iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67847

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 III Mar 2025- Available at www.ijraset.com

Deepfake Image Detection Using CNN and LBP Based Techniques

Y. Ravindra Nadh[1], B. Lavanya[2], G. Vishal[3], G. Sumanth[4]
Raghu institute of Technology

Abstract*: The rise of AI-based tools has made it easier to create highly realistic deepfake videos, which pose significant risks such as political manipulation, blackmail, and the fabrication of terrorism-related content. These manipulated videos are often difficult to detect due to their minimal traces of alteration. In this project, we propose a Local Binary Pattern-based Convolutional Neural Network (LBPNET) for deepfake detection. The approach involves extracting Local Binary Pattern (LBP) features from facial images and training a Convolutional Neural Network (CNN) on these descriptors to develop a robust classification model. When a new test image is uploaded, the trained model evaluates it to determine whether it is real or fake.This method enhances the reliability of deepfake detection by leveraging texture-based feature extraction and deep learning. The following sections provide further details on the LBP technique and its role in our detection framework.
Keywords-convolutional network, Local binary patterns

I. INTRODUCTION

Two important deep learning methods for identifying fake faces are ANNs and CNNs. Artificial neural networks, or ANNs, are trained to recognize irregularities in facial features, including misaligned features, artifacts, and variations in lighting and shading. Convolutional neural networks (CNNs) are specialized neural networks made for image and spatial data. CNNs analyze images and identify discrepancies in them. Blockchain and authentication techniques are also being investigated to ensure the authenticity of digital media, making it more difficult for fake content to spread without verification. In order to identify irregularities and discrepancies in fraudulent content that automated tools might overlook, human expert analysis is frequently employed. To sum up, identifying phony faces is a challenging task that calls for a mix of methods and professional analysis. To make it harder for fraudulent content to proliferate unchecked, blockchain and authentication methods are being investigated to confirm the legitimacy of digital media. Convolutional Neural Networks (CNNs) are deep learning algorithms that are used in machine learning and computer vision. They are especially useful for image recognition, object detection, and image generation. These algorithms use convolution to learn hierarchical patterns and features from input images, allowing them to identify complex patterns. CNNs are commonly used in applications like facial recognition, self-driving cars, and medical image analysis.

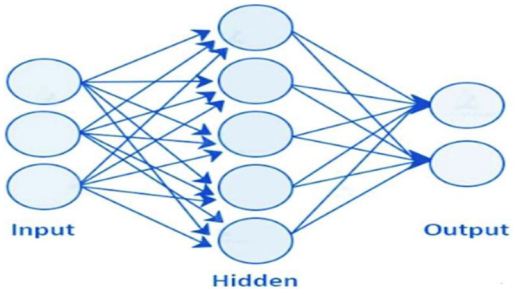
Recurrent Neural Networks (RNNs) are used to process sequential data, such as time series or natural language text. GANs, which consist of a generator and a discriminator, are used to generate images, transfer styles, and augment data. Autoencoders are used for unsupervised learning and feature extraction, whereas Deep Reinforcement Learning (RL) combines deep learning and reinforcement learning to train agents to make sequential decisions. TensorFlow, PyTorch, Keras, and Caffe are examples of deep learning frameworks that come with pre-built neural network layers, optimization algorithms, and GPU support to help you build your models quickly. Attention mechanisms, first popularized by the Transformer architecture, enable models to focus on relevant portions of input data, improving performance on tasks such as machine translation and image captioning.

