iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76618

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 13 Issue XII Dec 2025- Available at www.ijraset.com

A Comparative Study of Spatial and Frequency-Domain CNN Models for Detecting AI-Generated Images

Harshpratap Singh Rathore[1], Rishi Modh[2], Kanhaiya Suthar[3]
[1, 2, 3]*UG Student – B.V Patel Institute of Computer Science, Uka Tarsadia University, Bardoli, Gujarat-394350, India*

Abstract: ***The rapid advancement of generative artificial intelligence has led to the widespread availability of highly photorealistic AI-generated images, making it increasingly difficult to distinguish synthetic content from real images using manual inspection. This raises serious concerns related to misinformation, digital forgery, and loss of trust in visual media. To address this challenge, this paper presents a comparative study between two deep learning approaches for AI-generated image detection: a baseline Convolutional Neural Network (CNN) operating in the spatial domain and a hybrid CNN model incorporating Fast Fourier Transform (FFT) based frequency-domain features.***
A large and heterogeneous dataset consisting of real images and AI-generated images from multiple generative sources was constructed and carefully preprocessed to simulate real-world conditions such as resizing and compression. The dataset was split into training, validation, and test sets using strict separation to prevent data leakage. Both models were trained under identical experimental conditions for 15 epochs using the same optimization strategy. Experimental results show that the spatial-domain CNN achieved a test accuracy of approximately 85.3%, outperforming the CNN+FFT hybrid model, which achieved an accuracy in the range of 76–78%.
The findings demonstrate that, contrary to common assumptions, frequency-domain features do not necessarily enhance AI-generated image detection under realistic preprocessing constraints. This study highlights the robustness of spatial-domain CNNs and emphasizes the importance of empirical evaluation when designing deepfake detection systems.
Keywords: AI-Generated Images, Fake Image Detection, Convolutional Neural Network, FFT, Deep Learning, Image Forensics

I. INTRODUCTION

The rapid advancement of generative artificial intelligence has enabled the creation of highly photorealistic AI-generated images that closely resemble real-world photographs. While these developments support creative and industrial applications, they also raise serious concerns related to misinformation, digital forgery, and loss of trust in visual media. The widespread availability of such generative tools has made manual verification unreliable, creating a strong demand for automated and robust detection systems capable of distinguishing AI-generated images from real images.

Traditional image forensics approaches relied on handcrafted features such as noise inconsistencies, compression artifacts, and lighting irregularities. However, these methods struggle to generalize to modern AI-generated images, which are trained to replicate natural image statistics. As a result, deep learning-based methods, particularly Convolutional Neural Networks (CNNs), have become the dominant solution for fake image detection. CNNs learn hierarchical feature representations directly from pixel data, enabling effective discrimination between real and synthetic images under diverse conditions.

Recent studies have explored both spatial-domain and frequency-domain approaches for AI-generated image detection. Spatial-domain CNNs focus on learning semantic and structural image patterns, while frequency-domain methods attempt to capture spectral artifacts using transformations such as the Fast Fourier Transform (FFT). Although frequency-domain features are often assumed to enhance detection performance, their effectiveness under real-world preprocessing operations such as resizing and compression remains unclear. To address this gap, this paper presents a comparative analysis of a spatial-domain CNN and a CNN integrated with FFT-based frequency features, evaluated under identical experimental conditions to assess their practical effectiveness.

II. LITERATURE REVIEW

Recent studies on AI-generated image detection have primarily focused on deep learning-based approaches, especially Convolutional Neural Networks (CNNs), due to their strong ability to learn discriminative spatial features from images.

©IJRASET: All Rights are Reserved

Marra et al. [1] demonstrated that CNN-based models can effectively detect GAN-generated images shared over social networks by identifying subtle visual inconsistencies. Similarly, Wang et al. [2] showed that deep CNNs trained on synthetic datasets are capable of distinguishing real and fake images with high accuracy, establishing spatial-domain learning as a strong baseline for fake image forensics. Cozzolino et al. [3] further explored forensic analysis of deepfake images and highlighted the effectiveness of CNNs in capturing generation artifacts that are not easily perceptible to the human eye.

