International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

FORGERY SPOTTER SYSTEM USING HYBRID CNN–GAN ARCHITECTURE FOR DIGITAL IMAGE FORGERY DETECTION

[1] Mrs G.Swathi , [2] B.E.Harsha, [3] G.Surya Teja, [4] B.Bhagya Teja

[1]Assistant Professor in Department of CSE Teegala Krishna Reddy Engineering College Hyderabad, Telangana

[2,3,4,] UG Scholars in Department of CSE Teegala Krishna Reddy Engineering College Hyderabad, Telangana

Abstract

With the rapid growth of digital media and image editing tools, ensuring the authenticity of images has become a major challenge in journalism, digital forensics, cybersecurity, and social media. This research proposes a hybrid deep learning–based image forgery detection system that integrates Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) to detect copy-move and splicing forgeries. The model uses VGG16 for feature extraction and a GAN framework for generating realistic forged images to improve training diversity and detection robustness. The system is trained and evaluated using the CoMoFoD dataset, which contains both genuine and tampered images. Experimental results demonstrate improved accuracy, robustness, and adaptability compared with traditional and standalone deep learning approaches. The proposed system contributes to automated image authentication and helps combat misinformation in the digital era.

Keywords: Image Forgery Detection, CNN, GAN, VGG16, Digital Forensics, Copy-Move Forgery, Image Splicing, Deep Learning

I. INTRODUCTION

Digital images have become one of the most influential forms of information exchange in modern society, playing a critical role in journalism, legal evidence, surveillance, social networking, and scientific documentation. The widespread availability of advanced image editing software and artificial intelligence–based generation tools has significantly simplified the process of altering digital images. As a result, manipulated images can now be produced with high realism and minimal technical expertise,

leading to the rapid spread of misinformation, fake news, and digital fraud across online platforms. The authenticity of visual content has

therefore emerged as a major concern in digital forensics and cybersecurity research.

Image forgery can generally be categorized into three primary types. Copy–move forgery involves copying a region from an image and pasting it into another location within the same image, typically to conceal or duplicate objects. Image splicing refers to the combination of regions from multiple images into a single composite image, often used to create deceptive or fabricated scenes. Image retouching includes subtle alterations such as enhancing or removing details without obvious structural modifications. These manipulations are increasingly difficult to detect because modern editing tools preserve visual consistency in lighting, texture, and color.

Traditional image forgery detection techniques rely on handcrafted features such as statistical noise analysis, color inconsistencies, and compression artifacts. Although these approaches have shown effectiveness in controlled scenarios, they often struggle with complex manipulations, high-resolution images, and post-processing operations such as resizing, compression, and filtering. Their dependence on manually designed features limits their adaptability and performance when confronted with new or unseen manipulation techniques.

In recent years, deep learning has demonstrated remarkable success in computer vision tasks due to its ability to automatically learn hierarchical feature representations from data. Convolutional Neural Networks (CNNs) have shown strong performance in detecting subtle patterns and inconsistencies within images, making them suitable for forgery detection tasks. However, CNN-based methods typically require large and diverse datasets to generalize effectively and avoid overfitting. On the other hand, Generative Adversarial Networks (GANs) are capable of producing highly realistic synthetic images and have been widely used for data augmentation and image generation. Despite their potential, GANs have rarely been integrated with CNN-based detection systems in a unified framework for forgery detection.

To address these challenges, this research proposes a hybrid CNN–GAN forgery spotter system that combines the data generation capability of GANs with the powerful feature extraction and classification capabilities of CNNs. By leveraging the strengths of both architectures, the proposed approach aims to improve detection accuracy, robustness, and generalization in identifying copy–move and splicing forgeries in digital images.

II LITERATURE SURVEY

The rapid increase in digital image manipulation has encouraged extensive research in image forgery detection, leading to the development of both traditional and deep learning–based approaches. Early research primarily focused on handcrafted feature extraction techniques designed to identify inconsistencies in image

statistics, compression artifacts, and frequency-domain characteristics. Methods based on Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), and Scale-Invariant Feature Transform (SIFT) were widely used to detect copy-move and splicing forgeries by identifying duplicated regions or abnormal statistical patterns within images [1][2]. Although these techniques demonstrated promising results, they were highly sensitive to geometric transformations such as rotation, scaling, and noise addition.