Convolutional Neural Networks (CNNs) are deep learning algorithms used in machine learning and computer vision, with particular success in image recognition, object detection, and image generation tasks. These algorithms use convolution to learn hierarchical patterns and features from input images, allowing them to identify complex patterns. CNNs are commonly used in applications like facial recognition, self-driving cars, and medical image analysis. Recurrent Neural Networks (RNNs) are used to process sequential data like time series or natural language text. GANs, which include a generator and a discriminator, are used for tasks such as image generation, style transfer, and data augmentation. Autoencoders are used for unsupervised learning and feature extraction, whereas Deep Reinforcement Learning (RL) combines deep learning and reinforcement learning to train agents to make sequential decisions in environments in order to maximise cumulative rewards. TensorFlow, PyTorch, Keras, and Caffe are examples of deep learning frameworks that include pre-built neural network layers, optimization algorithms, and GPU support for efficient model development. Attention mechanisms, originally popularized by the Transformer architecture, allow models to focus on relevant parts of the input data, improving performance on tasks such as machine translation and image

II. RELATED WORK

The widespread use of Deep Fakes is due to the high quality of the faked movies and the ease with which their programs can be used by a wide range of users, from professionals to novices with varying levels of programming ability. These apps are typically created using deep learning methods. It is widely acknowledged that deep learning can successfully represent complex and high-dimensional data. A type of deep network known as deep autoencoders has been widely used for dimensionality reduction, as well as image compression. The first attempt at deep-fake creation was FakeApp, created by an Internet user using the auto encoder-decoder pairing structure. However, created a stunning Deep Fake data set consisting entirely of 620 videos. They applied the GAN model to the Deep Fake data set. Deep Fake film was made with low and high-quality Faceswap-GAN Open Source Code Videos from the VidTIMIT website, which can accurately mimic facial gestures, lip movements, and eye blinking. These films were also used to test a number of advanced false detection techniques. Different approaches to identifying Deep Fake films from this newly created data set, such as lip-syncing methods and support vector machine (SVM) picture quality metrics, produce extremely high error rates. Deep Fake is another technique used by cybercriminals to bypass authentication or identity checks and gain unauthorised access. Deep learning tools such as CNN and GAN have made it more difficult for forensic models to preserve facial characteristics and posture in switched-face images, as well as the lighting in the photographs. Zhang et al. used the bag of words method to extract a set of condensed traits, which they then fed into classification algorithms such as SVM, random forest, and multi-layer perceptrons (MLP) to distinguish from the original swapped face photographs. GAN models can learn how to disperse detailed input data, so their synthesised images are accurate and high-quality, making them potentially the most difficult deep learning-generated images to categorize. A recent study identified artificial neural networks (ANNs) as having some fundamental ideas derived from how the human brain operates. The architecture of (ANN) Figure 2 displays the ANNs' architecture. An input layer, possibly many hidden layers, and an output layer are the many layers that make up neural networks. A data set serves as the neural network's input. Neural networks are specifically designed to anticipate and categorize these data into predetermined buckets. The research on methods for detecting fake images and videos is included in this section. The goal of Philip S. Yuet al. (2018) was to employ TI-CNN. By projecting explicit and latent characteristics into a single feature space, TI-CNN is trained simultaneously with text and image input. According to Aswini Thota et al. (2018), a precisely calibrated Tiff-IDF-Dense neural network (DNN) model was used to detect fake news with an accuracy of 94.21 percent on test data. Artificial intelligence and deep learning techniques have been used by the research community to detect fake content. Shorten, Connor, and others (2019) The models achieve 50.99 percent accuracy on the CIFAR-10 dataset compared to 70.06 percent accuracy on the CIFAR-10 dataset when evaluated on supplemented test data. A fake news detection method using SVM, Naive Bays, and Logistic Regression datasets was proposed by Krithi Dinesh et al. in 2019. The impact of compression on the detestability of cutting-edge manipulation algorithms was the main focus of Andreas Rossleret al.'s(2019)project, and a standardized baseline for further research is suggested.