Alongside spatial-domain methods, several researchers have investigated frequency-domain analysis for detecting AI-generated images. Durall et al. [7] revealed that many generative models fail to accurately reproduce natural image spectral distributions, suggesting that frequency-domain representations can expose hidden generation artifacts. Building on this insight, Zhang et al. [5] proposed frequency-aware detection methods using spectral features and reported improved performance on controlled datasets. Frank et al. [6] also explored the use of frequency-domain information combined with CNNs, concluding that spectral features may provide complementary cues for identifying synthetic images under certain conditions.

Despite these promising findings, the reliability of frequency-domain features remains inconsistent across studies. Many existing works evaluate their methods on high-resolution or minimally processed images, which do not accurately represent real-world scenarios where images are commonly resized, compressed, and re-encoded before analysis. FaceForensics++ [4] highlighted the impact of preprocessing on detection performance, showing that model accuracy can degrade significantly under realistic transformations. This reveals a clear research gap: limited comparative evaluation of spatial and frequency-domain learning under identical training conditions and realistic preprocessing constraints. To address this gap, the present study performs a controlled comparison between a spatial-domain CNN and a CNN integrated with FFT-based frequency features, using the same dataset, preprocessing pipeline, and experimental setup.

III. DATASET AND PREPROCESSING

A. *Dataset Construction*

To evaluate AI-generated image detection under realistic conditions, a large and heterogeneous dataset was constructed consisting of real images and AI-generated images. Real images were collected from publicly available natural image sources, while AI-generated images were obtained from multiple modern generative models to ensure diversity in texture, structure, and visual artifacts. The dataset was deliberately balanced to avoid class bias during model training and evaluation.

Table 1DATASET COMPOSITION

CLASS	NUMBER OF IMAGES	PERCENTAGE (%)	DESCRIPTION
REAL IMAGES	30,000	50%	Natural photographs with variations in lighting, resolution, and compression
AI-GENERATED IMAGES	30,000	50%	Images generated using multiple AI generative models
TOTAL	60,000	100%	Combined dataset used for experiments

B. *Preprocessing and Optimization*

To ensure compatibility with CNN architectures and simulate real-world image pipelines, all images were subjected to a standardized preprocessing workflow. Images were resized to a fixed resolution and converted to RGB format to maintain consistency. Pixel values were normalized to stabilize training and accelerate convergence. These preprocessing steps reflect common transformations applied during social media uploads and online sharing.

As a result of resizing and optimization, the effective dataset size was reduced from approximately **52 GB to 6.24 GB**, enabling efficient model training while preserving discriminative visual information.

Table 2PREPROCESSING PIPELINE

STEP	DESCRIPTION
IMAGE RESIZING	All images resized to a fixed resolution compatible with CNN input
COLOR NORMALIZATION	Images converted to RGB format

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

PIXEL NORMALIZATION	Pixel values scaled to a standard range
DATASET OPTIMIZATION	Reduced storage size from 52 GB to 6.24 GB

C. *Train–Validation–Test Split*

To prevent data leakage and ensure fair evaluation, the dataset was strictly divided into **training**, **validation**, and **test** sets. Each subset maintained an equal distribution of real and AI-generated images. No overlap was allowed between the splits, ensuring that all performance metrics reflect true generalization on unseen data.

Table 3Train–Validation–Test Split Distribution

DATASET SPLIT	REAL IMAGES	AI-GENERATED IMAGES	TOTAL IMAGES
TRAINING SET (70%)	21,000	21,000	42,000
VALIDATION SET (15%)	4,500	4,500	9,000
TEST SET (15%)	4,500	4,500	9,000
TOTAL	30,000	30,000	60,000

IV. METHODOLOGY

This study evaluates two deep learning approaches for AI-generated image detection:
(1)a baseline spatial-domain CNN, and
(2) a hybrid CNN model integrating frequency-domain features using Fast Fourier Transform (FFT).
Both models were trained and evaluated under identical conditions to ensure a fair and controlled comparison.