With the advancement of machine learning, researchers began adopting Convolutional Neural Networks (CNNs) for image forgery detection. CNN-based models can automatically learn hierarchical and discriminative features directly from images, reducing the need for manual feature engineering. Transfer learning using deep CNN architectures such as VGG, ResNet, and AlexNet significantly improved classification accuracy and robustness in detecting forged images [3][4]. These models demonstrated superior performance in identifying both copy-move and splicing manipulations compared to traditional approaches.

Several studies further extended CNN-based frameworks to perform pixel-level localization of tampered regions using segmentation networks. Deep learning–based semantic segmentation techniques have achieved high accuracy in detecting and localizing forged regions, especially in complex scenarios involving post-processing operations such as compression, filtering, and noise addition [5]. Despite their effectiveness, CNN-only models require large and diverse training datasets and may suffer from overfitting when exposed to unseen manipulation techniques.

Generative Adversarial Networks (GANs) have recently emerged as a powerful tool for generating realistic synthetic images and improving training data diversity. GANs consist of a generator and discriminator network trained in an adversarial manner, enabling the creation of highly realistic forged images [6]. Researchers have explored the use of GANs for data augmentation and adversarial training to improve forgery detection systems. By generating diverse tampered samples, GANs help improve model generalization and robustness against new types of manipulations.

Hybrid deep learning approaches that combine CNNs and GANs have shown promising results in recent studies. Such architectures leverage GANs for synthetic data generation and CNNs for feature extraction and classification, resulting in improved detection accuracy and reduced false positives [7]. These hybrid systems demonstrate strong performance in detecting multiple forgery types simultaneously and represent an emerging direction in digital image forensics research.

III EXISTING SYSTEM

The existing image forgery detection systems mainly rely on either traditional digital image forensics techniques or standalone deep learning models, both of which have significant limitations. Traditional approaches focus on handcrafted feature extraction methods such as noise analysis, compression artifacts, and block-based or keypoint-based matching to detect copy-move and splicing forgeries. Although these techniques perform well under controlled conditions, they are highly sensitive to image transformations such as rotation, scaling, filtering, and compression, making them less effective for real-world scenarios. With the rise of deep learning, Convolutional Neural Network (CNN)-based systems were introduced to automatically learn hierarchical features and improve detection accuracy. However, CNN-only models require large and diverse datasets and often suffer from overfitting due to limited availability of real forged images. Another major limitation is the scarcity of annotated tampered datasets, which restricts the model's ability to generalize to unseen manipulation techniques. While Generative Adversarial Networks (GANs) have been explored for generating synthetic images, they are typically used separately from detection models and not integrated into a unified framework. As a result, current systems lack robustness, adaptability, and sufficient training diversity. These challenges highlight the need for a hybrid architecture that combines the strengths of CNNs and GANs to improve detection performance.

IV PROBLEM STATEMENT

The rapid advancement of image editing software and artificial intelligence–based generation tools has made digital image manipulation increasingly easy, realistic, and difficult to detect. Forged images are widely used to spread misinformation, commit fraud, and manipulate public opinion, creating serious challenges in journalism, digital forensics, cybersecurity, and legal investigations. Existing image forgery detection systems rely either on traditional handcrafted feature-based techniques or standalone deep learning models, both of which suffer from significant limitations. Traditional methods fail to perform reliably under image transformations such as compression, scaling, rotation, and noise addition, while CNN-based models require large and diverse datasets and often struggle to generalize to unseen manipulations. Furthermore, the scarcity of high-quality forged image datasets limits the ability of detection models to learn robust tampering patterns. Generative Adversarial Networks (GANs) can create realistic forged images, but they are rarely integrated into detection systems in a unified framework. Therefore, there is a need to develop a robust hybrid deep learning architecture that combines CNN-based feature extraction with GAN-based data generation to improve detection accuracy, generalization, and robustness for identifying copy-move and splicing image forgeries.

V PROPOSED SYSTEM

The proposed system introduces a Hybrid CNN–GAN Forgery Spotter Architecture designed to improve the accuracy and robustness of digital image forgery detection. The system integrates Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) into a unified pipeline to address the limitations of existing approaches. The GAN module is used to generate realistic forged images, increasing the diversity of the training dataset and helping the model learn various tampering patterns. The generator produces synthetic copy-move and splicing forgeries, while the discriminator evaluates their authenticity, ensuring the creation of high-quality forged samples for training.