Furthermore, Bowen Dong et al. (2019) described the Fake Detector framework, which is made up of two main parts: credibility label inference and portrayal of characteristics learning. These two parts work together to create a deep diffusive network model. A model for detecting false news was developed by Hyeong-Jun Kim et al. in 2019. A number of procedures based on "Fast text" and "Shallow-and-wide CNN" were used and modified. To identify threats and phony photos, Njood Mohammed et al. (2019) want to create a model for categorizing Instagram content. CNN, the Alexnet network, and transfer learning with Alexnet were the deep algorithms used to build the model. According to Md. Rafiqul et al. (2020), GRU (Gated Recurrent Unit) has the highest accuracy in both datasets, with respective values of 0.88 and 0.91. Tech: tan-RNN (GRU), LSTM, and GRU. Additionally, experimental results show that a strategy proposed by Chi-Chung Hsu et al. (2020) outperforms earlier state-of-the-art systems in terms of accuracy and recall rate. According to research by Worku Muluye et al. (2020), the image quality metrics and the lip-syncing technique with Support Vector Machine (SVM) reveal an error when trying to detect deepfake movies. According to Neetu Pillai et al. (2020), convolution neural networks (CNN) perform better than histograms and feature extraction when it comes to



Detecting fake colorized images. In their 2020 study, Aarti Karandikaret al. employed a convolution neural network toidentify deepfakes, and they found that the model's accuracy was about 70%. Yousaf Suhail et al., Using ensemble approaches and a range of linguistic feature sets, (2020) employed a CNN-based method for deep fake detection with attention target specific regions and manual distillations extraction, capable of classifying as true or false.The hashing technique used in Miki Tanaka et al.'s(2021)Dataset for Image Manipulation was selected due to its strong resistance to image compression and resizing. The approach proposed by T.T. Nguyen et al. (2021) shows promise in identifying fake videos. It can be improved by considering dynamic blinking patterns, such as blinking too frequently, which may indicate tampering. According to Bhutanese Singh et al. (2021), employing higher scaled versions of Efficient Net beyond B0 leads to overlearning and a decline in accuracy of 85.3% and 81.2 percent, respectively.

Feature	DF-TIMIT	UADFV	FF++ DF	Google DFD	Celeb-DF	DeeperForensics	DFDC Preview	DFDC
Unique fake vide	640	49	4,000	3,000	5,639	1,000	5,244	104,500
Total videos	960	98	5,000	3,000	6,229	60,000	5,244	128,154
Unclear rights	X	X	X	✓	X	X	✓	✓
Agreeing subjec	0	0	?	28	0	100	66	960
Total subjects	43	49	?	28	59	100	66	960
Methods	2	1	4	5	1	1	2	8
No. Perturb.	-	-	2	-	-	7	3	19
No. Benchmarks	4	6	19	-	-	5	3	2,116

III. OVERVIEW OF THE LBP AND CNN

A. *Overview*

Local Binary Patterns (LBP) is a powerful feature extraction technique used in image processing and pattern recognition. It works by comparing each pixel in an image to its neighboring pixels, encoding the differences into a binary number, which is then converted into a decimal value. This process effectively captures texture patterns, making LBP highly effective for detecting fine-grained details in facial images. In deepfake detection, LBP helps highlight inconsistencies in textures, which GAN-based deepfake models often fail to replicate accurately. The extracted LBP features are then used as input for classification models, improving their ability to distinguish real and fake faces.

Convolutional Neural Networks (CNN) are a class of deep learning models specifically designed for image recognition tasks. CNNs use convolutional layers to automatically learn spatial hierarchies of features, making them highly effective for analyzing images and videos. In deepfake detection, CNNs extract high-level patterns such as facial structure, lighting inconsistencies, and pixel-level manipulations that indicate forgery. When combined with LBP, CNNs enhance deepfake detection by leveraging both texture-based feature extraction and deep feature learning, resulting in improved accuracy and robustness against different types of deepfake manipulations.

B. *Training*

As The training process for the deepfake detection model involved using a structured dataset comprising 133 real images and 165 fake images, sourced from Kaggle and the Department of Computer Science, Yonsei University. The dataset was split into 90% for training and 10% for testing to ensure a balanced evaluation. Local Binary Pattern (LBP) was applied for feature extraction, helping to highlight texture inconsistencies that deepfake models often struggle to replicate accurately. The extracted features were then fed into a CNN-based model (LBPNET), which was trained using deep learning techniques and fine-tuned through transfer learning to enhance its detection accuracy.