A. *Baseline CNN Architecture (Spatial Domain)*

The baseline model is a custom-designed Convolutional Neural Network trained directly on RGB images in the spatial domain. The architecture follows a hierarchical feature extraction strategy, where shallow layers capture low-level features such as edges and textures, while deeper layers learn higher-level semantic and structural representations.
The network consists of multiple convolutional blocks, each containing a convolutional layer followed by a ReLU activation function and a pooling layer to reduce spatial dimensions and mitigate overfitting. After feature extraction, the learned representations are flattened and passed through fully connected layers. A sigmoid activation function is used in the final output layer to perform binary classification between real and AI-generated images.

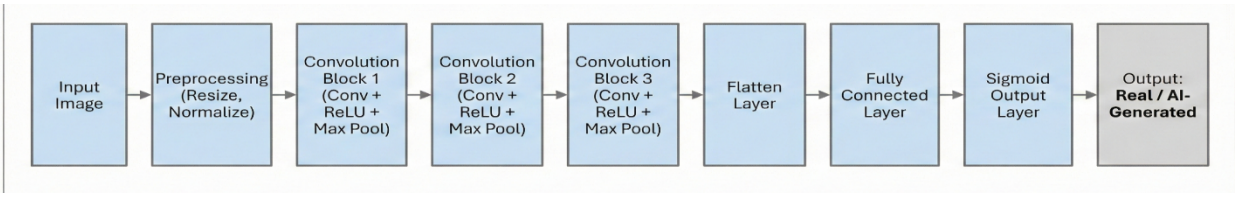


Figure 1Architecture of the spatial-domain CNN model

B. *CNN + FFT Hybrid Model (Spatial + Frequency Domain)*

To investigate the contribution of frequency-domain information, a hybrid CNN model was implemented by integrating FFT-based features with spatial-domain learning. For this approach, each input image is first transformed into the frequency domain using the Fast Fourier Transform. The magnitude spectrum of the FFT output is computed to capture frequency characteristics while discarding phase information.
The hybrid model consists of two parallel branches. The **spatial branch** follows the same CNN architecture used in the baseline model and processes the RGB image directly. The **frequency branch** processes the FFT magnitude spectrum using additional convolutional layers designed to learn discriminative frequency patterns. The outputs of both branches are concatenated and passed through fully connected layers for final classification.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

