HUMAN FACES GENERATION USING DCGAN

[1]SHAIK PARVEEN, [2]P.BOBBY SOWJANYA

[1]Students, Department of MCA, B V Raju College, Bhimavaram Ap

[2]Assistant Professor, Department of MCA, B V Raju College, Bhimavaram Ap

ABSTRACT

This project presents an advanced deep learning framework for generating realistic human faces and synthesizing images from textual descriptions. The system leverages Deep Convolutional Generative Adversarial Networks (DCGAN), consisting of a Generator and Discriminator, to produce high-quality human face images. The Generator creates synthetic images from learned patterns, while the Discriminator evaluates their authenticity by distinguishing between real and generated images. Additionally, the project integrates a hybrid CNN-LSTM architecture for text-to-image generation, where the LSTM extracts semantic features from textual input and the CNN utilizes these features to generate corresponding images. The model is trained on a limited dataset of approximately 8000 images, enabling it to generate contextually relevant outputs for predefined text inputs. The system includes multiple modules such as user authentication, DCGAN training and loading, face generation, and text-to-image synthesis, all accessible through a web-based interface. Experimental results demonstrate an accuracy of 96% in distinguishing real and generated images, with progressive improvement observed during training epochs. This project highlights the potential of combining generative models and sequence learning techniques for creative AI applications, while also addressing challenges related to limited data and computational constraints.

Keywords: DCGAN, CNN, LSTM, Generative Adversarial Networks, Human Face Generation, Text-to-Image Synthesis, Deep Learning, Image Generation, Artificial Intelligence.

I.INTRODUCTION

Artificial Intelligence (AI) and Deep Learning have revolutionized the field of computer vision by enabling machines to understand, analyze, and generate visual data. One of the most impactful innovations in this domain is the development of Generative Adversarial Networks (GANs), which can create realistic images by learning patterns from large datasets. GANs consist of two competing neural networks—a Generator and a Discriminator—that work together to improve the quality of generated outputs. This project utilizes a specialized form of GAN known as Deep Convolutional Generative Adversarial Network (DCGAN), which incorporates convolutional layers to enhance feature extraction and image

quality. The ability of DCGAN to generate realistic human faces has opened new possibilities in areas such as virtual avatars, entertainment, and data augmentation.

In addition to face generation, the project also focuses on text-to-image synthesis, which is an emerging area in deep learning. This technique allows the system to generate images based on textual descriptions, bridging the gap between natural language processing and computer vision. The project implements a hybrid model combining Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). The LSTM processes input text and extracts meaningful semantic features, while the CNN uses these features to generate corresponding images. Although the model is trained on a limited dataset of around 8000 images, it is capable of producing relevant outputs for predefined inputs, demonstrating the feasibility of such systems even with constrained resources.

The system is designed with multiple functional modules, including user authentication, DCGAN training and loading, human face generation, and text-to-image conversion. A web-based interface is developed to ensure ease of use and accessibility for users. The training results indicate that the model achieves high accuracy in distinguishing real and generated images, with performance improving over training epochs. Despite limitations such as dataset size and computational constraints, the project successfully demonstrates the integration of generative models and sequence learning techniques. Overall, this work highlights the potential of combining DCGAN and CNN-LSTM architectures in building intelligent systems capable of creative image generation and real-world applications.

II SURVEY OF RESEARCH

The study by Ian Goodfellow et al. (2014) [1] introduced Generative Adversarial Networks (GANs), a groundbreaking framework for generating realistic data. The methodology involves two neural networks, namely the Generator and Discriminator, trained in a minimax game setting. Results demonstrated the ability to generate high-quality synthetic images resembling real data. However, the model suffers from issues such as training instability and mode collapse. This research laid the foundation for advanced generative models like DCGAN used in this project.

The study by Alec Radford et al. (2015) [2] proposed Deep Convolutional GAN (DCGAN) to improve image generation quality. The methodology replaces fully connected layers with convolutional layers to enhance feature extraction. Results show that DCGAN can generate realistic human faces with stable training compared to traditional GANs. However, it requires careful tuning of

hyperparameters and large datasets. This work directly supports the face generation module in the proposed system.

The study by Sepp Hochreiter and Jürgen Schmidhuber (1997) [3] introduced Long Short-Term Memory (LSTM) networks for sequence learning tasks. The methodology uses memory cells and gating mechanisms to capture long-term dependencies in data. Results showed significant improvements in text processing and sequence prediction tasks. However, LSTMs can be computationally expensive and slow to train. This research forms the basis for extracting meaningful features from text in the text-to-image module.

