International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 11 | November - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

GENERATION AND DETECTION OF ARTIFICIAL FACIAL IMAGES: AN EXPLORATION

Dr. Suvarna Pawar[1], Sudarshan Vetal[2], Sumit Sharma[3], Avishkar Pagare[4], Chetan Patil[5]

*1 Associate Professor, Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune, Maharashtra, India
*2,3,4,5 Department of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune, Maharashtra, India

---

Abstract - This project investigates the surge in AI-generated images, particularly the realistic portraits of human faces. While offering creative and entertainment possibilities, it raises concerns about trust and authenticity in digital content due to the potential misuse of these AI-generated contents which can be used for creating deceptive identities. The dual-edge nature that this technology provides, are the critical tools that can distinguish artificial faces from genuine ones. Our objective is a comprehensive exploration of techniques for both generating and detecting artificial facial images, evaluating accuracy score for each models that we list below.

Generation: DALL-E2, Mid Journey, Stable Diffusion

Detection: Resnet-50, VGG-16, Inception, Xception, AlexNet.

The outcome of this research will help the researchers and students all around the world with their ongoing discussions on responsible AI development.

***Key Words*:** AI-generated images, trust and authenticity, deceptive identities, models, responsible AI development

1. INTRODUCTION

The project, Generation and Detection of Artificial Facial Images: An Exploration, involves a thorough exploration of Artificial Intelligence (AI) and Machine Learning (ML) techniques for generation and detection of artificial facial images. This exploratory approach is aimed to advance our understanding and capabilities in two critical areas: the generation of artificial facial images using AI techniques and the subsequent detection of such artificially generated facial images.

In the aspect of image generation, this project focuses into AI technologies to generate realistic facial images. Using deep learning models and other AI methodologies, we aim to achieve generation of lifelike facial representations. This component of the project holds significant use for applications ranging from digital art and entertainment to avatar creation.

The detection aspect of the project addresses the ethical and security challenges posed by AI-generated facial images, providing multiple ways of detection and demonstrating insights and findings using visualization. This exploration aims to develop and enhance detection methods capable of distinguishing between authentic and AI-generated facial images. This research oriented exploratory approach contributes to analysis of multiple techniques for detection that can be applicable in various aspects such as identity theft and authentication.

In summary, the central objectives of this research-oriented exploratory approach encompass a multifaceted investigation, which can be described as follows:

1. **Image Generation:** To leverage AI technologies, including deep learning models and neural networks, to create highly realistic facial images that are almost similar to real human faces. The generated images have the potential to be employed in various applications, including digital art, entertainment and avatar creation.

2. **Image Detection:** To develop, refine and expand the techniques for detecting AI-generated facial images. These techniques address ethical and security concerns and have relevance in areas such as identity theft prevention and authenticity verification.

This project seeks to explore and experiment capacity of AI to generate incredibly convincing artificial facial images and the need for robust mechanisms to detect and differentiate between authentic and artificially generated facial images. By doing so, it contributes to the broader discourse on the applications and implications of AI in our increasingly visual and interconnected world.

2. LITERATURE SURVEY

The rapid advancement of deep learning techniques in recent years has revolutionized the generation of artificial facial images. A vast progress has been made in generating realistic artificial facial images and along with it many challenges have been encountered, associated with detection of such AI-generated images.

There was a study on DALL-E2 which uses Dataset with 27 text prompts, five AI-Text-To-Image generators. And it found that DALL-E-2 produces most appealing and realistic images. Also Existing models show limited performance for AI-generated images.[1]

CelebA-HQ and FFHQ datasets were used by Style-based generator, Variational Autoencoder and found that ResNet50 excels in binary and multi-class scenarios, While VGG16, VGG19 perform well in other scenarios.[2]

A Prior work used a number of datasets in which there was a comparative analysis of major deepfake algorithms and variants. It used the deepfake generation algorithms using GANs, CNN, RNN, LSTM for detection.[3]

A particular study which was published just recently in IEEE 2023, used Xception model, CelebA-HQ and Style-GAN datasets, preprocessing (resizing), training for different epochs and found that, testing accuracy of 98.77%, Xception outperforms other CNN models.[5]

Categorization of GAN face detection reviewed approaches in deep learning physical, physiological, and human visual performance. This study highlights challenges, breakthroughs, and importance of human visual performance.[6]

3. METHODOLOGY

3.1 Functional Analysis

1. **User Authentication and Access Control:** User authentication to ensure authorized access to the system.

2. **Model Selection:** Model selection involves choosing an appropriate Generative Model among (Dall-E2, Mid Journey, Stable Diffusion) and among (Resnet-50, AlexNet, VGG-16, Inception, Xception) to serve as the foundation for detecting artificial facial images.

3. **Data Input and Output:** Users would be able to input real or artificial facial images for generation and detection analysis. The system would display the results for generated image or image detection authenticity analysis, including whether the image is real or artificial.

