Indira Gandhi Delhi Technical University For Women

Department Of Information Technology, IGDTUW

Summer Internship On Generative AI

Project Report

Data Augmentation Using Generative AI: Synthetic Data Production:Automatic human facial expression image generator using generative AI

Internship Period:3rd June,2024 - 12th July,2024

Date of Submission:

Intern: Medhavi Bhardwaj(03201192022), B.Tech. (AI-ML), 2nd year

DECLARATION

We, the undersigned, declare that the project report entitled "**Automatic human facial expression image generator using generative AI**" is the result of our own collective work and investigation completed during our Six Week Summer Internship on "**Generative AI and Prompt Engineering"**. This work is original and has not been copied or reproduced from any other source, except where specifically acknowledged.

The project has been conducted under the supervision of Dr. Santanoo Pattnaik . We affirm that we have adhered to all academic and ethical guidelines prescribed by the institution. We acknowledge all sources of assistance and contributions received during the preparation of this report. Any false statement in this declaration may lead to disciplinary action as per the institution's regulations.

Interns:

Medhavi Bhardwaj [03201192022]

Kalpana Mishra [2022004703]

Ananya Bansal [00401022023]

Madhulika Chaudhary [03001192022]

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to all those who have contributed to the successful completion of this project : **“Automatic human facial expression image generator using generative AI”**.
First and foremost, we extend our deepest appreciation to our project supervisor, Dr. Santanoo Pattnaik, for their unwavering support, guidance, and valuable insights throughout the entire project.
We would also like to thank our fellow team members for their dedication, hard work, and collaborative spirit.
We are grateful to all the mentors for providing their valuable feedback and suggestions that have helped refine the contents of this report.
I would like to acknowledge the support and cooperation received from Indira Gandhi Delhi Technical University For Women, as well as the resources and facilities provided, which have significantly contributed to the successful execution of this project.
Last but not least, I would like to express my deep appreciation to my family and friends for their understanding, encouragement, and patience throughout this project journey. Their unwavering support has been a constant source of motivation.
Each of the above individual’s contributions has played a crucial role in shaping this report and enhancing our learning experience.

ABSTRACT

This project aims to generate images of human facial expressions using generative AI techniques to enhance machine learning models for sentiment analysis and emotion identification. The significance of this project lies in its potential to provide high-quality, diverse datasets for training sentiment analysis models, improving their accuracy and reliability in real-world applications. The project addresses the challenge of obtaining diverse and representative datasets, which are crucial for developing robust sentiment analysis models.

OBJECTIVES

Generate a diverse set of human facial expressions using generative AI.Create a high-quality dataset suitable for training sentiment analysis models.Evaluate the effectiveness of the generated dataset in improving the performance of sentiment analysis and emotion identification models.

EXPECTED OUTCOMES

Combination of data augmentation with an image-to-image random facial expression generation generative AI model is capable of generating high quality images of realistic and diverse facial expressions while addressing the limited availability of datasets and model training. The model will be evaluated on its ability to generate a collection of contextually correct and realistic images.

Image Generation

- Expression Specification: Define a method for specifying the desired emotion for image generation. This could involve encoding emotions as input vectors.

- Image Generation Process: Use the trained generative models to produce images based on specified emotions. Implement mechanisms to control the diversity and quality of generated images.

TABLE OF CONTENTS

1. Chapters

3. Figures

CHAPTER 1: INTRODUCTION

1.1 Background and Motivation

Currently we are living in the era of Generative AI. Generative AI refers to the use of AI to generate new content like text, images, audios, videos, etc. It fetches content from different places and arrange them according to our need. It all started long time back, in around 1960s with the introduction of chatbots. Slowly many advancements were made but it was not until 2014, with the introduction of generative adversarial networks , or GANS that generative AI could create convincingly authentic images, videos and audio of real people.

Image to image generative ai models are an important constituent of generative ai models. These take an image as input and produce the output images as per the training and requirement.

1.2 Problem Statement

Human facial expressions play a critical role in communication, influencing emotional understanding in various applications such as virtual reality, animation, and social robotics. However, existing generative models often fail to produce realistic and diverse facial expressions, resulting in limitations in emotional accuracy and user engagement.