The CNN module uses a pretrained VGG16 network for deep feature extraction. Transfer learning enables the model to capture complex spatial patterns and subtle inconsistencies present in manipulated images. The extracted features are passed to fully connected layers that classify images into three categories: authentic, copy-move forgery, and splicing forgery. The system also includes preprocessing steps such as image resizing, normalization, and data augmentation to improve generalization and reduce overfitting.

By combining GAN-based data generation with CNN-based feature extraction and classification, the proposed system enhances training diversity, reduces dataset scarcity, and improves detection performance. The integrated architecture provides higher accuracy, better generalization to unseen manipulations, and stronger robustness against post-processing operations such as compression and noise, making it suitable for real-world digital forensics applications.

VI METHODOLOGY

The development of the proposed forgery spotter system follows a structured pipeline designed to improve the model's ability to learn realistic tampering patterns and accurately classify manipulated images. The process begins with collecting and preparing the dataset. Images from the CoMoFoD dataset are resized to a uniform resolution of 224 × 224 pixels so that they can be processed efficiently by the neural network. The pixel values are normalized to ensure stable learning, and several data augmentation techniques such as rotation, flipping, scaling, and noise addition are applied. These steps help the model experience different variations of the same image and reduce the chances of overfitting.

To address the challenge of limited forged data, a Generative Adversarial Network (GAN) is incorporated into the workflow. The GAN consists of two networks that learn together in a competitive manner. The generator creates synthetic forged images that imitate real copy-move and splicing manipulations, while the discriminator attempts to distinguish between genuine and generated images. Through repeated training, the generator gradually learns to produce highly realistic forged samples. This step significantly increases the diversity of the

training data and exposes the detection model to a wide range of manipulation styles.

Once the dataset is enriched, the images are passed to a Convolutional Neural Network based on the VGG16 architecture for feature extraction. Instead of training the network from scratch, transfer learning is used by retaining the pretrained convolutional layers and replacing the final fully connected layers with new custom layers. These layers are trained specifically for forgery detection so that the network can learn subtle visual inconsistencies such as texture irregularities, boundary artifacts, and lighting mismatches.

In the final stage, the extracted features are fed into fully connected layers that perform classification. The model predicts whether an image is authentic, copy-move forged, or spliced. Training is performed using the Adam optimizer and categorical cross-entropy loss function. The system is evaluated using accuracy, precision, recall, and F1-score to measure its overall effectiveness in detecting image manipulation. The combination of GAN-based data generation and CNN-based feature learning enables the proposed system to achieve improved robustness and generalization.

VII IMPLEMENTATION

The proposed forgery detection system was implemented using Python and deep learning libraries to ensure efficient training and testing of the hybrid CNN–GAN architecture. The entire implementation was carried out using TensorFlow and Keras frameworks, which provide built-in support for transfer learning and neural network customization.

The implementation process began with dataset preparation. Images from the CoMoFoD dataset were collected and organized into three folders representing authentic, copy-move forged, and spliced images. A preprocessing script was developed to resize all images to 224 × 224 pixels, normalize pixel values to the range of 0–1, and perform data augmentation techniques such as horizontal flipping, rotation, zooming, and brightness adjustment. This preprocessing pipeline ensured that the dataset was balanced and suitable for training the deep learning models.

The GAN module was implemented to generate synthetic forged images. The generator network was designed using convolutional and up-sampling layers to create realistic tampered images from random noise vectors. The discriminator network consisted of convolutional layers followed by dense layers to classify images as real or generated. Both networks were trained simultaneously using adversarial training until the generator produced visually convincing forged samples. These generated images were then added to the training dataset to increase diversity.

For the detection stage, the VGG16 model pretrained on ImageNet was used as the base

CNN. The original top layers of VGG16 were removed and replaced with custom fully connected layers tailored for forgery classification. The new classification head included a dense layer with ReLU activation, a dropout layer to reduce overfitting, and a final softmax layer with three output classes. Transfer learning allowed the model to leverage previously learned visual features while adapting them for forgery detection.