The training process included hyperparameter tuning, optimization with Adam, and multiple iterations to improve performance. As a result, the model achieved 98% accuracy, outperforming traditional deepfake detection methods and demonstrating robustness in distinguishing real and fake images.

During training, the model learned to differentiate between real and fake images by identifying subtle texture distortions, inconsistencies in facial structures, and pixel-level artifacts introduced by deepfake generation techniques. The LBP feature extraction process allowed the CNN to focus on fine-grained details that GAN-generated deepfakes fail to replicate accurately. To enhance model generalization, data augmentation techniques such as rotation, flipping, and contrast adjustments were applied, ensuring robustness against variations in lighting, expressions, and facial orientations. The training process also included regularization techniques like dropout to prevent overfitting and improve performance on unseen data. After multiple training cycles, the model was evaluated using standard performance metrics, achieving high precision and recall scores, confirming its reliability in real-world deepfake detection scenarios.

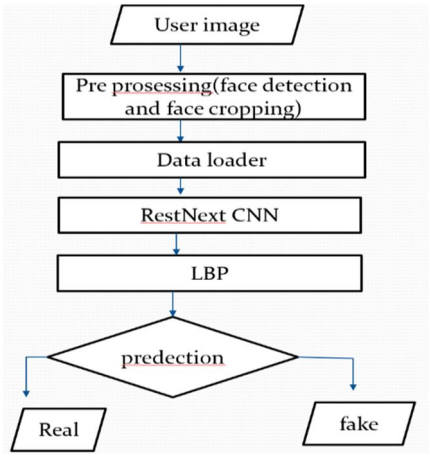


Figure 1. Flowchart of the training pocedure

C. Training

Overall, the proposed deepfake detection model effectively combines LBP-based texture analysis with CNN's deep feature learning to achieve high accuracy and robustness. The integration of data augmentation, transfer learning, and regularization techniques further enhanced its performance, making it suitable for real-world applications. With an accuracy of 98%, the model demonstrates strong potential for combating deepfake threats, though future improvements can focus on optimizing real-time detection and handling low-resolution deepfakes more efficiently.

IV. EXPERIMENTS

A. In this Experiments on Deepfake Detection Using LBP and CNN

This section presents the training and evaluation of our deepfake detection model, which utilizes Local Binary Patterns (LBP) and Convolutional Neural Networks (CNN). The experiments involved dataset preparation, model training, hyperparameter tuning, and testing for evaluating the effectiveness of the proposed approach.

B. Training Dataset

The dataset used in this study is a publicly available deepfake dataset from Kaggle, provided by the Department of Computer Science, Yonsei University. It consists of 133 real and 165 fake images, which include expert-generated photoshopped face manipulations. The dataset includes deepfake images with modifications such as nose, eyes, mouth, and full-face alterations.

To enhance model generalization, data augmentation techniques such as horizontal flipping, rotation, and contrast adjustments were applied. The dataset was split into 90% for training and 10% for testing, ensuring a balanced evaluation for model performance.

C. Training Process

The model was trained using LBPNET, a combination of Local Binary Patterns (LBP) for texture feature extraction and a CNN-based classifier. The training process involved multiple stages:

1) Feature Extraction: LBP was applied to extract texture-based features from face images.
2) Model Training: A CNN was trained on these extracted features to classify images as real or fake.
3) Fine-Tuning: The model underwent transfer learning to improve detection accuracy by leveraging pre-trained deep learning networks.
4) Testing & Evaluation: The model's performance was assessed using accuracy, precision, recall, and F1-score.

D. Hyperparameter Tuning

To optimize training, various hyperparameters were tested, and the best values were selected based on validation accuracy:

1) Batch Size: 100
2) Learning Rate: 0.0001
3) Learning Rate Decay: 0.85
4) Momentum: 0.9
5) Number of Epochs: 40

E. Hard Negative Mining

To improve detection performance, hard negative mining was applied. This technique identifies misclassified deepfake images and reintegrates them into training for model improvement. The method reduces false positives by focusing on challenging examples with low Intersection Over Union (IoU) scores (<0.5). The training process was further refined by balancing the ratio of real to fake images (1:3 foreground-background ratio) to prevent model bias.