1.3 Objectives of the Project

Facial expressions play a crucial role in human communication, but capturing diverse expressions for data augmentation in AI training is challenging. Our AI model generates high-quality images of human facial expressions, enhancing datasets and improving AI training.

This AI model aims to transform input images into outputs: newly generated images of facial expressions depicting the recognised facial expressions using deep learning techniques and algorithms like GANs. By using a dataset of facial expressions, its objective is to generate realistic variations such as happiness, sadness, surprise, and more.

This image-to-image generator AI model represents a significant advancement in AI-driven visual content creation, offering diverse applications across industries and research disciplines.

1.4 Scope Of The Project

The scope of our gen ai model for human facial expressions is to basically create a wider database of human facial expressions that can be further used to train deep learning models on emotion and sentiment analysis. The model can significantly impact various industries by addressing emotion generation, real time processing, contextual adaptation ,etc.

1.5 Structure of the Report

The report includes the literature review,introduction, background, methodology, output, interpretation, references

CHAPTER 2: LITERATURE REVIEW

2.1 Overview of Generative AI

Generative AI is a subfield of artificial intelligence focused on creating systems that can generate new content, such as text, images, music, and even code, that is indistinguishable from human-created content. It involves using machine learning models to generate data that mimics the distribution of a given dataset.

Generative AI is a subset of artificial intelligence that focuses on creating new content, rather than analyzing existing data. It's capable of generating various forms of content, including text, images, audio, video, and even code.

How Does it Work?

Generative AI models are trained on massive amounts of data. They learn patterns and relationships within the data, enabling them to generate new content that resembles the training data. This process often involves deep learning techniques, particularly neural networks.

Key Components and Techniques

- **Large Language Models (LLMs):** These models, like GPT-4, are trained on vast amounts of text data and can generate human-quality text, translate languages, write different kinds of creative content, and answer your questions in an informative way.
- **Generative Adversarial Networks (GANs):** These models consist of two neural networks competing against each other: a generator that creates new data instances, and a discriminator that classifies them as real or fake. This adversarial process leads to the generation of highly realistic content.
- **Diffusion Models:** These models start with random noise and gradually refine it into a desired output, such as an image, by removing noise step by step.

2.2 Recent Developments and Applications

Applications of Generative AI

The applications of generative AI are vast and continue to expand:

- **Content Creation:** Generating articles, poems, scripts, marketing copy, social media posts, and more.
- **Image and Video Generation:** Creating realistic images, art, and videos, including deepfakes.
- **Music and Audio Generation:** Composing music, creating sound effects, and generating speech.
- **Drug Discovery:** Designing new molecules for potential drug candidates.
- **Code Generation:** Automating parts of the coding process, suggesting code snippets, or even generating entire programs.

Challenges and Considerations

While generative AI offers immense potential, it also presents challenges:

- **Bias:** Generative AI models can perpetuate biases present in the training data.
- **Misinformation:** The creation of deepfakes and other misleading content can pose significant risks.
- **Ethical Implications:** Questions about copyright, ownership, and the potential misuse of generative AI need to be addressed.

The Future of Generative AI

Generative AI is rapidly evolving, with new advancements and applications emerging constantly. It has the potential to revolutionize various industries and change the way we interact with technology.

2.3 The applications of of generative AI in data augmentation

Generative AI has emerged as a powerful tool for enhancing data augmentation, significantly improving the performance of machine learning models. Here's a breakdown of its applications:

Image and Video Data Augmentation

- **Generating new images:** Creating diverse images through transformations like rotations, flips, crops, and color adjustments.
- **Creating realistic image variations:** Generating images with different lighting conditions, backgrounds, and object positions.
- **Synthetic image generation:** Generating entirely new images based on existing data, such as generating faces or objects.
- **Video augmentation:** Enhancing video datasets by creating variations in speed, frame rate, and adding noise.

Text Data Augmentation

- **Generating synthetic text:** Creating new text samples by paraphrasing, back translation, and synonym replacement.
- **Augmenting text with noise:** Introducing typos, misspellings, or grammatical errors to improve model robustness.
- **Data augmentation for specific tasks:** Generating text for tasks like sentiment analysis, text summarization, and machine translation.

Audio Data Augmentation

- **Generating synthetic audio:** Creating new audio samples by adding noise, changing pitch, and altering speed.
- **Augmenting speech data:** Generating different accents, background noises, and speech variations.

