(Wavelet transform and Intrinsic signature based real and synthetic image classification)

Rumman Shaikh, Akshata Thorkar, Sanika Thakare, Atharv Mhatre, Vipin Yadav

Department of Electronics and Telecommunication Engineering,School of Electronics and Telecommunication Engineering, MIT Academy Of Engineering Alandi, Pune.

***Abstract*— The ”Deep Fake Detection using Intrinsic Signatures” is a method use for spotting fake images/videos by analyzing their basic and natural attributes. This method is use to aim classify manipulated content from genuine media without depending on external data. By identifying features within the video/image, this method improves the accuracy and quality of deep fake identification. This approach holds promise in addressing the increasing concern over misleading digital media and contributes to the development of more reliable tools for identifying manipulated content**

Intrinsic signature analysis looks for small clues and inconsistencies in deepfake images that give them away as fake. These clues are like unique patterns that differ from real images. Key terms in this area include deep learning, neural networks, and analysis like wavelet transform and azimuthal averaging.

The review also talks about the challenges researchers face, such as deepfake creators trying to outsmart detection methods and the need for big datasets to train accurate detection models. It touches on the ethical side too, emphasizing the responsible use of deepfake detection technology.

Keywords —Deepfake detection, intrinsic signature analysis, Wavelet transform, Azimuthal averaging, convolutional neural networks, artificial neural networks, machine learning,.

INTRODUCTION

With the rise of fake videos, it's becoming harder to tell what's real online. Common detection methods using outside clues are not as good anymore. The "Deep Fake Detection using Intrinsic Signatures" project wants to find a better way. Instead of looking at outside info, it focuses on the stuff inside the video itself. This project aims to create a new and strong method to spot fake videos, even when the usual clues don't work well. It's all about making sure we can trust what we see online and stop fake videos from fooling us.The "Deep Fake Detection using Intrinsic Signatures" project addresses the growing challenge posed by fake videos in the online landscape. As technology advances, distinguishing between authentic and manipulated content has become increasingly difficult. Traditional detection methods that rely on external clues are becoming less reliable. Hence, there is a critical need for innovative approaches that focus on the intrinsic characteristics of the videos themselves.

"Deep Fake Detection using Intrinsic Signatures" aims to face misleading digital content by focusing on intrinsic characteristics within media. This approach provides accurate and reliable detection techniques,maintaining trust in digital media and combating manipulated content spread.The reason, behind this effort is based on the need to maintain trust in media. As fake content becomes more prevalent its impact goes beyond spreading misinformation.

It poses a threat to society by eroding trust in visual information, which can create an atmosphere of doubt and skepticism. This project strives to be a measure that assures individuals they can consume content knowing that advanced detection methods are, in use to protect against deception.Despite advances in image fraud detection, identifying fake images remains a difficult task. Most current methods depend on deep learning techniques that require large amounts of labeled training data, which is often hard to obtain. This challenge is especially pronounced with fake faces, which can be used in various malicious contexts, from identity theft to misinformation campaigns

Our method demonstrates simple and interpretable features that can be effectively used for deepfake detection, offering a compelling alternative to complex deep learning models. By utilizing frequency domain characteristics such as wavelet transforms which converts image to 2D spectrum and converting these into 1D spectra through azimuthal averaging, the method achieves high classification accuracy.For classification we have used various machine learning algorithms.By focusing on frequency components, our approach provides a robust way to detect fake images without requiring large amounts of labeled data. The intrinsic signatures in the frequency domain offer a reliable means of identifying fake content, even when the images are highly realistic.

I. PROBLEM STATEMENT

The rise of tools like GANs has made it easy to create realistic fake images, especially fake faces, which are hard to detect. Current detection methods need a lot of labeled data and often fail to catch these sophisticated fakes. Our project aims to solve this problem by using a new approach that looks at the unique patterns in the image's frequency domain. By analyzing their frequency features with the Wavelet Transform, we can train a system to tell real faces from fake ones. This method promises better accuracy and works even with limited labeled data, helping to ensure the authenticity of digital images.

