IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13346

Detection Of AI Generated Images Using Machine Learning and Deep Learning Models

Akshatha Nayak, Harsha, Prajeet Chendekar, Shreevatsan A, Sunil Kumar S*

Dept. of Artificial Intelligence & Machine Learning, Mangalore Institute of Technology & Engineering, Moodbidri, Karnataka, India

*Corresponding Author

Abstract: Artificial intelligence (AI)-generated images intended to incite social and economic unrest have become more widely shared since the introduction of advanced AI tools. AI-generated images using Generative Adversarial Networks (GANs) are frequently used to create content that makes it difficult to discern between real and artificial content. As a result, false information is spread along with an increase in cybercrimes. The goal of this proposed work is to detect these AI-generated images by building a Convolutional Neural Network (CNN) model. This CNN model will be trained to distinguish between real and AI-generated images. This strategy will support the preservation of social and economic stability, which may be jeopardized by improper use of images produced by artificial intelligence in informational campaigns. It will also aid in the prevention of cybercrimes like image forgery and impersonation that are caused by AI-generated images.

Keywords: Generative Adversarial Networks (GANs), Convolutional Neural Network (CNN), AI-Generated Images

I. INTRODUCTION

AI images generated using advanced Generative Adversarial networks (GANs) are a potential problem in today's digital environment. It is difficult to discern what is genuine and what is not because of these images. False and misleading information is disseminated due to the ease with which these AI generated images are created. This creates difficulties in determining what to believe thus leading to social, political, religious, and financial concerns.

The primary objective of the proposed work is to address the proliferation of false AI-generated images by focusing solely on the detection of such content using a specialized Deep Learning model.
By leveraging Convolutional Neural Networks (CNN), the proposed approach aims to effectively identify AI-generated images designed to disrupt societal harmony. By narrowing the scope to a single model, we streamline the detection process, enabling more efficient and accurate identification of deceptive content.

Through the exclusive utilization of CNN, the proposed initiative seeks to develop a robust framework capable of differentiating between genuine and AI-generated images. By eliminating the complexity introduced by multiple models, we enhance the reliability and effectiveness of the detection mechanism.

When utilized improperly, the rise in AI-generated content can be a serious danger to dependability and confidence in the digital sphere. By using the proposed initiative, which can differentiate between real and artificial intelligence-generated content, we help to maintain social peace, financial stability and the credibility of information.

II. LITERATURE REVIEW

In [11] deep learning-based Computer Generated Face Identification model, employing a customized Convolutional Neural Network (CNN) architecture. Utilizing PGGAN and BEGAN models for generating deepfake images from CelebA dataset, they tackled imbalanced data concerns through an Imbalanced Framework (IF-CGFace) by training AdaBoost and eXtreme Gradient Boosting (XGB) with features extracted from CGFace layers. A notable work by Deng Pan; Lixian Sun; Rui Wang[8] , the researchers employed Xception and MobileNet, leveraging their capabilities for classification tasks to automatically detect deepfake videos. They utilized training and evaluation datasets from FaceForensics++, encompassing four datasets generated with distinct deepfake technologies. The research achieved highaccuracy, ranging from 91% to 98%, across all datasets.

DOI: 10.17148/IJARCCE.2024.13346

A notable study by Weize Quan , Kai Wang, Dong-Ming Yan[6] the successful application of CNNs in various multimedia security tasks, we adopted a 3D filter group in the initial layer of our network to accommodate RGB images effectively. Unlike previous methods with fixed or constrained layers, our approach allows for flexible training of the 3D

convolutional filters, enabling automatic extraction of discriminative features tailored to our classification task. Furthermore, we conducted a comparative analysis with a recent parallel work, StatsNet, focusing on network design, architecture, and forensic performance, particularly on challenging datasets like Columbia.

In [12] It focuses on identifying artifacts created during the generation of DF videos using Generative Adversarial Networks (GANs). By detecting inconsistencies in facial features warped to fit the source's face, and comparing them with their surroundings, the method can identify DF artifacts. Temporal discrepancies between frames are captured using Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM), which helps in recognizing anomalies introduced during GAN reconstruction. The approach is streamlined by training a ResNet CNN model directly on the resolution inconsistencies observed in affine face wrappings. Limitations are Limited Training Data, Artefacts from Affine Warping.

