International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Synthetic Image Identification using Deep Learning: A Review

Vaibhav Pawar[1], Pratiksha Pawar[2], Aditya Dawghat[3], Shreyash Somvanshi[4], Rajashree Ghule[5]

[1]*Department of Artificial Intelligence and Data Science, VPKBIET, Baramati*
[2]*Department of Artificial Intelligence and Data Science, VPKBIET, Baramati*
[3]*Department of Artificial Intelligence and Data Science, VPKBIET, Baramati*
[4]*Department of Artificial Intelligence and Data Science, VPKBIET, Baramati*
[5]*Department of Artificial Intelligence and Data Science, VPKBIET, Baramati*

Abstract - Recent developments in Artificial Intelligence (AI) have made it possible to create images of such high quality that people are unable to distinguish them from photographs taken in the real world. Given the vital importance of data authenticity and reliability, our system suggests a way to improve our capacity for computer-based AI image recognition. It is simple to imagine scenarios in which AI-generated images are exploited to create political unrest, fabricate acts of terrorism, and blackmail individuals. In this work, we have described a new deep learning-based method that can effectively distinguish AI-generated images from real images. Our method is capable of automatically detecting the authenticity of images present. Our system uses VGG-19, ResNet-50 and EfficientNet-B0 as the extractor of basic representation. We test our approach on a sizable dataset in order to replicate real-time events and improve the model's performance on real-time data.

Key Words: Synthetic Images, A.I. generated, Res-Net, Convolution Neural Network, CIFAKE, Explainable A.I.

1. INTRODUCTION

Our focus is on developing an image classification model to distinguish between real images and those generated by A.I. algorithms. It can address the rising concerns surrounding the authenticity of digital content and specify misinformation. By using advanced machine learning techniques, we aim to contribute to the image forensics, ensuring the credibility and trustworthiness of visual media in the era of AI-generated content.

Artificial intelligence (AI) is used in machine learning, which gives computers the capacity to autonomously learn from experience and get better without explicit programming. The creation of computer programs that can access data and utilize it to learn for themselves is the main goal of machine learning. In order to find patterns in data and use the examples we give to make better judgments in the future, learning starts with data, such as examples, first hand experience, or instruction. The main goal is to enable requirements.

The creation of Generative Adversarial Networks (GAN) has allowed computers to produce realistic facial images that can readily fool the public. These produced false faces will inevitably present major societal problems, such as false information and security risks. Therefore, it would be ideal to have strong methods for identifying these false faces. Nevertheless, unlike the extensive research in GANs, our challenge of fake face detection is still not well understood, and our understanding of generated faces is somewhat superficial. In real-world situations, faces come from various unidentified sources, such as distinct GANs, and they could have unidentified picture distortions, such as JPEG compression, noise, down sampling, and blur, which increases the difficulty of this task.

2. Synthetic Image Identification

A. Convolutional Neural Network

In the context of synthetic image classification, Convolutional Neural Networks (CNNs) play a pivotal role, offering a robust framework for extracting intricate features from complex datasets. The process initiates with the curation of a labeled dataset, comprising synthetically generated images, laying the groundwork for CNN training.

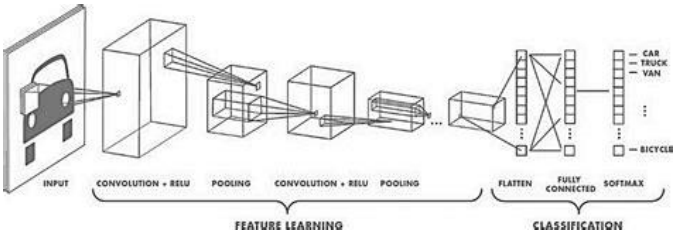

Fig -1: CNN Architecture

The architecture of the CNN is meticulously designed, incorporating convolutional layers for feature extraction, activation functions like Rectified Linear Unit (ReLU) for non-linearity, and pooling layers for dimensionality reduction. The model's compilation involves selecting a suitable loss function, such as categorical cross entropy, and an optimizer, often the adaptive moment estimation (Adam) algorithm. During the training phase, the dataset is divided into training and validation sets, with hyperparameters fine-tuned iteratively to optimize convergence. Post-training, the CNN undergoes evaluation on a separate test set to gauge its ability to generalize to unseen data. This iterative process allows for refinement, and if necessary,

fine-tuning of hyperparameters and architecture for optimal performance. The culmination of this journey sees the deployment of the trained CNN, integrated into applications to autonomously classify synthetically generated images with a high degree of accuracy.