II. RELATED WORK

In this Paper[1],The technique described in this study uses high-frequency pattern analysis to identify artificial intelligence (AI)-generated phony facial photos. The method is incredibly effective; for high- and medium-resolution deepfakes produced by different GAN models, it achieves 100% accuracy. 91% accuracy is maintained by the approach even for low-resolution photos. The method is a potent tool for guaranteeing the authenticity of digital photos since it can quickly identify genuine faces from fakes by concentrating on the frequency domain of the images. This technique works especially well because it doesn't need a lot of labeled data, which makes it practical and efficient for use in real-world scenarios.

.

In this paper [2],The authors go over a number of techniques used in agriculture for precise nutrient predictions and crop selection, including SVM algorithms, photo processing, and bio-inspired frameworks. The importance of feature selection and classification in crop prediction is also covered in this paper, along with the application of deep learning techniques like RNN and LSTM for crop yield prediction. They present Boruta and RFE feature selection techniques. The study addresses the difficulties in analyzing agricultural data and highlights the use of devices and sensors for agricultural monitoring. The use of deep neural techniques for fake news detection, such as LSTM, is also mentioned by the authors.

In this paper[3].Experiments show that slight training variations, like initialization, produce distinct fingerprints across a GAN's images. These fingerprints enable precise image and model attribution, remain stable across frequencies and patch sizes, and resist typical GAN distortions. Despite vulnerability to perturbation attacks, fine-tuning effectively restores their reliability. Our fingerprints consistently surpass recent methods and Inception features in identifying images from different sources.

In this paper[4],Experiments show that even small training variations, like how a GAN is initialized, create unique fingerprints in the images it generates. These fingerprints are very useful for accurately identifying both the images and the models that created them. They remain consistent across different frequencies and image patch sizes and are not easily affected by common GAN distortions. While these fingerprints can be vulnerable to specific perturbation attacks, fine-tuning the model helps restore their reliability.

Our fingerprinting method consistently outperforms recent techniques and Inception features in identifying the source of images. This means that our approach can more accurately trace back which GAN model produced a particular image, making it a powerful tool for detecting deepfakes and attributing them to their original models. Overall, our method provides a robust and efficient solution for image and model attribution in the context of GAN-generated content

In this paper[5], they developed a method for detecting CG images using a ResNet-50 CNN and transfer learning. By extracting 2048-dimension bottleneck features, they tested training ResNet-50 from scratch and full transfer learning.method achieved over 97% accuracy, matching state-of-the-art results, and eliminating manual feature extraction. t-SNE improved class separability, but the SVM classifier's learning curve remained unstable with more data, suggesting room for further accuracy improvements. The method struggles with highly realistic CG images, as shown by tests with Holmes et al. and Carvalho et al.

In this paper[6],The document explains how to distinguish between real photos and computer-made ones by analyzing specific patterns in the images. It focuses on demosaicing, a process cameras use to determine colors in a picture. The authors have developed a method to detect areas in images that might have been altered by examining these color patterns. Their approach uses a special filter to highlight these patterns. By studying how colors vary across the picture and developing predictive models for nutrient management. The integration of data from different sources and the applying some mathematical analysis, they can identify these patterns.

Although their example focuses on green parts of the picture, they claim that the method works for other colors as well. The document also highlights the importance of demosaicing in photography and its impact on picture quality. Cameras guess the missing colors in a picture by using the colors nearby. Their method can detect changes in real photos taken by regular cameras, even if there are some complexities in the image.

Additionally, the document notes that computer-generated images are becoming increasingly realistic, particularly in movies, making it harder to differentiate them from real photos. The overall message is that their method provides a way to check if a picture has been altered, which is crucial in a world where fake images are becoming more common.