III. SCOPE AND METHODOLOGY

Scope

The focus of this project is to develop a robust detection system using Convolutional Neural Networks (CNN) to identify AI-generated images that pose a threat to societal harmony. By leveraging CNN exclusively, the project aims to streamline the detection process, ensuring more efficient identification of false information spread through AI-generated content. Through the utilization of a single model, we simplify the detection mechanism, enhancing its reliability and effectiveness. By narrowing the scope to CNN-based detection, the project aims to address the proliferation of false information propagated by AI-generated images. By accurately distinguishing between genuine and synthetic content, the proposed system contributes to maintaining social and economic stability, fostering trust and reliability in digital information, and mitigating the risks associated with cybercrimes such as image forgery and impersonation.

The implementation of CNN-based detection represents a strategic response to the rising threat posed by AI-generated content. By harnessing the capabilities of Deep Learning tailored to image recognition tasks, the project aims to bolster social peace, financial stability, and the credibility of digital information.

Methodology

The methodology commences with the compilation of a diverse dataset comprising predominantly AI-generated images sourced from the Leonardo.AI tool, supplemented with real-world images. This dataset was categorized into two classes:AI-generated and real images.

Concurrently, for the Convolutional Neural Network (CNN) the data preprocessing included resizing and normalization steps applied to the raw image data. Following the dataset preparation, the model underwent training, with performance evaluation conducted using a suite of metrics including F1 score, Precision and Accuracy. Additionally, heatmap analysis, training and validation accuracy graphs, and confusion matrices were utilized to comprehensively assess the model's performance.

For the CNN model, a customized architecture was implemented using TensorFlow and relevant libraries. Further, model configurations were optimized with the help of hyperparameter tuning. Overall, the methodology encompassed dataset collection, feature extraction, model training, and thorough evaluation using diverse metrics and visualizations, ensuring comprehensive assessment of model performance.

IV. DESIGN AND IMPLEMENTATION

System design is a crucial phase following the Software Development Life Cycle (SDLC). It starts with requirement analysis and progresses to designing the overall system structure. The main goal of system design is to outline the architecture, modules, their relationships, purposes, and how they integrate. This phase provides a comprehensive overview of the system flow and architecture.

Architectural Design

Architectural design in software engineering entails the detailed description and visualization of a system's structure, serving as a blueprint for understanding its components, attributes, and interactions. This comprehensive process involves defining the system's building blocks, including modules, layers, and components, along with their externally visible

DOI: 10.17148/IJARCCE.2024.13346

properties and behaviors. Through architectural design, the relationships and dependencies among these components are meticulously delineated, guiding the system's behavior and functionality.