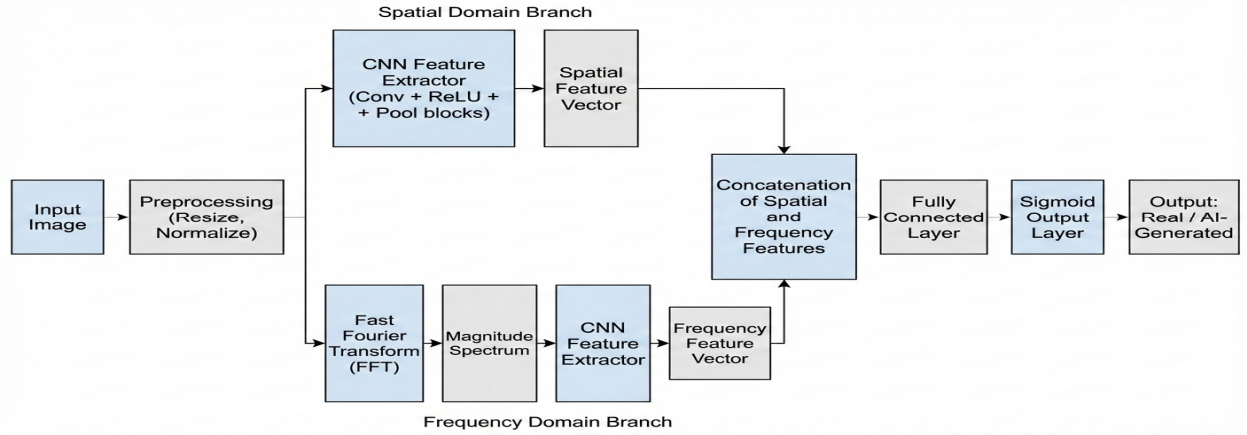


Figure 2Architecture of the CNN+FFT hybrid model

C. Model Comparison Strategy

To ensure a fair comparison between the two approaches, both models were trained using the same dataset, preprocessing pipeline, and training configuration. No additional data augmentation or class rebalancing techniques were applied. Performance differences are therefore attributed solely to architectural design choices rather than implementation bias.

Table 4MODEL COMPARISON PROTOCOL

PARAMETER	BASELINE CNN	CNN + FFT
DATASET	Same	Same
PREPROCESSING	Same	Same
OPTIMIZER	Same	Same
EPOCHS	15	15
EVALUATION METRICS	Accuracy, Precision, Recall, F1-score	Same

V. EXPERIMENTAL SETUP

All experiments were conducted under a controlled and consistent environment to ensure a fair comparison between the spatial-domain CNN and the CNN+FFT hybrid model. Both models were implemented using the same deep learning framework, dataset, preprocessing pipeline, and training configuration. This ensures that any observed performance differences are solely due to architectural design choices rather than experimental bias.

A. Implementation Environment

The models were implemented using Python and a deep learning framework suitable for convolutional neural network training. Training and evaluation were performed in a GPU-enabled environment to handle the large-scale dataset efficiently. The dataset was stored on cloud-based storage and accessed during training.

Table 5IMPLEMENTATION ENVIRONMENT

COMPONENT	DESCRIPTION
PROGRAMMING LANGUAGE	Python
DEEP LEARNING FRAMEWORK	TensorFlow / Keras
EXECUTION PLATFORM	GPU-enabled environment
DATASET STORAGE	Cloud-based storage
REPRODUCIBILITY	Fixed random seed

VI. RESULTS AND DISCUSSION

This section presents the experimental results obtained from the spatial-domain CNN and the CNN+FFT hybrid model, followed by a detailed comparative analysis. Both models were evaluated under identical conditions using the same dataset, preprocessing pipeline, and training configuration to ensure a fair comparison.

B. Baseline CNN Results (Spatial Domain)

The baseline spatial-domain CNN demonstrated stable learning behavior across training epochs. Validation performance steadily improved and reached its peak near the final epochs, indicating effective feature learning without severe overfitting. When evaluated on the unseen test set, the model achieved a test accuracy of approximately **85.3%**, demonstrating strong generalization performance.