In this paper[7],The document discusses using capsule networks to detect fake images and videos. Capsule networks are a type of neural network designed to recognize and preserve spatial relationships within images. This makes them particularly good at identifying subtle features that distinguish real content from fakes. The method works by analyzing the structure and patterns in the images and videos, which helps to spot inconsistencies created during the generation of fake content. Compared to traditional neural networks, capsule networks can more accurately identify manipulated images and videos because they maintain the spatial hierarchies in the data. This approach is especially useful for detecting deepfakes, which are becoming increasingly realistic and harder to identify with conventional methods. Overall, using capsule networks provides a powerful tool for ensuring the authenticity of visual media in an era where fake content is rapidly increasing.

The paper [8] describes how to analyze convolutional traces in photos to identify deepfakes. The term "convolutional traces" describes the minute patterns and artifacts that deepfake images and videos produced using convolutional neural networks (CNNs) leave behind. The goal of this approach is to find these traces in order to distinguish between authentic and fraudulent content. The method can detect deepfakes by looking at the finer characteristics in the image, like irregularities in texture and pixel distribution.

The method is identifying the distinct convolutional signatures that the generative processes of deepfake models leave behind by employing sophisticated algorithms. When compared to conventional detection techniques, this enables a more accurate identification of modified images and movies. The resilience of the approach rests in its capacity to identify even expertly constructed deepfakes that could deceive

The paper [9] discusses the application of machine learning techniques in nutrient management in agriculture. It reviews various methods and approaches developed over the last decade for estimating fertilizer and nutrient status. The study explores detection and classification techniques, including machine learning algorithms, feature selection, and performance metrics such as accuracy, recall, precision, F-measure, and area-under-the-curve. Additionally, the paper highlights the use of technologies like soil spectroscopy, chemometrics, neural networks, and PLSR in developing predictive models for nutrient management. The integration of data from different sources and the importance of data quality and quantity are emphasized to enhance the accuracy and practical application of machine learning in agricultural nutrient management.

In order to develop practical answers, this paper [10] addresses the difficulties presented by Deepfake technology and the Deepfake Detection Challenge. It suggests a deep convolution GAN detection model as a remedy, which makes use of noise for data diversity and performs well even with little datasets. As the number of training iterations increases, the model continues to perform exceptionally well and consistently detects fakes. Accuracy is maximized for parameters such as epoch cycles, batch size, noise level, and model layers. The model's high criteria are confirmed by evaluation metrics such as Fréchet Inception Distance and Inception Score. Nevertheless, there are still difficulties in managing tiny datasets and resolving GAN restrictions like as mode collapse and convergence problems. In order to improve performance, future research should concentrate on strengthening these areas and expanding the generalization of GAN models.

III. METHODOLOGY

In the following section. we describe our approach for implementing classification of real and computer generated images.

1. Data Collection and Description:

There is no publicly available dataset exists at this time that offers high-resolution photos with labels identifying real and fake images. In order to remedy this, we have used photos from existing datasets to construct our own dataset, Faces-HQ2. To guarantee a broad array of faces, we used pictures from the 100K Faces project [1], the Flickr-Faces-HQ dataset [15], the CelebA-HQ dataset [14], and www.thispersondoesnotexist.com. We have collected 40,000 high-quality photos in total, divided equally between 20,000 real and 20,000 fake images

2. Applying Wavelet transform:

Input Images are converted to the 2D spectrum frequency domain for analysis using the wavelet transform. The wavelet transform yields details about the image's spatial localization and frequency content. The wavelet transform is highly effective in capturing abrupt changes in color and crisp edges in photographs.A wavelet transform condenses an image's energy into a small number of important coefficients.Wavelets of different sorts, such as (Haar, Daubechies, and Coiflets) can be selected according to the particular needs of the image analysis task.we have used Haar wavelet for the same.