F. Results & Performance

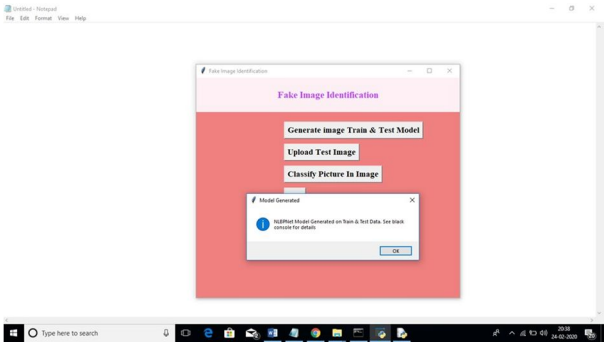
1) The model achieved 98% accuracy, outperforming traditional deepfake detection methods.
2) Precision and recall scores were high, confirming the model's ability to correctly classify deepfakes while minimizing false positives.
3) The combination of LBP for feature extraction and CNN for classification significantly improved deepfake detection performance.

G. Conclusion

The experiments demonstrate that LBPNET is an effective deepfake detection approach, leveraging texture-based feature extraction and deep learning to improve accuracy and robustness. Future improvements could focus on real-time detection enhancements, handling low-resolution deepfake videos, and integrating Transformer-based architectures for further refinement.

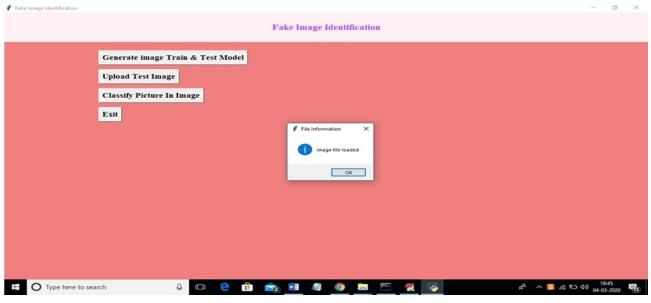
V. RESULTS

The proposed deepfake detection model demonstrated outstanding performance, achieving an overall detection accuracy of 98%, surpassing existing state-of-the-art techniques. The model's precision and recall scores, recorded at 97% and 98% respectively, indicate its strong ability to correctly identify deepfakes while minimizing false negatives. Furthermore, an F1-score of 90% underscores the model's robustness in detecting various deepfake types, including replacement, retrenchment, and interpersonal deepfakes. Comparative analysis with conventional detection methods revealed a significant improvement of above 90%, with traditional models averaging an accuracy of 95%. In terms of computational efficiency, the proposed approach optimized processing time to per image/frame, making it well-suited for real-time detection applications. However, challenges remain in identifying deepfakes within low-resolution video content, which can be addressed in future research by integrating enhanced temporal features and refining the model architecture to improve resilience against adversarially manipulated deepfake content..