- **Data augmentation for music:** Creating new music samples by manipulating tempo, pitch, and rhythm.

Medical Image Data Augmentation

- **Generating synthetic medical images:** Creating artificial MRI, CT, and X-ray images for training medical image analysis models.
- **Augmenting existing medical images:** Applying transformations like rotations, flips, and noise to increase data diversity.

Benefits of Generative AI for Data Augmentation

- **Increased data volume:** Generating synthetic data to expand limited datasets.
- **Improved model performance:** Enhancing model generalization and robustness.
- **Addressing data imbalance:** Creating synthetic data for underrepresented classes.
- **Privacy preservation:** Generating synthetic data while protecting sensitive information.

Intelligent fault diagnosis, detection, and prognostics (DDP) for complex equipment prognostics and health management (PHM) have achieved remarkable breakthroughs. Equipment in industrial scenarios often operates in normal conditions, resulting in missing anomalies, limited failures, and incomplete degradation paths. Thus the limited information on the state of the equipment collected from sensor readings severely hinders the cognitive capabilities of discriminative artificial intelligence (AI) for PHM. Data augmentation and generation (DA&G) techniques, represented by generative AI, have shown great promise in overcoming the limitations of PHM application scenarios (Liu, et.al., 3034).

Tiwary Chaudhary, et. al. (2023) have conducted a study which considers shape, complexion and other identity related information separate from certain specified muscle movements which are specific for emotion recognition. This is done by a novel Emotion-Generative Adversarial Network, which saves a lot of effort and simplifies the Facial Expression Recognition process. Then Scale Invariant Feature Transformation and Vola John's face extraction method for pre-processing and face image extraction from background were applied which enabled the model to be evaluated accurately irrespective of scale, orientation, illumination etc and with very less training samples. The feature extracted facial image is fed to an attention-based Convolutional Neural Network. This will ensure more emphasis on critical areas for expression recognition of facial image. Finally, we have used Local Binary Pattern for classification of the input image to a particular emotion class. Facial Expressions are quite personalized and may look different for different individuals. Whereas there are certain facial muscles which show some common features for certain human expressions and facial shapes. Convolution Neural Networks have shown tremendous success in Facial Expression Recognition tasks. Most of Facial Expression Recognition (FER) systems rely on machine learning approaches that require large databases (DBs) for effective training.Since they are not easily available, a good solution is to augment the DBs with appropriate techniques, which are typically based on either geometric transformation or deep learning based technologies (e.g., Generative Adversarial Networks (GANs)). Whereas the first category of techniques have been fairly adopted in the past, studies that use of GAN-based techniques are limited for FER systems. To advance in this respect, we evaluate the impact of the GAN techniques by creating a new DB containing the generated synthetic images. The face images contained in the KDEF DB are used as the base to create novel synthetic images using the facial features of 2 images selected from the YouTube-Faces DB(Porcu, Simone, et. al., 2020).

CHAPTER 3: METHODOLOGY

3.1 Background

The project utilizes Generative Adversarial Networks (GANs) to generate realistic images of human facial expressions from existing images. GANs are a class of machine learning models comprising two neural networks: the generator and the discriminator. These two networks engage in a competitive process, where the generator attempts to create convincing fake images, and the discriminator strives to distinguish between real and generated images.

The generator takes random noise as input and produces images that mimic the real data distribution. The discriminator, on the other hand, evaluates images and outputs the probability that an image is real or fake. Through iterative training, the generator improves its ability to produce realistic images, while the discriminator becomes more adept at identifying fake images.

This adversarial process drives both networks to enhance their performance, resulting in high-quality and realistic generated images. This capability makes GANs particularly valuable for applications like training machine learning models for sentiment analysis and emotion recognition, where diverse and accurate facial expression images are crucial for effective model training and evaluation.

Effective training of neural networks requires much data. In the low-data regime, parameters are underdetermined, and learnt networks generalise poorly. Data Augmentation alleviates this by using existing data more effectively. However standard data augmentation produces only limited plausible alternative data. Given there is potential to generate a much broader set of augmentations, we design and train a generative model to do data augmentation. The model, based on image conditional Generative Adversarial Networks, takes data from a source domain and learns to take any data item and generalise it to generate other within-class data items(Storky, Edwards,et. al.,2017)

3.2 Implementation Details

1. **Data Preparation:**

- Images were sourced from the CelebA-HQ dataset, which contains high-quality images of human faces.