3. Azimuthal Averaging:

azimuthally averaging is a method for doing a rotationally symmetric analysis of an image's frequency components. the wavelet transform output provides a powerful tool for simplifying and analyzing the frequency content of images It involves averaging the spectral data over concentric circles centered at the frequency domain's origin when applied to the output of a wavelet transform image. By converting the complex 2D frequency data into a straightforward 1D profile, this technique facilitates the analysis of certain features such as radial frequency content.

4. Model Development and Evaluation:

For classification of real and computer generated images , a number of machine learning algorithms were assessed, including Artificial Neural Networks (ANN),Convolutional Neural Networks(CNN),Random Forest,Support Vector Machines (SVM),Logistic regression(LR).These algorithms were selected for accurate and reliable classification.

Artificial Neural Networks (ANN): This algorithm used for ability to learn complex patterns and features from data In this algorithm,data is preprocessed to match input neurons with extracted features in order to use an Artificial Neural Network (ANN) to identify real and computer-generated images.ReLU-activated dense layers & softmax/sigmoid output layer for binary classification are created then the Adam/SGD optimizer, binary cross-entropy,recall, accuracy, and precision are employed. To avoid overfitting, divide the data into train, validation, and test sets. Then, specify the batch sizes and epochs.

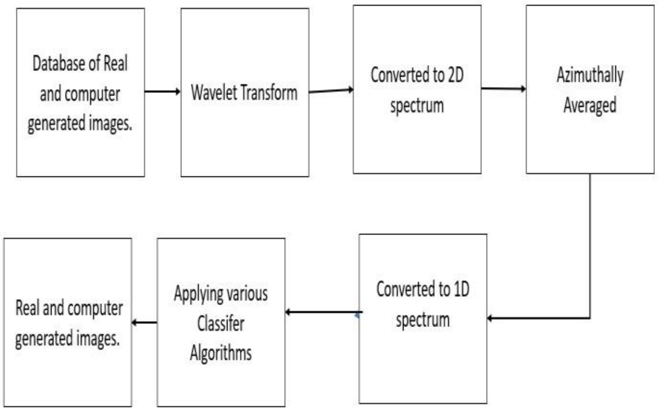
Convolutional Neural Networks(CNN): CNNs provide robust and accurate results, making them indispensable in image classification. An input layer have been preprocessed is where the CNN begins. ReLU activation is used by sequential convolutional layers to provide non-linearity while applying filters to extract textures and edges. Max-pooling layers use feature map downsampling to reduce complexity. The next layer is dense completely connected, which combines features for categorization.for classification of real and computer generated images, it uses sigmoid. The usual loss function, binary cross-entropy, is optimized using SGD or Adam to reduce mistakes.

Random forest: Random Forests use ensemble learning and decision trees to provide a reliable and efficient method for classifying both real and computer-generated images. Random Forests are ensembles of decision trees trained independently on subsets of both data and features through Bootstrap Aggregation (Bagging). This method ensures diversity among trees, enhancing robustness. During training, each node randomly selects features to split, reducing overfitting. Each tree predicts independently, and final classification uses majority voting across all trees, leveraging collective insights for accurate classification of real and computer-generated images.

Support Vector Machines(SVM): SVMs classify real and fake images by extracting critical features like frequency which are then used to train the model with various kernel functions selected based on data complexity. Cross-validation optimizes hyperparameters such as the penalty parameter (C) and kernel settings. SVMs seek to find the best hyperplane that maximally separates image classes. Their significance in digital media forensics lies in their capability to manage complex datasets through meticulous parameter tuning and feature extraction, ensuring robust classification.

Logistic regression: It starts by extracting features from images—converting them into numerical vectors encompassing features. These features, along with labeled data, form the training dataset, which is split into training and validation sets. Logistic regression models the likelihood of binary outcomes using these features, optimizing weights via methods like gradient descent or maximum likelihood estimation. After training, the model predicts whetherimages are artificial or real based on a predefined crite

IV. BLOCK DIAGRAM


Database of Real and computer generated images.
Wavelet Transform
Converted to 2D spectrum
Azimuthally Averaged
Converted to 1D spectrum
Applying various Classifer Algorithms
Real and computer generated images.