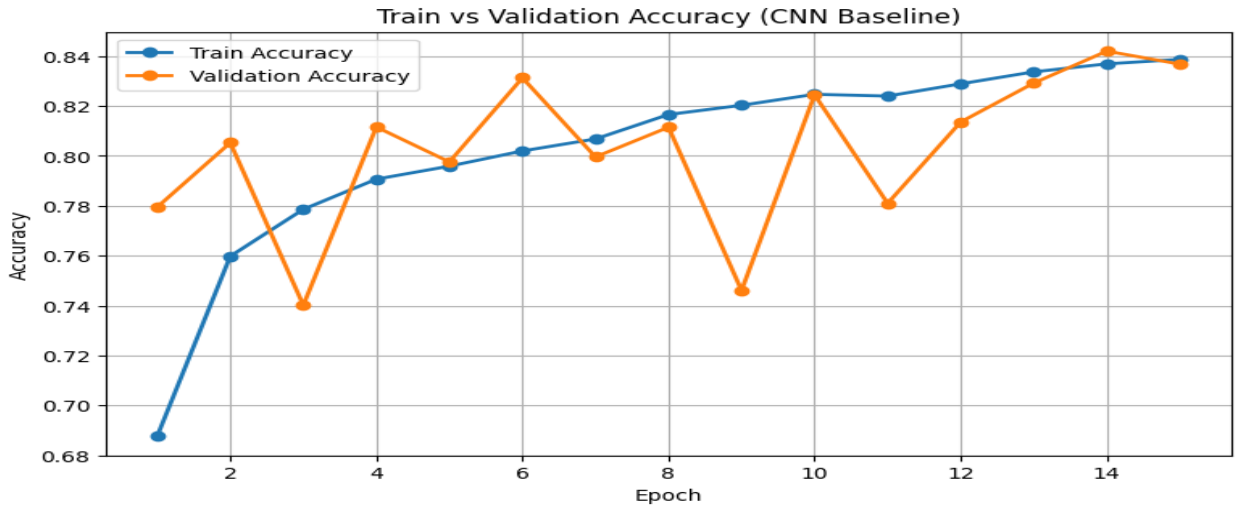


Figure 3Training and validation accuracy of the spatial-domain CNN

The relatively small gap between training and validation accuracy suggests that the spatial-domain CNN learned robust and transferable features. These results indicate that spatial cues alone are sufficient to capture meaningful differences between real and AI-generated images under realistic preprocessing conditions.

C. CNN + FFT Hybrid Model Results

The CNN+FFT hybrid model was trained under the same configuration for 15 epochs. Despite the additional frequency-domain information and increased architectural complexity, the hybrid model achieved a lower test accuracy in the range of **76–78%**. Compared to the baseline CNN, validation accuracy showed less stability across epochs, indicating difficulty in effectively integrating spatial and frequency-domain features.

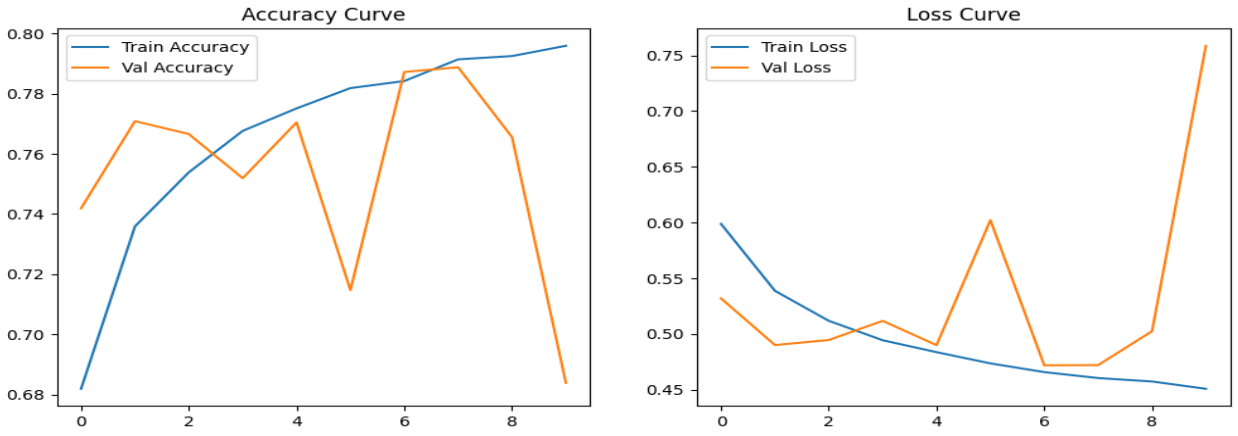


Figure 4 Training and validation accuracy of the CNN+FFT hybrid model

Although the hybrid model was able to learn discriminative patterns to some extent, the inclusion of FFT-based features did not translate into improved performance. This suggests that frequency-domain information, when applied to resized and heterogeneous datasets, may lose discriminative power or introduce noise that negatively affects learning.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Comparative Analysis and Discussion

A direct comparison between the two models reveals that the spatial-domain CNN consistently outperformed the CNN+FFT hybrid model across all evaluation metrics. This outcome contradicts the commonly held assumption that frequency-domain features inherently enhance AI-generated image detection.