The study by Yoshua Bengio et al. (2009) [4] explored Convolutional Neural Networks (CNNs) for image processing and recognition tasks. The methodology utilizes convolution and pooling layers to automatically learn spatial hierarchies of features. Results demonstrated high accuracy in image classification and feature extraction. However, CNNs require large labeled datasets for optimal performance. This research supports the image generation process using extracted text features.

The study by Scott Reed et al. (2016) [5] proposed a text-to-image synthesis model using GANs conditioned on textual descriptions. The methodology combines text embeddings with image generation networks to produce images aligned with input text. Results showed that the model could generate plausible images from simple descriptions. However, generated images lacked fine details and required large datasets. This study is closely related to the text-to-image module in the proposed project.

The study by Han Zhang et al. (2017) [6] introduced StackGAN, a two-stage GAN architecture for high-resolution image synthesis from text. The methodology generates images in stages, refining details progressively. Results showed significant improvements in image quality and realism. However, the model increases computational complexity and training time. This research highlights advancements in improving text-to-image generation quality beyond basic GAN models.

III. WORKING METHODOLOGY

The proposed system follows a deep learning-based approach integrating DCGAN and CNN-LSTM architectures for image generation tasks. Initially, the dataset consisting of human face images is preprocessed through resizing, normalization, and augmentation to improve model performance. The DCGAN model is then trained using two neural networks: the Generator and the Discriminator. The Generator creates synthetic images from random noise, while the Discriminator evaluates whether the images are real or generated. Both networks are trained simultaneously in an adversarial manner, improving each other's performance iteratively.

As training progresses, the Generator becomes capable of producing realistic human faces, while the Discriminator enhances its classification accuracy. The system monitors performance using loss and accuracy graphs across training epochs to ensure proper convergence.

Once the DCGAN model is trained, it is deployed within the system for real-time human face generation. Users can upload input images through the web interface, and the system processes these images to compare them with generated outputs. The Generator produces multiple synthetic face images, and the Discriminator classifies them as real or fake. The output includes the original image along with several generated images, clearly indicating their classification results. This module demonstrates the effectiveness of generative models in producing realistic visual content. Although the system is trained on a relatively small dataset, it is optimized to provide acceptable results with minimal computational resources, making it suitable for practical applications and academic purposes.

In parallel, the text-to-image generation module utilizes a hybrid CNN-LSTM architecture to convert textual descriptions into visual representations. The LSTM network processes the input text and extracts meaningful semantic features by capturing contextual dependencies in the sequence. These extracted features are then passed to the CNN model, which generates images corresponding to the textual input. The model is trained on a limited dataset; therefore, it performs best with predefined or sample input sentences. The system provides an interactive interface where users can input text and view generated images. This integration of natural language processing and computer vision demonstrates the potential of multimodal deep learning systems in creative AI applications.

IV RESULTS EXPLANATIONS

In propose project we are employing DCGAN (deep convolution generative adversarial network) to generate human faces. DCGAN consists of two different networks such as Generator and Discriminator where generator will generate faces from trained model and then Discriminator will predict weather uploaded and generated faces are real or fake.

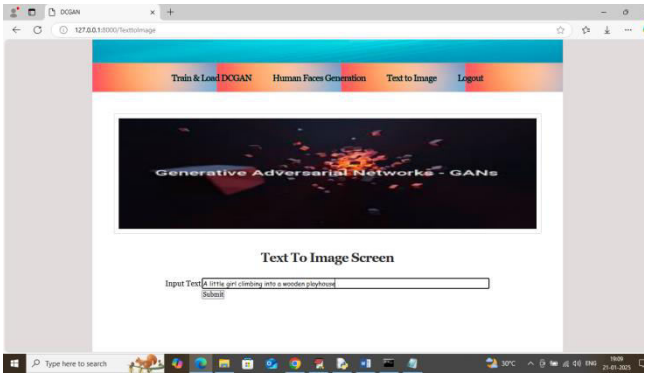
Text to Image Generation

In this module we have used combination of CNN and LSTM algorithm to generate images from text where CNN will be responsible to generate images by utilizing text features predicted by LSTM. This model trained on Image Net dataset and this model will be applied on LSTM features to generated related image. In real time they will train on millions of images so generation will be accurate but we trained the model on 8000 images only so it

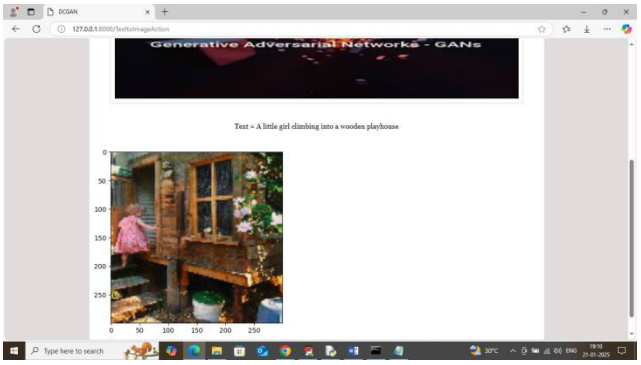
will work for few sentences which we have given in 'samples_input_text.txt' file.