V. FUTURE SCOPE

Due to the advancement in technology and era many of the crimes are coming under these Deepfakes resulting in degradation of ones image and personality in public.

The future of deepfake detection, particularly in cyber cells, plays significant potential for rapid and accurate image and video analysis. As deepfake technology advances, so must the tools used to tackle it. Coming era Cyber cells will increasingly rely on advanced AI and machine learning algorithms for real-time detection of deepfakes. Improved techniques, such as intrinsic signature analysis with wavelet transformations, can provide more strong defenses by quickly identifying hazardous manipulations in media.

The combination of these technologies into cybersecurity frameworks will enhance the ability to monitor and protect digital platforms. Automated systems capable of swift detection and alerting can help reduce the spread of fake media, protecting individuals and organizations from misinformation and fraud. Future research will likely focus on refining these methods, reducing computational costs, and increasing detection accuracy, making deepfake detection an integral part of cyber defense strategies.

VI. RESULTS AND OBSERVATIONS

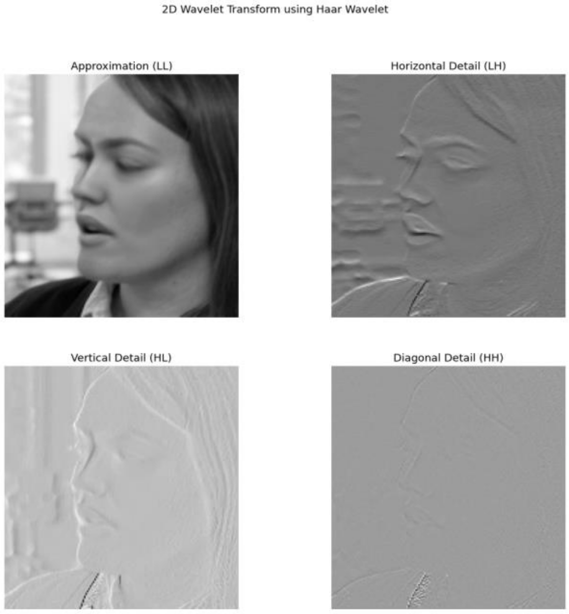


Figure 1. 2D Wavelet Transform using Haar Wavelet

After applying Haar Wavelet Transform on input images, output is obtained as shown in Figure 1.It tells about approximate,horizontal,vertical and diagonal details of input images.Wavelet tranforms input image to 2D spectrum.