One plausible explanation for this behavior lies in the preprocessing pipeline. Image resizing, compression, and normalization—common in real-world image sharing platforms—can suppress or distort high-frequency components that frequency-based methods rely on. As a result, FFT-derived features may become less informative or even misleading when applied to heterogeneous datasets.

In contrast, the spatial-domain CNN appears to learn more robust visual patterns that remain effective even after aggressive preprocessing. These findings highlight that increased model complexity does not necessarily lead to improved performance and emphasize the importance of empirical evaluation when designing practical AI-generated image detection systems.

Table 6COMPARATIVE PERFORMANCE OF PROPOSED MODELS

MODEL	TEST ACCURACY	KEY OBSERVATION
SPATIAL-DOMAIN CNN	85.3%	Robust and stable under realistic preprocessing
CNN + FFT HYBRID	76–78%	Frequency features did not improve performance

VII. CONCLUSION

This study conducted a comparative evaluation of two deep learning approaches for AI-generated image detection: a spatial-domain Convolutional Neural Network (CNN) and a hybrid CNN model incorporating Fast Fourier Transform (FFT) based frequency-domain features. Both models were trained and tested under identical conditions using a large, balanced, and heterogeneous dataset designed to reflect realistic image preprocessing scenarios. The experimental results showed that the spatial-domain CNN achieved a higher test accuracy of approximately **85.3%**, whereas the CNN+FFT hybrid model achieved a lower accuracy in the range of **76–78%**, indicating inferior generalization performance.

The findings demonstrate that, under practical conditions involving image resizing and normalization, frequency-domain features do not necessarily provide additional discriminative power for AI-generated image detection. Instead, spatial-domain CNNs were observed to learn more robust and transferable visual patterns, leading to improved detection accuracy. This study emphasizes the importance of empirical validation over theoretical assumptions and confirms that simpler spatial-domain architectures can outperform more complex hybrid models in real-world deployment scenarios.

REFERENCES

[1] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, “Detection of GAN-generated fake images over social networks,” Proc. IEEE Conf. Multimedia Information Processing and Retrieval (MIPR), 2018, pp. 384–389, doi: 10.1109/MIPR.2018.00084.

[2] H. Wang, Z. Wang, J. Zhang, and Y. Wang, “CNN-based detection of synthetic images generated by GANs,” IEEE Access, vol. 7, pp. 140774–140785, 2019, doi: 10.1109/ACCESS.2019.2944107.

[3] D. Cozzolino, J. Thies, A. Rössler, M. Nießner, and L. Verdoliva, “Forensic analysis of image-based deepfake,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–922, 2020, doi: 10.1109/JSTSP.2020.3002111.

[4] A. Rössler et al., “FaceForensics++: Learning to detect manipulated facial images,” Proc. IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1–11, doi: 10.1109/ICCV.2019.00009.

[5] Y. Zhang, F. Ding, and S. Kwong, “Frequency-aware GAN detection using spectral features,” Pattern Recognition, vol. 122, 2022, doi: 10.1016/j.patcog.2021.108306.

[6] H. Frank, O. Alshaabi, and P. Cunningham, “Leveraging frequency domain information for detecting AI-generated images,” Signal Processing: Image Communication, vol. 99, 2021, doi: 10.1016/j.image.2021.116450.