B. Texture Extraction Network

In the realm of texture extraction, a dedicated Texture Extraction Network takes center stage. Commencing with the assembly of a meticulously curated dataset, rich in diverse textures and labeled for supervised learning, this network is designed to unravel and characterize intricate patterns within images. Leveraging convolutional layers, specifically tailored filters are deployed to capture nuanced textural elements, offering local detection across varying scales and orientations. Non-linear activation functions, such as Rectified Linear Unit (ReLU), introduce complexity to the model, allowing it to capture intricate relationships inherent in diverse textures. Pooling layers, typically in the form of max pooling, contribute to spatial dimension reduction, ensuring computational efficiency while retaining crucial textural information. Fully connected layers may be incorporated for complex texture classification tasks, synthesizing learned features for high-level predictions.

C. Explainable A.I.

The methodology employed by XAI is notably model-agnostic, allowing its principles to be universally applied across diverse machine learning architectures. This adaptability ensures that the quest for transparency is not confined to specific model types but can be embraced across the AI spectrum. The interpretability provided by XAI is both local and global. On a local scale, it elucidates the rationale behind specific predictions for individual instances, providing users with insights into the model's decision-making for specific cases. On a global scale, it unveils overarching patterns within the entire model, enabling a holistic comprehension of its behavior.

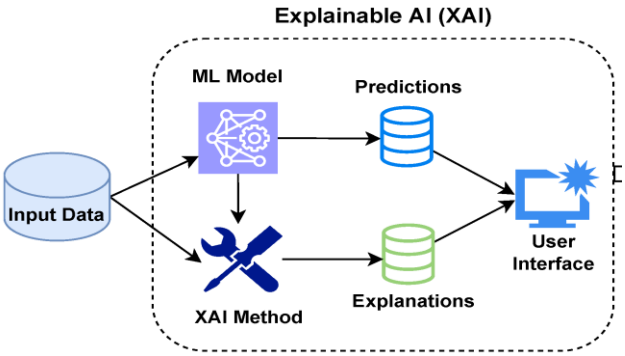

Fig -1: Explainable AI Architecture

3. Literature Survey

[1] The paper titled Robust Texture-Aware Computer-Generated Image Forensic: Benchmark and Algorithm, by author Weiming Bai et al was published in the year 2021. It focuses on the discrimination between CG and photographic (PG) images. This study describes the new benchmark called Large Scale CG images Benchmark (LSCGB), which contains a large number of CG and PG images with expert-annotated labels. The benchmark has high diversity and small bias between CG and PG images. Additionally, the authors propose a texture-aware network that extracts texture information from multiple levels of features and captures the relations among feature channels using a gram matrix. The proposed algorithm utilizes CNNs as the backbone network for feature extraction. Different backbone networks, such as VGG-19, ResNet-50, and EfficientNet-B0, are evaluated to assess the effectiveness of the texture-aware approach. Additionally, it is mentioned that the proposed method achieves an accuracy score of 99.94% on the Rahmouni benchmark, which is part of the LSCGB.

[2] A paper titled CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images by author Jordan J. Bird, and Ahmad Lotfi, was published in the year 2023. It discusses the use of diffusion models for generating high-fidelity images and the potential for these models to generate audio as well. The survey also mentions the importance of data authenticity and trustworthiness, as AI-generated images can be used for misinformation, fake news, and cybersecurity attacks. The CNNs are used as feature extractors to recognize the authenticity of the images. Additionally, Explainable AI (XAI) is employed using Gradient Class Activation Mapping (Grad-CAM) to interpret the model predictions and provide visual cues for the authenticity of the images.

[3] Another paper delving into the realm of image classification titled Deep Learning Based Method to Discriminate Between Photorealistic Computer Generated Images and Photographic Image, by author Kunj Bihari Meena, and Vipin Tyagi was published in the year 2020. It proposes a deep learning-based technique to differentiate between photorealistic computer-generated (CG) images and photographic images. The technique utilizes transfer learning, where the weights of a pre-trained deep convolutional neural network (DenseNet-201) are transferred to train a support vector machine (SVM) for classification. Experimental results on the DSTok dataset show that the proposed technique outperforms existing methods, achieving a detection accuracy of 94.12%. Overall, the proposed technique offers a promising solution to the challenge of distinguishing between CG and photographic images in the context of multimedia tools and digital forensics.