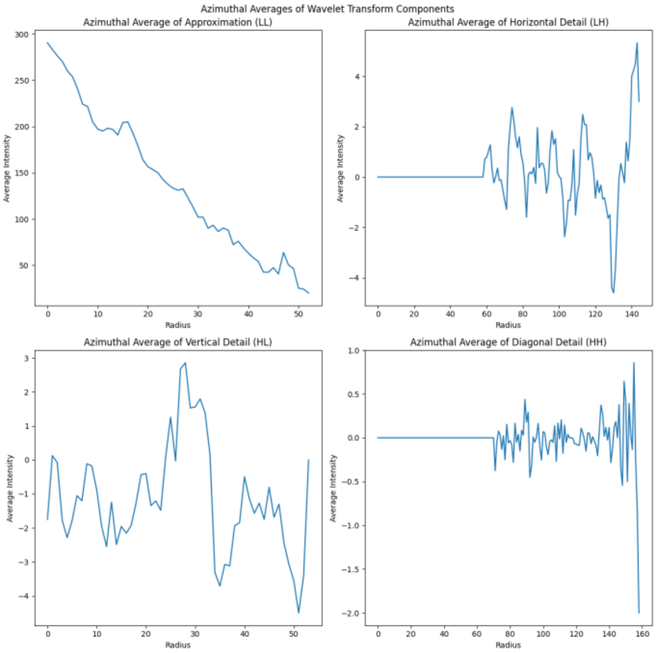


Figure 2. Azimuthal averaging of fake image

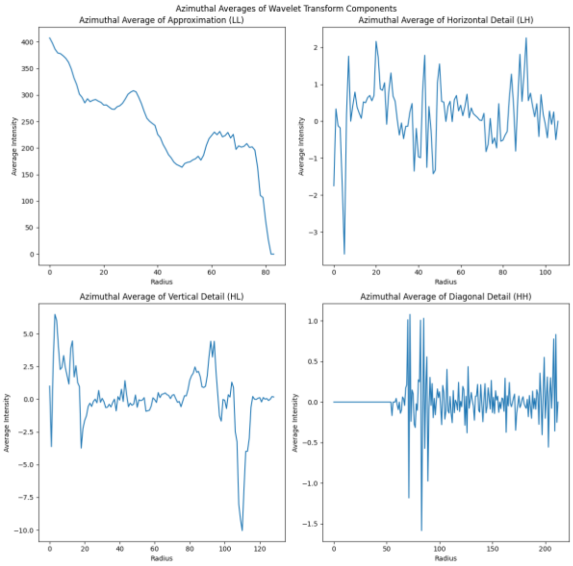
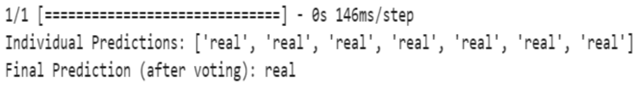


Figure 3. Azimuthal averaging of Real image

On the output of wavelet transform,we have done azimuthal averaging which converts 2D spectrum into 1D spectrum.Figure 2 shows azimuthal averaging of fake image and figure 3 shows azimuthal averaging of real image.It’s a method for doing a rotationally symmetric analysis of an image's frequency components.

After azimuthal averaging we have applied various classifier algorithms on these 1D spectra, achieving different levels of accuracy: Support Vector Machine (SVM) with 92% accuracy, Logistic Regression (LR) with 83% accuracy, Random Forest with 93% accuracy,K Nearest Neighbour with 93.03% accuracy,Artificial neural networks(ANN) with 92% accuaracy This approach demonstrates the potential of wavelet transform-based feature extraction combined with standard classifiers to effectively differentiate between real and computer-generated images.



```
1/1 [==============================] - 0s 146ms/step
Individual Predictions: ['real', 'real', 'real', 'real', 'real', 'real', 'real']
Final Prediction (after voting): real
```

Figure 4.Output after performing Voting

We have also performed voting to predict using all machine learning models that we have implemented Figure 4 shows output after performing Voting

VII. CONCLUSION

The study successfully demonstrates that simple and interpretable features can be effectively used for deepfake detection, offering a compelling alternative to complex deep learning models. By utilizing basic image statistics and frequency domain characteristics, this approach leverages techniques such as wavelet transforms, which convert images into 2D spectra. These 2D spectra are then transformed into 1D spectra through a process known as azimuthal averaging, where the image data is averaged along circular paths centered at the image's origin. This innovative method captures essential patterns and irregularities often found in deepfake images.

The effectiveness of these features is validated through the implementation of several classification algorithms. For instance, a Support Vector Machine (SVM) achieves a high classification accuracy of 92%, indicating its ability to create a robust boundary between real and fake images based on the derived features. Logistic Regression (LR), another classifier, reaches an accuracy of 83%, demonstrating that even simpler linear models can discern subtle differences between real and computer-generated images when fed with well-chosen features. Random Forest, known for its ensemble learning capability, attains a 93% accuracy, showcasing its strength in handling varied data distributions and providing a reliable classification through multiple decision trees.

Moreover, the K Nearest Neighbour (KNN) algorithm also matches this high performance with a 93.03% accuracy, highlighting its effectiveness in instance-based learning where it compares new data points directly with stored instances. Artificial Neural Networks (ANNs) also show a commendable performance with a 92% accuracy, benefiting from their ability to capture complex non-linear relationships in the data despite being a simpler form of neural networks compared to deep learning models.