- Each image was resized to 64x64 pixels and normalized to a range of [-1, 1] for better performance during training.

2. **Model Design**:

- **Generator**: Designed to create images from random noise, it utilizes multiple layers, including dense, batch normalization, LeakyReLU, and Conv2DTranspose layers to generate 64x64 pixel images.

- **Discriminator**: Designed to distinguish real images from those generated by the generator. It employs Conv2D, LeakyReLU, Dropout, and Dense layers.

3. **Loss Functions and Optimizers:**

- **Loss Functions**: Binary cross-entropy loss was used for both generator and discriminator.

- **Optimizers:** Adam optimizer was employed for both the generator and discriminator with a learning rate of 0.0001.

4. **Training:**

- The model was trained for 50 epochs. During each epoch, the generator produced images from random noise, and the discriminator evaluated both real and generated images.

- Gradients were computed and applied to update the generator and discriminator weights.

CHAPTER 4: RESULTS AND ANALYSIS

4.1 Results

The images have been produced in batches, and the model ran for 50 epochs thus generating 50 batches of images each with varying degrees of clarity and quality.

These images can further be separated out by cropping to form a part of a database. These databases can further be used to train machine learning models for identifying facial expressions and what emotions are being shown in them. Thus these databases can also be used for sentiment analysis, which finds applications in a wide variety of fields.

Figure 1: Images generated as output

4.2 Interpretation

The AI model has been trained to generate high-quality images of human facial expressions using a dataset of facial expressions, primarily leveraging Generative Adversarial Networks (GANs). Here's an interpretation of the training and its results:

Epochs and Training Process

- **Epoch Completion:** The model was trained for 50 epochs, indicating 50 cycles through the entire training dataset. Each epoch represents one full pass through all the training data, allowing the model to learn and improve incrementally. The consistent completion of epochs from 19 to 50 suggests a steady progression in the training process.

Image Generation and Quality

- **Image Format and Quality:** The generated images are in Portable Network Graphic (PNG) format with a bit depth of 32, which includes true color with an alpha channel for transparency. The images have dimensions of 329 x 328 pixels, which is suitable for detailed facial expressions.
- **Software and Resolution:** The images were created using Matplotlib version 3.8.0, a popular Python library for generating static, animated, and interactive visualizations. The images have a resolution of 100 pixels per inch (PPI), which ensures clarity and detail.

Model Components

- **Generator Network:** Designed to create images from random noise, the generator uses various layers like dense, batch normalization, LeakyReLU, and Conv2DTranspose layers to produce 64x64 pixel images. The aim is to generate realistic images that mimic real facial expressions.
- **Discriminator Network:** Responsible for distinguishing between real and generated images, the discriminator employs layers such as Conv2D, LeakyReLU, Dropout, and Dense layers. It helps in evaluating the quality of the generated images and provides feedback to the generator for improvement.

Training Techniques

- **Loss Functions:** Binary cross-entropy loss was used for both the generator and discriminator, a standard choice for binary classification problems where the goal is to distinguish between two classes (real and fake images).
- **Optimizers:** The Adam optimizer with a learning rate of 0.0001 was employed for both networks, facilitating efficient and adaptive learning.

4.3 Inferences

Based on the successful training and the quality of the generated images, several inferences can be made:

Model Efficacy: The AI model demonstrates a high capability in generating realistic and diverse facial expressions.

Improved Datasets: This technology can significantly enhance datasets used for training other AI models, particularly in areas requiring nuanced emotional recognition.

Versatility: The model's applications across entertainment, human-computer interaction, research, and accessibility indicate its versatility and potential to impact multiple industries.

Future Potential: Continued advancements in GANs and image generation techniques can lead to even more sophisticated models, further expanding the possibilities for AI-driven visual content creation.

CHAPTER 5: DISCUSSION

5.1 Key Findings

The model successfully generates a wide range of realistic facial expressions, demonstrating the ability to capture subtle emotional variations such as joy, sadness, and surprise thus enhancing its scope of application. By training on diverse datasets, the model effectively represents various ethnicities and age groups, addressing biases often seen in generative models. The generated data can serve as valuable training data for improving emotion recognition systems, contributing to advancements in AI understanding of human emotions.