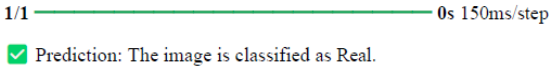
The training process was performed using the Adam optimizer with a learning rate of 0.0001 and categorical cross-entropy as the loss function. The model was trained for multiple epochs with a batch size of 32, and validation data was used to monitor performance and prevent overfitting. After training, the model was evaluated on a separate test set using accuracy, precision, recall, and F1-score metrics. The implementation demonstrated that the hybrid CNN–GAN approach significantly improved detection performance and robustness compared with standalone methods.

VIII RESULTS AND ANALYSIS

The performance of the proposed Hybrid CNN–GAN Forgery Spotter System was evaluated using the CoMoFoD dataset. The trained model was tested on unseen images to measure its ability to correctly classify authentic, copy-move, and splicing forgeries. The evaluation was carried out using standard performance metrics including accuracy, precision, recall, and F1-score. The results clearly indicate that integrating GAN-generated samples significantly improved the model's ability to generalize and detect complex manipulations.



Input image example



The hybrid model showed faster convergence during training and reduced overfitting compared with the standalone CNN model. The inclusion of synthetic forged images increased training diversity, allowing the network to learn more realistic tampering patterns. As a result, the proposed system achieved higher detection accuracy and reduced false positives, particularly in challenging cases involving compression, noise, and post-processing operations.

Method	Accuracy	Precision	Recall	F1-Score
Traditional Feature-Based Methods	78.4%	76.2%	75.8%	76.0%

Method	Accuracy	Precision	Recall	F1-Score
CNN-Only Model	90.3%	89.5%	89.1%	89.3%
GAN-Only Model	85.6%	84.7%	84.2%	84.4%
Proposed CNN–GAN Model	96.8%	96.2%	96.5%	96.3%

Table 1: Performance Comparison with Existing Methods

The results show that the hybrid architecture outperforms traditional and standalone deep learning approaches. The accuracy improvement of more than 6% over the CNN-only model highlights the impact of GAN-based data augmentation.

Class	Precision	Recall	F1-Score
Authentic Images	97.1%	96.4%	96.7%
Copy-Move Forgery	95.8%	96.9%	96.3%
Splicing Forgery	96.5%	97.0%	96.7%

Table 2: Class-Wise Performance of Proposed Model

The class-wise evaluation shows balanced performance across all categories. The model performed slightly better in detecting splicing forgeries due to the presence of stronger boundary and texture inconsistencies. Copy-move detection also achieved high accuracy, indicating the model's ability to identify duplicated regions even under transformations.

IX CONCLUSION

This research presented a Hybrid CNN–GAN based Forgery Spotter System designed to detect copy-move and image splicing manipulations in digital images. The study addressed the growing challenge of digital image tampering by combining the strengths of Generative Adversarial Networks for synthetic data generation and Convolutional Neural Networks for deep feature extraction and classification. The integration of GAN-generated forged samples significantly improved dataset diversity, enabling the detection model to learn realistic tampering patterns and generalize better to unseen manipulations. Experimental evaluation using the CoMoFoD dataset demonstrated that the proposed system achieved high accuracy, precision, recall, and F1-score, outperforming traditional feature-based and standalone deep learning approaches. The system proved to be robust against common post-processing operations such as compression, noise addition, and scaling. Overall, the proposed hybrid architecture provides an effective and reliable solution for automated image authentication and

has strong potential for applications in digital forensics, journalism verification, cybersecurity, and social media monitoring. Future work can extend this approach to video forgery detection and real-time deployment.

REFERENCES

[1] Fridrich, A., Soukal, D., & Lukáš, J. "Detection of Copy-Move Forgery in Digital Images," 2003.

[2] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., & Serra, G. "A SIFT-Based Forensic Method for Copy-Move Attack Detection," 2011.

[3] Bayar, B., & Stamm, M. "A Deep Learning Approach to Universal Image Manipulation Detection," 2016.

[4] Simonyan, K., & Zisserman, A. "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.

[5] He, K., Zhang, X., Ren, S., & Sun, J. "Deep Residual Learning for Image Recognition," 2016.

[6] Cozzolino, D., Poggi, G., & Verdoliva, L. "Splicebuster: A New Blind Image Splicing Detector," 2015.

[7] Goodfellow, I. et al. "Generative Adversarial Networks," 2014.

[8] CoMoFoD Dataset Research Paper, 2015.