[4] The paper titled "BDC-GAN: Bidirectional Conversion Between Computer-Generated and Natural Facial Images for Anti-Forensics" by authors Fei Peng et.al was published in the year 2020, introduces a novel method for discriminating computer-generated (CG) and natural facial images using a bidirectional generative adversarial network (GAN). The proposed BDC-GAN synthesizes the noise of one domain with the content of another domain to generate facial images that deceive existing forensic methods. It highlights the limitations of existing unidirectional CG facial image anti-forensic methods and bidirectional domain adaptation methods, and demonstrates that BDC-GAN achieves better visual quality and

deception ability.

[5] The paper titled Distinguishing Computer-Generated Images from Natural Images Using Channel and Pixel Correlation by authors Zhang RS et. al, was published in the year 2020. It proposes a convolutional neural network (CNN) model called ScNet (Self-coding Network) to distinguish computer generated (CG) images from natural images (NIs). The model utilizes the correlation between color channels and pixels to extract features that differentiate CG images from NIs. The paper validates the robustness of the model against post processing techniques, such as JPEG compression, and evaluates its generalization capability using a challenging dataset. The authors also introduce a self-coding module in the network to explicitly extract correlation information between color channels, further improving the discrimination capacity.

[6] The paper titled Incremental learning for the detection and classification of GAN-generated images by Francesco Marra et al.in 2019 recognized the limitations of existing methods and proposed an innovative solution grounded in incremental learning principles. This incremental learning approach allows the model to adapt to the dynamic landscape of GAN-generated images, continuously updating its knowledge without forgetting previously learned information. While it provides a significant step forward in addressing the challenges associated with GAN-generated content detection, the paper also raises important questions about the adaptability of such methods to entirely new GAN architectures.

[7] The paper titled Detecting GAN generated fake images using co-occurrence matrices by authors Lakshmanan Nataraj et. al, was published in the year 2019. It leverages classical steganalysis techniques, particularly co-occurrence matrices, to identify deviations from natural image statistics induced by GANs. Instead of using handcrafted features, they directly pass co-occurrence matrices through a deep convolutional neural network (CNN), allowing the network to learn essential features. The proposed method is applied to GAN datasets, including CycleGAN and StarGAN, achieving promising results with over 99% classification accuracy in both datasets. The approach proves promising in diverse and challenging datasets, highlighting its potential significance in the field of digital image forensics.

[8] The paper titled Distinguishing between natural and computer-generated images using convolutional neural networks by author Weize Quan et al. was published in the year 2018. It proposed a convolutional neural network (CNN) framework for effectively distinguishing between natural images (NIs) and computer-generated (CG) images in image forensics, showcasing superior performance and robustness in challenging scenarios with heterogeneous image origins. The direct use of co-occurrence matrices and their integration with deep learning in the proposed method offers a novel and efficient strategy for the detection of GAN-generated fake images.

[9] The paper titled Exposing computer-generated images by using deep convolutional neural networks by authors Edmar R.S. de Rezende et al. (2018) presents a novel method for detecting computer-generated (CG) images utilizing deep convolutional neural networks (CNNs). The approach is thoroughly evaluated on diverse datasets, demonstrating robustness against various image processing operations, including JPEG compression, noise addition, and blur. The study contributes significantly to the literature on CG image detection, providing valuable insights into the effectiveness of deep learning methodologies in addressing challenges associated with synthesized images.

[10] The paper titled Computer graphics identification combining convolutional and recurrent neural networks by authors Peisong He et.al. (2018) introduces a deep learning-based pipeline for distinguishing between photo-graphics (PG) and computer-generated graphics (CG). The method combines convolution al neural networks (CNNs) and recurrent neural networks (RNNs), incorporating preprocessing steps such as color space transformation and the Schmid filter bank. The dual-path CNN architecture is designed to extract joint feature representations of local patches, considering both color and texture characteristics. The directed acyclic graph RNN is then applied to model spatial dependencies and extract global artifacts. The proposed framework achieves better identification ability for CGs, especially in cases of low-resolution images.