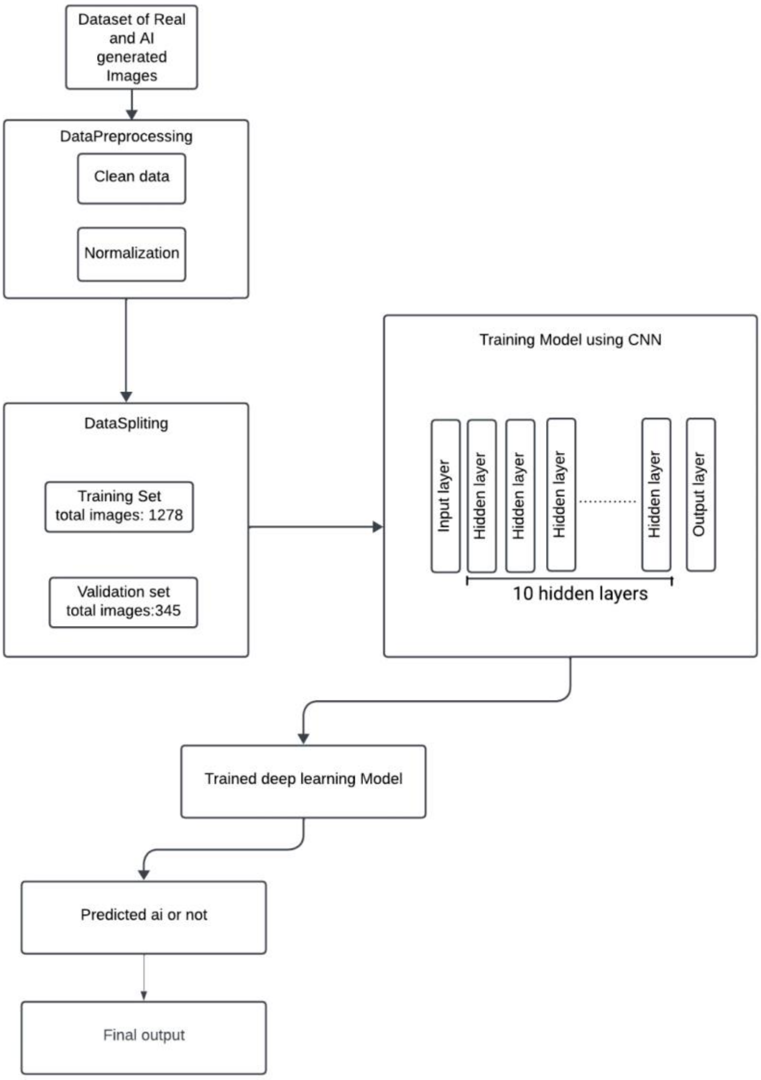
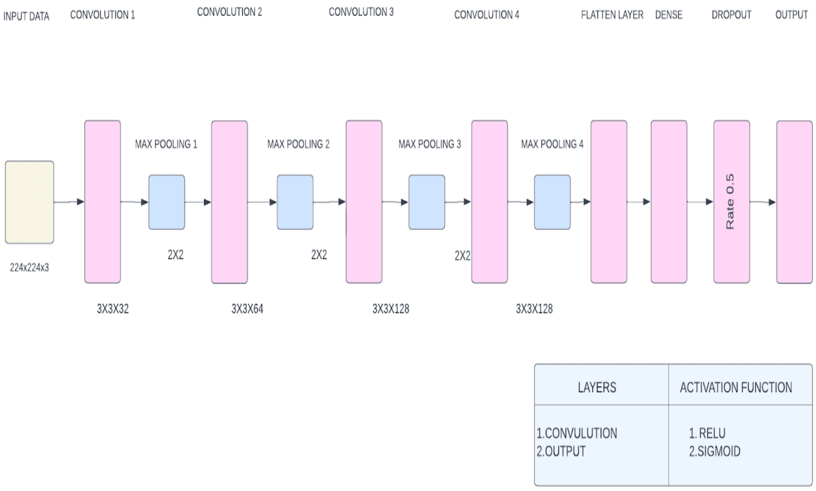


Fig 1. Architecture diagram of Proposed work



LAYERS	ACTIVATION FUNCTION
1.CONVULUTION 2.OUTPUT	1. RELU 2.SIGMOID

Fig 2. CNN Architecture

DOI: 10.17148/IJARCCE.2024.13346

V. RESULT AND CONCLUSION

Result

Our study focused on evaluating the effectiveness of a CNN model in detecting whether an image is AI-generated or not. The evaluation utilized a dataset containing both AI-generated and non-AI-generated images. Through comprehensive experimentation and validation, the CNN model demonstrated notable capabilities in distinguishing between the two types of images.
During the experimental phase, the CNN model was trained and validated using a dedicated training set. The validation process revealed promising results, with the CNN model achieving a validation accuracy of 96%. This high accuracy underscores the CNN model's proficiency in discerning AI-generated images from non-AI-generated ones.

These findings highlight the CNN model's heightened sensitivity and specificity in accurately identifying AI-generated images. Leveraging its advanced architecture and ability to capture intricate patterns within images, the CNN model emerges as a reliable tool for combating the dissemination of false information propagated through AI-generated content.