This approach not only proves to be computationally efficient but also robust against common post-processing techniques such as compression and noise addition, which are often used to evade detection. The simplicity of the features used ensures that the model remains interpretable, making it easier for analysts to understand and trust the detection process. Additionally, this method's low computational requirements make it highly suitable for real-time deepfake detection in resource-constrained environments, such as on mobile devices or in scenarios with limited processing power.

VIII. References

1. Durall, Ricard, Margret Keuper, Franz-Josef Pfreundt, and Janis Keuper. "Unmasking deepfakes with simple features." *arXiv preprint arXiv:1911.00686* (2019).

2. Tariang, Diangarti Bhalang, Prithviraj Senguptab, Aniket Roy, Rajat Subhra Chakraborty, and Ruchira Naskar. "Classification of Computer Generated and Natural Images based on Efficient Deep Convolutional Recurrent Attention Model." In *CVPR workshops*, pp. 146-152. 2019..

3. Yu, Ning, Larry S. Davis, and Mario Fritz. "Attributing fake images to gans: Learning and analyzing gan fingerprints." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7556-7566. 2019..

4. Wang, Sheng-Yu, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A. Efros. "CNN-generated images are surprisingly easy to spot... for now." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8695-8704. 2020.

5. De Rezende, Edmar RS, Guilherme CS Ruppert, Antonio Theophilo, Eric K. Tokuda, and Tiago Carvalho. "Exposing computer generated images by using deep convolutional neural networks." *Signal Processing: Image Communication* 66 (2018): 113-126..

6. Malik, Asad, Minoru Kuribayashi, Sani M. Abdullahi, and Ahmad Neyaz Khan. "DeepFake detection for human face images and videos: A survey." *Ieee Access* 10 (2022): 18757-18775.

7. Guarnera, Luca, Oliver Giudice, and Sebastiano Battiato. "Deepfake detection by analyzing convolutional traces." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 666-667. 2020.

8. Guo, Ke, and Rangding Wang. "A method for identifying computer images and real images." In *2011 International Conference on Electronics, Communications and Control (ICECC)*, pp. 638-641. IEEE, 2011.

9. Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Use of a capsule network to detect fake images and videos." *arXiv preprint arXiv:1910.12467* (2019)..

10. Ni, Xuan, Linqiang Chen, Lifeng Yuan, Guohua Wu, and Ye Yao. "An evaluation of deep learning-based computer generated image detection approaches." *IEEE Access* 7 (2019): 130830-130840..

11. Dirik, Ahmet Emir, Sevinç Bayram, Husrev T. Sencar, and Nasir Memon. "New features to identify computer generated images." In *2007 IEEE International Conference on Image Processing*, vol. 4, pp. IV-433. IEEE, 2007.

12. Sychandran, C. S., and R. Shreelekshmi. "Performance Comparison of Deep Learning Models for Computer Generated Image Detection." In *2023 International Conference on Control, Communication and Computing (ICCC)*, pp. 1-5. IEEE, 2023.

13. AlShariah, Njood Mohammed, A. Khader, and J. Saudagar. "Detecting fake images on social media using machine learning." *International Journal of Advanced Computer Science and Applications* 10, no. 12 (2019): 170-176.

14. Kumar, Manoj, and Hitesh Kumar Sharma. "A GAN-based model of deepfake detection in social media." *Procedia Computer Science* 218 (2023): 2153-2162..

15. Swaminathan, Ashwin, Min Wu, and KJ Ray Liu. "Digital image forensics via intrinsic fingerprints." *IEEE transactions on information forensics and security* 3, no. 1 (2008): 101-117.

16. Gallagher, Andrew C., and Tsuhan Chen. "Image authentication by detecting traces of demosaicing." In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1-8. IEEE, 2008..