4. Survey Comparison

Table -1: Survey of relevant research papers

Author	Techniques	Gaps
Weiming Bai, et al. 2021 [1]	VGG-19, Texture Extraction Network	More Diverse dataset is required, Accuracy could be increased.
J.J Bird, Ahmad Lotfi, et al. 2023 [2]	CNN, Computer Vision, Explainable AI	Can't classify for Human faces, Accuracy can be increased.
Kunj Bihari Meena et al. 2022 [3]	DenseNet-201 Network, Transfer Learning, SVM Classifier	Modification is required when images are treated by operations like noise addition, image blurring, and contrast enhancement.
Fei Peng et.al. 2020 [4]	BDC-GAN	It generally reduces the quality of generated images while preserving the anti-forensic performance.
Zhang RS et. al. 2020 [5]	CNN, Channel and Pixel Correlation	Can't apply Pixel and Channel Correlation to other tasks only works for images generated by specified approach.
Francesco Marra et.al. 2019 [6]	Incremental learning, Representation Learning	Can not detect images generated by new GANs

Lakshmanan Nataraj et.al.2019 [7]	Deep CNN, Co-occurrence matrices	Can't localize the manipulated pixels in GAN generated images
Weize Quan et.al. 2018 [8]	BDC-GAN	Need to continue work on CNNs for image forensics.
Edmar R.S. de Rezende et. al. 2018 [9]	Deep CNN, ResNet50	Focus on different architectures for features extraction
Peisong He et.al.2018 [10]	CNN and RNN	Need to extend framework for GAN images

5. CONCLUSION

In summary, an overview of the latest developments in generative A.I. unveils a paradigm shift in the field of image classification. CNN has transformed the domain of image classification and analysis. Leveraging texture differences, a new network outperforms existing methods, indicating broader potential. The system's speed and efficiency in image classification provide a practical solution for processing large volumes of visual data rapidly and accurately. Extensive testing, including usability evaluations and security assessments, has confirmed the system's user-friendliness and resilience against potential threats. The system's capabilities and performance metrics make it suitable for real-world deployment in scenarios requiring image classification and verification. Integrating this classification model with Explainable A.I.(XAI) will increase model interpretability and can give us more understanding of its working. In summary, our project has not only addressed a significant challenge in image classification but has also paved the way for practical applications of AI-generated and real image differentiation.

REFERENCES

1. Weiming Bai et al. (2021). "Robust Texture-Aware Computer Generated Image Forensic: Benchmark and Algorithm". In: IEEE Transactions on Image Processing 30, pp. 8439–8453. doi:10.1109/TIP.2021.3114989.
2. Jordan J. Bird, and Ahmad Lotfi (2023). CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. arXiv: 2303.14126 [cs.CV]
3. De Rezende, Edmar RS et al. (2018). "Exposing compute rgenerated images by using deep convolutional neural networks". In: Signal Processing: Image Communication 66, pp. 113–126.
4. .Peisong He, et al. (2018). "Computer graphics identification combining convolutional and recurrent neural networks". In: IEEE Signal Processing Letters 25.9, pp. 1369–1373.
5. Francesco Marra, et al. (2019). Incremental learning for the detection and classification of GAN-generated images. arXiv: 1910.01568 [cs.CV].
6. Kunj Bihari Meena, and Vipin Tyagi (July 2020). "A Deep Learning Based Method to Discriminate Between Photorealistic Computer-Generated Images and Photographic Images". In: pp. 212–223. ISBN: 978-981-15-6633-2. doi: 10.1007/978-981-15-6634-9 20.
7. Lakshmanan Nataraj, et al. (2019). "Detecting GAN generated fake images using co-occurrence matrices". In: arXiv preprint arXiv:1903.06836.
8. Quan, Weize, K. Wang, Dong-Ming Yan and Xiao peng Zhang. "Distinguishing Between Natural and Computer-Generated Images Using Convolutional Neural Networks." IEEE Transactions on Information
9. Edmar R.S. de Rezende, Guilherme C.S. Ruppert, et al. "Exposing computer generated images by using deep convolutional neural networks", in Signal Processing: Image Communication,vol.66,2018,https://doi.org/10.1016/j.image.2018.04.006.
10. Ed P. He, X. Jiang, T. Sun and H. Li, "Computer Graphics Identification Combining Convolutional and Recurrent Neural Networks," in IEEE Signal Processing Letters, vol. 25, no. 9, pp. 1369-1373, Sept. 2018, doi: 10.1109/LSP.2018.2855566.