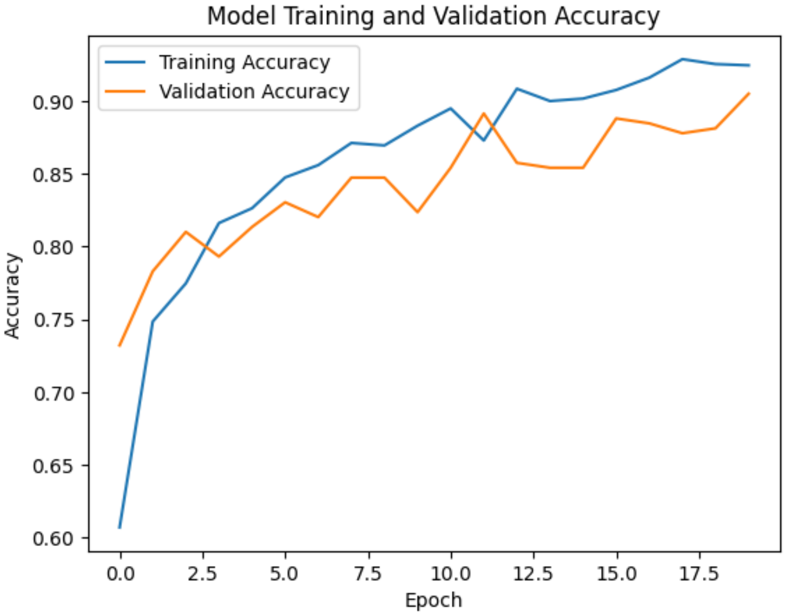


Fig 3. CNN model Result graph

CONCLUSION

In conclusion, our project signifies a significant advancement in leveraging Convolutional Neural Networks (CNN) for the detection of AI-generated images. Through meticulous experimentation and validation, we have showcased the efficacy of CNN in accurately discerning between AI-generated and genuine images. With a validation accuracy of 96%, CNN demonstrates remarkable sensitivity and specificity in identifying AI-generated content.

The integration of a CNN model into our detection system has proven instrumental in enhancing its robustness and reliability. By relying solely on CNN for image classification, we have achieved a streamlined approach that prioritizes accuracy and efficiency. The utilization of CNN as the sole model eliminates complexities associated with ensemble methods and underscores the standalone capabilities of CNN in tackling the challenges posed by AI-generated content.

Looking ahead, our study lays the groundwork for further advancements in CNN-based detection systems for combating the proliferation of false information. By continuing to refine and optimize CNN models, we can ensure their seamless integration into digital platforms and social media networks. Collaborative efforts between researchers, policymakers, and technology stakeholders are essential for harnessing the full potential of CNN-based solutions and safeguarding the integrity of digital information landscapes.

DOI: 10.17148/IJARCCE.2024.13346

REFERENCES

[1]. Diallo, B., Urruty, T., Bourdon, P., & Fernandez-Maloigne, C. (2020). Robust forgery detection for compressed images using CNN supervision. Forensic Science International Reports, 2, 100112.
[2]. Chen, H., Chang, C., Shi, Z., & Lyu, Y. (2022). Hybrid features and semantic reinforcement network for image forgery detection. Multimedia Systems, 28(2), 363-374.
[3]. El Biach, F. Z., Iala, I., Laanaya, H., & Minaoui, K. (2021). Encoder-decoder based convolutional neural networks for image forgery detection. Multimedia Tools and Applications, 1-18.
[4]. Aria, M., Hashemzadeh, M., & Farajzadeh, N. (2022). QDL-CMFD: a quality-independent and deep learning-based copy-move image forgery detection method. Neurocomputing, 511, 213-236.
[5]. Ali, S. S., Ganapathi, I. I., Vu, N. S., Ali, S. D., Saxena, N., & Werghi, N. (2022). Image forgery detection using deep learning by recompressing images. Electronics, 11(3), 403.
[6]. Quan, W., Wang, K., Yan, D. M., & Zhang, X. (2018). Distinguishing between natural and computer-generated images using convolutional neural networks. IEEE Transactions on Information Forensics and Security, 13(11), 2772-2787.
[7]. Qazi, E. U. H., Zia, T., & Almorjan, A. (2022). Deep learning-based digital image forgery detection system. Applied Sciences, 12(6), 2851.
[8]. Deepfake Detection through Deep Learning , Publisher: IEE Deng Pan; Lixian Sun; Rui Wang; Xingjian Zhang; Richard O. Sinnott
[9]. Deepfakes: Detecting forged and synthetic media content using machine learning. In Artificial Intelligence in Cyber Security: Impact and Implications; Springer: Berlin/Heidelberg, Germany, 2021; pp. 177–201.
[10]. Online Detection of AI-Generated Images , David C. Epstein, Ishan Jain, Oliver Wang, Richard Zhang; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2023, pp. 382-392.
[11]. Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review February 2021. [12]. DEEPFAKE DETECTION THROUGH DEEP LEARNING 1Kanchan Warke, 2Nilam Dalavi