To implement this project we have designed following modules

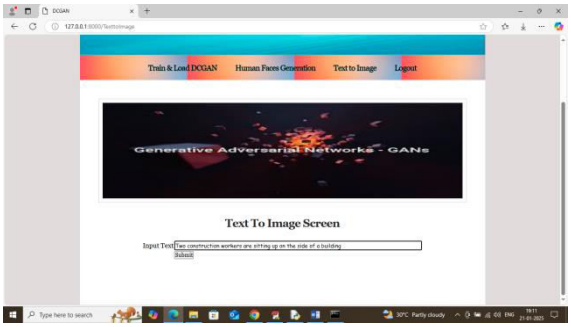
1) User Login: user can login to system using username and password as 'admin and admin'.
2) Train & Load DCGAN: after login user will use this module to generate and load DCGAN model and then perform prediction on test data to calculate prediction accuracy
3) Human Faces Generation: using this module user can upload any image and then DCGAN will generate some images and then apply Discriminator to detect those images as Real or Fake (Ai generated).
4) Text to Image: in this module user can upload some text and then application will generate related image.



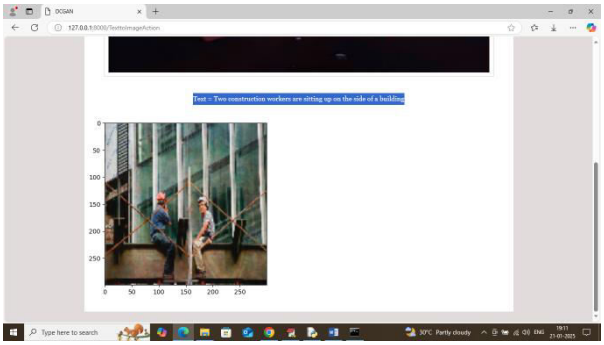
In above screen entering some text and then click on 'Submit' button to get below page



In above screen can see generated image from given text and similarly you can use sentences given in 'samples_input.txt' file. Below is another example



In above screen entering some other text and then will get below output



In above screen can see generated image from given text and you can try other sentences.

V.CONCLUSION

The proposed project successfully demonstrates the application of deep learning techniques for human face generation and text-to-image synthesis. By utilizing Deep

Convolutional Generative Adversarial Networks (DCGAN), the system is capable of generating realistic human face images through adversarial training between the Generator and Discriminator. Additionally, the integration of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks enables the system to generate images from textual descriptions, bridging the gap between natural language processing and computer vision. The system achieves promising results with an accuracy of 96% in distinguishing real and generated images, validating the effectiveness of the implemented models even with a relatively small dataset.

Despite achieving good performance, the system has certain limitations. The quality of generated images is constrained by the limited dataset of 8000 images, which affects the diversity and realism of outputs. The text-to-image module works effectively only for predefined or simple sentences and struggles with complex or unseen inputs. Additionally, the system requires moderate computational time for generating images, which may not be suitable for real-time applications. These limitations highlight the need for larger datasets, better model optimization, and improved architectures to enhance overall performance.

In future work, the system can be improved by incorporating advanced models such as StyleGAN or diffusion-based models for higher-quality image generation. Expanding the dataset to include millions of images can significantly improve accuracy and diversity. The text-to-image module can be enhanced using transformer-based architectures for better semantic understanding. Furthermore, optimizing the system for faster processing and deploying it on cloud platforms can make it more scalable and accessible. These enhancements will enable the system to achieve higher performance and broader real-world applicability.

RE.FERENCES

[1] Ian Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Yoshua Bengio, “Generative Adversarial Networks,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[2] Alec Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” *International Conference on Learning Representations (ICLR)*, 2016.

[3] Sepp Hochreiter and Jürgen Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] Y. Yann LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.

[5] Scott Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative Adversarial Text-to-Image Synthesis," *International Conference on Machine Learning (ICML)*, 2016.

[6] Han Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas, "StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks," *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NeurIPS*, 2012.

[8] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *ICLR*, 2014.

[9] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.