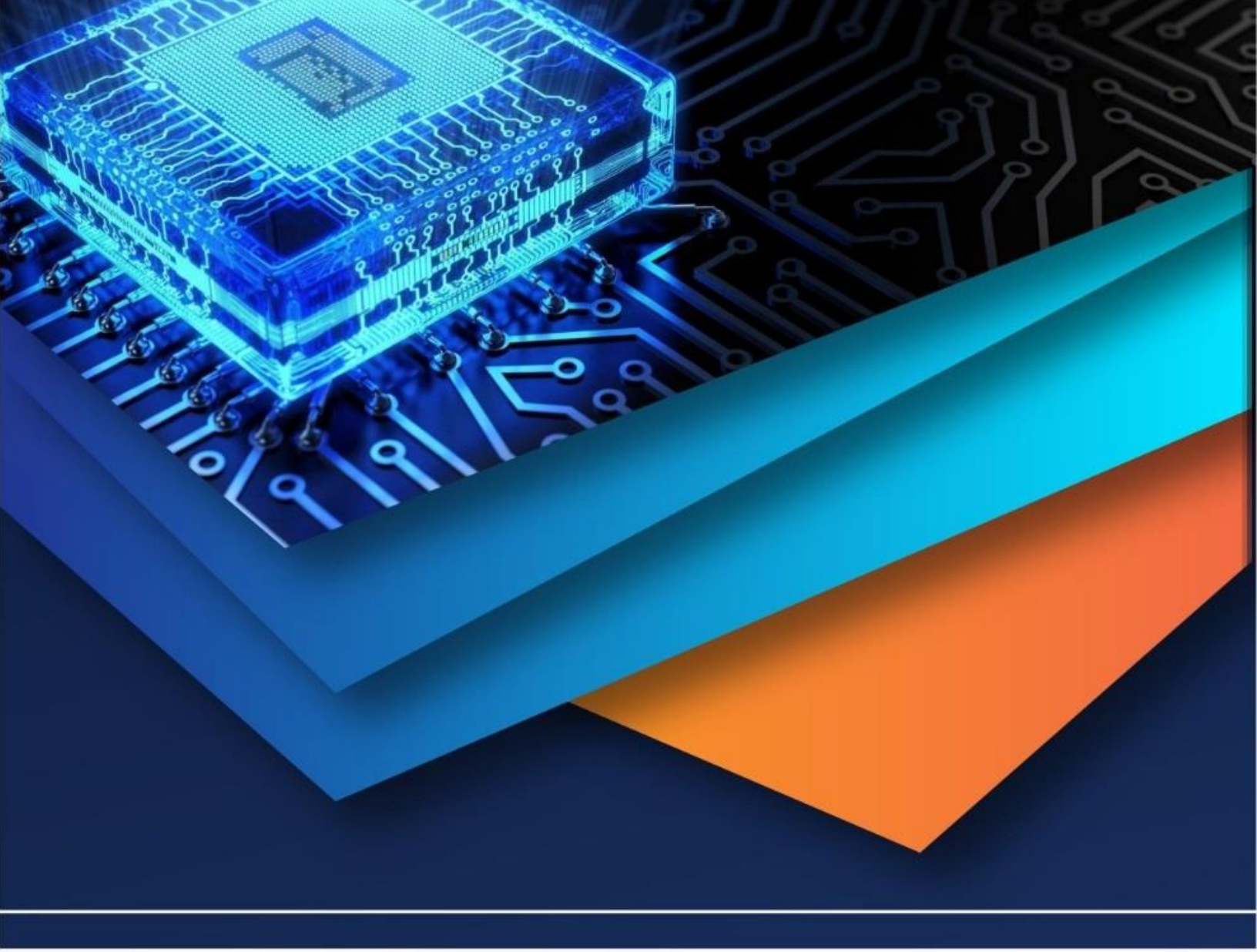
[7] J. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, “Watch your up-convolution: CNN-based generative deep neural networks are failing to reproduce spectral distributions,” Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 7890–7899, doi: 10.1109/CVPR42600.2020.00791.

[8] R. Durall, C. Keuper, and J. Keuper, “On the detection of GAN-generated images using frequency analysis,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 2962–2966.

[9] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. Woo, “GAN-generated face detection using CNNs,” IEEE Access, vol. 7, pp. 45356–45365, 2019, doi: 10.1109/ACCESS.2019.2909030.

[10] L. Verdoliva, “Media forensics and deepfakes: An overview,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020, doi: 10.1109/JSTSP.2020.3002101.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)