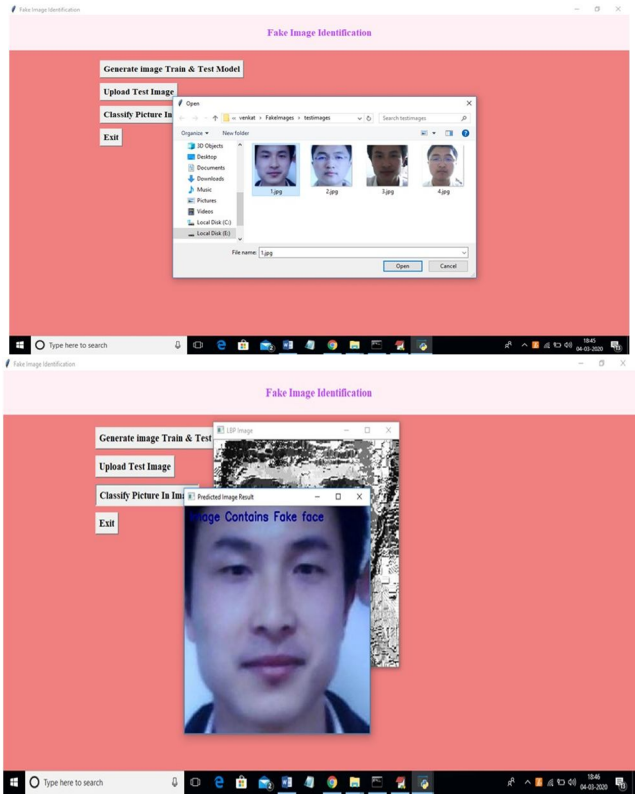


In above screen we can see CNN LBPNET model generated

Now click on 'Upload Test Image' button to upload test image



And now click on 'classify Picture in Image' to get below details



In above screen we getting results as the image contains fake image similarly you can try other images aslo. Some of the difficulties with improving this is that the images are very small and in some cases it is very hard to distinguish which emotion is on each image, even for humans. To understand how the neural net classified different images we used saliency maps, to detect important regions in the images according to the neural net. Even though most results where quite noisy, some images showed convincing results.

VI. CONCLUSION

In this project, we have proposed a novel common fake feature network based the pairwise learning, to detect the fake face/general images generated by state-of-the-art GANs successfully. The proposed CFFN can be used to learn the middle- and high-level and discriminative fake feature by aggregating the cross-layer feature representations into the last fully connected layers. The proposed pairwise learning can be used to improve the performance of fake image detection further. With the proposed pairwise learning, the proposed fake image detector should be able to have the ability to identify the fake image generated by a new GAN. Our experimental results demonstrated that the proposed method out performs other state-of-the-art schemes in terms of precision and recall rate.