Thus it can be integrated with various deep learning models making them intelligent enough to do sentiment analysis for various purposes.

5.2 Implications

Some of the implications highlighting the potential of our generative ai model are:

Entertainment: Creating animated characters with lifelike expressions in movies, games, and virtual reality.

Human-Computer Interaction: Enhancing user interfaces by providing emotionally responsive avatars or virtual assistants.

Research: Studying human emotions and behaviors by generating controlled facial expressions for psychological experiments.

Accessibility: Supporting individuals with communication challenges by creating personalized avatars capable of expressing a wide range of emotions.

5.3 Limitations

Although considerably fine output is generated however, the quality and resolution of the output images is not very good. We aimed of making a high quality image generating ai model but the resultant images are blurred and not very clear.

Challenges and Considerations

- **Quality of synthetic data:** Ensuring generated data is realistic and representative.
- **Computational cost:** Training generative models can be computationally intensive.
- **Ethical implications:** Addressing biases and potential misuse of generated data.

By effectively leveraging generative AI for data augmentation, organizations can significantly enhance the performance of their machine learning models and unlock new possibilities in various domains.

5.4. Recommendation for Future Work

Evaluation and Improvement

• Quality Assessment: Use metrics like FID (Frechet Inception Distance) to evaluate the quality of generated images.

- Diversity Assessment: Ensure the generated images are diverse across different demographics.
- Iterative Improvement: Based on evaluation results, refine and retrain the models to improve performance.
- Performance Evaluation: Evaluate the performance of these datasets using ML models and adjust the image generation process as needed.

A lot more can be created with this gen ai model of ours like a more expanded dataset collection which consists of many more facial expressions making it more realistic and applicable.

We can also focus on optimizing the model for faster processing times, enabling seamless real-time applications in interactive environments. also on generating more nuanced and subtle emotional expressions to capture complex human feelings and improve authenticity.

We can also explore the use of generated expressions in AR/VR environments, enhancing user immersion and emotional connection.

CHAPTER 6: CONCLUSION

6.1 Summary of the Work

This project focused on generating realistic human facial expressions using Generative Adversarial Networks (GANs) to enhance datasets for sentiment analysis and emotion recognition. The model was trained on high-quality datasets, producing diverse and realistic emotional expressions. Images were generated in sets- one set for each epoch.

6.2 Achievements of the Project

- Developed a GAN-based model to generate high-quality, diverse facial expressions.
- Enhanced datasets for training sentiment analysis models, improving accuracy and reliability.
- Addressed biases by including various ethnicities and age groups in the training data.

6.3 Concluding Remarks

The project successfully generated realistic facial expressions, enhancing AI models for emotion recognition. While image clarity needs improvement, future work will focus on optimizing real-time applications and expanding emotional expression range, advancing AI-driven visual content creation.

6.4 Future Directions:

By increasing the quality of the generated images, by using larger sample input dataset size, for the training data to train the generative AI model, or fine tuning parameters used to generate the images, we can get better quality synthetic data (images) of human facial expressions.

REFERENCES

1. Tiwary, Gyanendra & Chauhan, Shivani & Goyal, Krishan. (2023). Facial Expression Recognition Using Expression Generative Adversarial Network and Attention CNN. 447-454.

2. Porcu, Simone, Alessandro Floris, and Luigi Atzori. 2020. "Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems" *Electronics* 9, no. 11: 1892. https://doi.org/10.3390/electronics9111892

3. https://www.techtarget.com/searchenterpriseai/definition/generative-AI#:~:text=Generative%20AI%20was%20introduced%20in,and%20audio%20of%20real%20people.

4. Liu, S., Chen, J., Feng, Y., Xie, Z., Pan, T., & Xie, J. (2024). Generative artificial intelligence and data augmentation for prognostic and health management: Taxonomy, progress, and prospects. *Expert Systems with Applications*, *255*, 124511.

5. Antoniou, A., Storkey, A., & Edwards, H. (2017). Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*.

6. https://www.aiacceleratorinstitute.com/the-5-primary-generative-ai-applications-and-how-they-work/