4. **Image Generation Functionality:** The system will be able to generate artificial facial images from a given input, using machine learning models. Users will have the option to specify parameters for image generation, such as age, gender, or ethnicity or can directly input a facial image to generate corresponding artificial facial image**.**

5. **Image Detection Functionality:** The system must be capable of analyzing and determining the authenticity of facial images, classifying them as real or artificially generated. It will provide a probability of authenticity for each image.

6. **Feedback and Reporting:** Allows users to provide feedback on the system's performance and report issues or concerns.

3.2 Datasets

We start with acquiring and collecting all the datasets for the project. It involves a combination of searching, downloading and possibly preprocessing the data. Here is a detail list of sources through which we are going to do this:

- **Public Repositories:** Many datasets, such as Celeb-A and LFW, are available on public repositories like Kaggle, GitHub, or the UCI Machine Learning Repository.

- **Official Websites:** Some datasets, like FFHQ and DFDC, have official websites where we can find information on how to download the data.

- **Research Papers:** Checking the research papers that introduced or used the datasets we're interested in. Authors often provide download links or instructions in the papers.

- **Data Request:** In some cases, we need to request access to a dataset, if it contains sensitive information. Following the procedure outlined by the dataset provider will get the job done.

- **APIs:** Some datasets offer APIs for easy access. For example, Celeb-A can be accessed through the CelebA-HQ API.

3.3 External Interface Requirements

We will be getting the subscription of these models. By integrating these in our system. After which, we will be able to achieve the generative functionality in our system.

DALL-E2: DALL-E2 are text-to-image and image to image models developed by OpenAI using deep learning methodologies to generate digital images from natural language descriptions, called “prompts” and “images”.

Mid Journey: Mid journey is a generative artificial intelligence program and service created and hosted by San Francisco-based independent research lab Mid journey, Inc. Mid journey is currently only accessible through a Discord bot on their official Discord server, by direct messaging the bot, or by inviting the bot to a third-party server.

Stable Diffusion: Stable Diffusion is a deep learning, text-to-image model released in 2022 based on diffusion techniques. It is primarily used to generate detailed images conditioned on text descriptions and generating image-to-image translations guided by a text prompt.

3.4 Project Process Modeling

Agile Model

The Agile model, a well-established software development approach, finds remarkable applicability in our project, "Generation and Detection of Artificial Facial Images". In an era characterized by rapid technological advancements and ever-evolving user demands, this model's inherent flexibility and adaptability prove invaluable. Our project, centered on the creation of lifelike facial images using AI technologies and the development of robust detection methods, benefits greatly from the Agile methodology.

Agile Implementation for the project:

1. Backlog Creation
2. Breaking down tasks
3. Progress updates and planning
4. Iterative Development
5. Review and Retrospective
6. Collaboration
7. Continuous Integration and Testing
8. User Involvement for feedback
9. Backlog Refinement
10. Documentation
11. Continuous Improvement

4. PROPOSED SYSTEM

Architecture

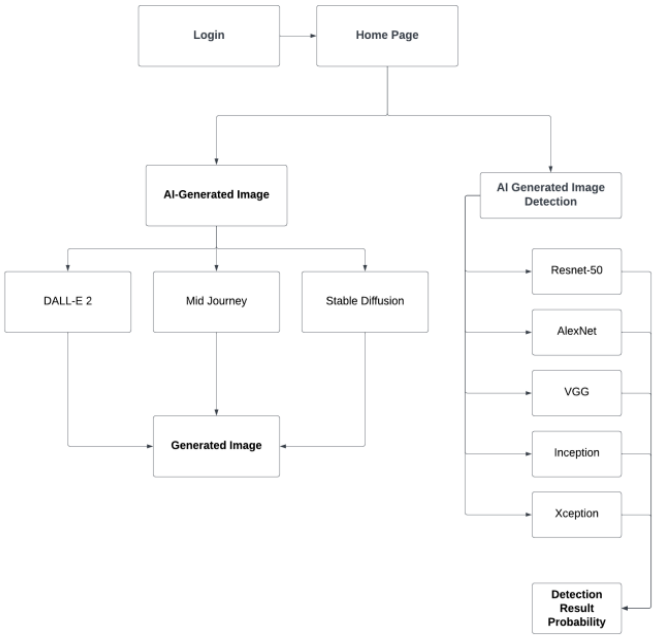

Figure 4.1: Architecture Diagram

System Overview

The system is Web-Based and the project aims to explore and develop a comprehensive system for generating artificial facial images and detecting AI-generated facial images and their authenticity using multiple detection models. Comparison in the form of visualization and metrics will be provided with other models.

Key Components:

1. Artificial Facial Image Generation Mechanism: Generation of artificial facial images. Training the model on a dataset of real facial images to produce convincing artificial counterparts. 2. Artificial Facial Image Detection Mechanism: Implementing image detection techniques using machine learning approaches, for artificial image detection and authentication.

2. Artificial Facial Image Detection Mechanism: Implementing image detection techniques using machine learning approaches, for artificial image detection and authentication.

3. Facial Images Dataset: Collecting and curating a diverse dataset of real and artificial facial images for training and testing purposes. Planned Datasets, but not limited to : FFHQ, Face Forensics, Forgery Net, Celeb-A, etc.

4. User Interface (UI): Has a user-friendly interface for users to interact with the system. Enable users to input and analyze facial images for authenticity.

5. Performance Evaluation: Conducting the comprehensive performance evaluation to assess the accuracy and reliability of the multiple detection mechanisms using standard evaluation metrics to measure the performance.

6. Research and Development: Investigating the latest techniques and advancements in image generation and detection to improve system capabilities. At the same time, exploring the ethical implications and potential applications of the technology.

5. CONCLUSIONS

The project "Generation and Detection of Artificial Facial Images: An Exploration" has successfully addressed the complex and multifaceted domain of artificial image generation and detection. It has achieved the core objectives of creating highly realistic artificial facial images and developing effective detection methods for distinguishing AI-generated images from authentic ones. The project has implemented three advanced generation models and five robust detection models, demonstrating versatility in addressing the challenges of image manipulation and identification. The research-oriented exploratory approach has contributed to a deeper understanding of AI's capabilities and ethical implications in the context of image generation and detection.

ACKNOWLEDGEMENT

The implementation and documentation of this project would not be succeeded without the kind support from individuals. First of all, we would like to express our gratitude to Dr. Suvarna Pawar, who always gives us valuable advice and kind assistance throughout this project. Last but not least, we would like to thank the Faculty of Computer Engineering, Savitribai Phule Pune University for giving us the great knowledge.

REFERENCES

Book:

[1] Christian Rathgeb, Ruben Tolosana, Ruben Vera-Rodriguez, Christoph Busch, “Handbook of Digital Face Manipulation and Detection”, Springer, 2022.

IEEE Papers:

[1] Steve Göring, Rakesh Rao Ramachandra Rao, Rasmus Merten and Alexander Raake, "Evaluation of AI-Generated Image Appeal and Realism", 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2023.

[2] Mohammed Berrahal, Mohammed Boukabous, Idriss Idrissi, "A Comparative Analysis of Fake Image Detection in Generative Adversarial Networks and Variational Autoencoders", International Conference on Decision Aid Sciences and Applications (DASA), 2023.

[3] Nikhil Sontakke, Sejal Utekar, Shivansh Rastogi, Shriraj Sonawane, "Comparative Analysis of Deep-Fake Algorithms", International Journal of Computer Science Trends and Technology (IJCST) – Volume 11 Issue 4, 2023.

[4] Galamo Monkam, Weifeng Xu, Jie Yan, "A GAN-based Approach to Detect AIGenerated Images", IEEE International Conference on Machine Learning and Applications (ICMLA), 2022.

[5] Bagus Izzan Muafy, Febryanti Sthevanie, Kurniawan Nur Ramadhani, "Generated AI Face Detection using Xception Model", International Conference on Advances in Electrical, Computer and Communication Technologies (ICAECCT), 2022.

[6] Xin Wang, Hui Guo, Shu Hu, Ming-Ching Chang, Siwei Lyu, "GAN-generated Faces Detection: A Survey and New Perspectives," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

[7] Mohana, D. M. Shariff, A. H and A. D, "Artificial (or) Fake Human Face Generator using Generative Adversarial Network (GAN) Machine Learning Model", 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2021.

[8] Nguyen, Thanh Thi, et al., "Deep learning for deepfakes creation and detection: A Survey", 2022 Computer Vision and Image Understanding, 2022.

[9] Yang Yu, Rongrong Ni and Yao Zhao, Senior Member, "Mining Generalized Features for Detecting AI-Manipulated Fake Faces", IEEE, 2022.

[10] D. Yadav, S. Salmani, "Deepfake: A survey on facial forgery technique using generative adversarial network", 2019 International Conference on Intelligent Computing and Control Systems (ICCS), IEEE, 2019.

[11] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, "Deep learning for deepfakes creation and detection: A survey", arXiv preprint arXiv:1909.11573.

[12] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, "Deepfakes and beyond: A Survey of face manipulation and fake detection", Information Fusion, 2020.

[13] S. Agarwal, N. Girdhar and H. Raghav, "A Novel Neural Model based Framework for Detection of GAN Generated Fake Images", 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021.

[14] S. Ramachandran, A. V. Nadimpalli and A. Rattani, "An Experimental Evaluation on Deepfake Detection using Deep Face Recognition", International Carnahan Conference on Security Technology (ICCST), Hatfield, United Kingdom, 2021.

[15] Leandro A. Passos, Danilo Jodas, Kelton A. P. da Costa, Luis A. Souza Júnior, Douglas Rodrigues, Javier Del Ser, David Camacho, João Paulo Papa, "A Review of Deep Learning-based Approaches for Deepfake Content Detection", 2022.