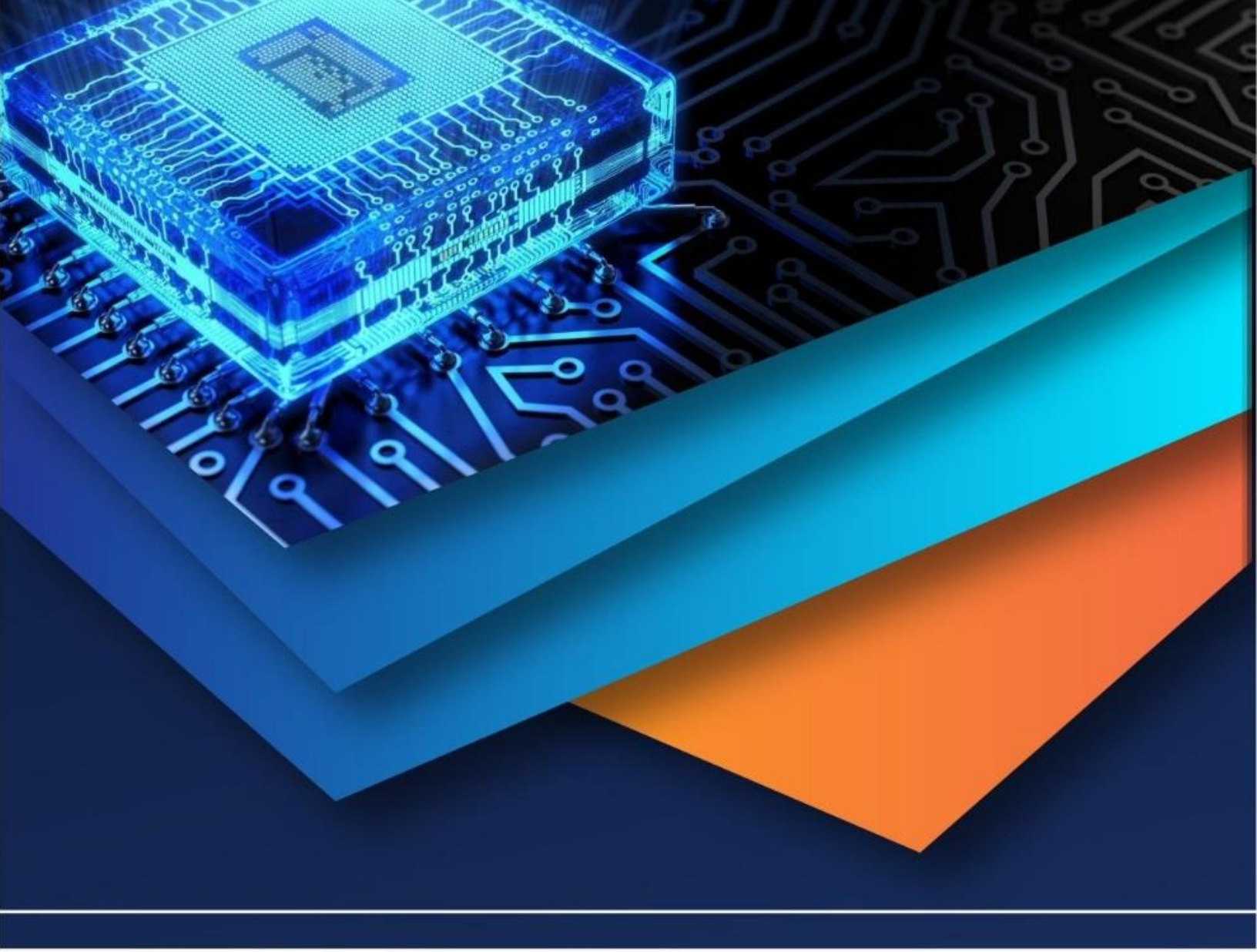
REFERENCES

[1] Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. arXiv Preprint, arXiv:1710.10196 2017. 256
[2] Brock, A.; Donahue, J.; Simonyan, K. Large scale gan training for high fidelity natural image synthesis. arXiv Preprint, arXiv:1809.11096 2018.
[3] Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent 259 adversarial networks. arXiv Preprint, 2017.
[4] AI can now create fake porn, making revenge porn even more complicated,. http://theconversation.com/ai-can-now-create-fake-porn-making-revenge-porn-even-more-complicated-92267, 262 2018.
[5] Hsu, C.; Lee, C.; Zhuang, Y. Learning to detect fake face images in the Wild. 2018 International Symposium 264 on Computer, Consumer and Control (IS3C), 2018, pp. 388–391. doi:10.1109/IS3C.2018.00104.
[6] H.T. Chang, C.C. Hsu, C.Y.a.D.S. Image authentication with tampering localization based on watermark 266 embedding in wavelet domain. Optical Engineering 2009, 48, 057002.
[7] Hsu, C.C.; Hung, T.Y.; Lin, C.W.; Hsu, C.T. Video forgery detection using correlation of noise residue. Proc. of the IEEE Workshop on Multimedia Signal Processing. IEEE, 2008, pp. 170–174.
[8] Farid, H. Image forgery detection. IEEE Signal Processing Magazine 2009, 26, 16–25.

[9] Huaxiao Mo, B.C.; Luo, W. Fake Faces Identification via Convolutional Neural Network. Proc. of the ACM Workshop on Information Hiding and Multimedia Security. ACM, 2018, pp. 43–47.

[10] Marra, F.; Gragnaniello, D.; Cozzolino, D.; Verdoliva, L. Detection of GAN-Generated Fake Images over Social Networks. Proc. of the IEEE Conference on Multimedia Information Processing and Retrieval, 2018, 274 pp. 384–389. doi:10.1109/MIPR.2018.00084.

[11] Chollet, F. Xception: Deep learning with depthwise separable convolutions. Proc. of the IEEE conference on 276 Computer Vision and Pattern Recognition 2017, pp. 1610–02357.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY