Face Generation using GAN

Sakshi Singh, Neha Lalit, Ameesha Roy, Akanksha Chauhan, Konika Rani, Neha Gautam

Chandigarh University, Mohali, Punjab, India

Corresponding author: Sakshi Singh, Email: sakshisingh.ssg@gmail.com

With applications ranging from entertainment and art to AI-assisted human- computer interaction, the need for realistic face generation has increased dramatically. By employing a multi-modal system that seamlessly blends generative adversarial networks (GANs), natural language processing (NLP), and an attention mechanism, this study offers a novel approach to solving this issue. The FFHQ annotated dataset is used in this work to extract subtle facial traits from user-provided text descriptions using NLP-based feature vectors. An attention technique is used to further improve the images produced by a Conditional GAN, directing the model to concentrate on text-conditioned regions. The method generates diverse, high-quality face images that closely match user-specified criteria by utilizing multi-GAN capabilities. The efficacy of this approach is demonstrated by the experimental results, which are qualitative with user-generated material and quantitative with measures such as Inception Score. This work advances the state of the art in the field of multi-modal face generation by providing a promising path for it.

Keywords: Natural language processing (NLP), vector encoding, Generative adversarial networks (GANs), Machine Learning (ML), Multi-GAN, Conditional GAN, Deep Convolutional Generative Adversarial Networks (DCGAN), Attention Mechanism.

in: *Advancements in Communication and Systems*. Ed. by Ashish Kumar Tripathi and Vivek Shrivastava. Computing and Intelligent Systems, SCRS, India., 2023, pp. 115–120. DOI: https://doi.org/10.56155/978-81-955020-7-3-10

1 Introduction

In artificial intelligence, creating realistic and diversified face images is a fascinating topic with applications ranging from improving human-computer interaction in chatbots and virtual assistants to character design in the gaming industry. While traditional generative models have achieved great progress in this area, the aim of smoothly converting user-provided text descriptions into photorealistic faces is still unattainable.

This research introduces an innovative approach to tackle the intricate problem of face generation by connecting the fields of computer vision, natural language processing (NLP), and deep learning. Our approach employs an annotated dataset from the Flickr-Faces-HQ (FFHQ) collection, which is a vast compilation of human faces with diverse features and expressions. Instead of traditional GANs, our method uses NLP-based feature vectors to enhance the synthesis process and enable users to provide text descriptions to steer the production process.

The incorporation of a Conditional Generative Adversarial Network (GAN) with an attention mechanism is a key component of our approach. Because of this combination of technologies, the model can generate high-resolution face photos while also selectively focusing on particular textual characteristics, producing images that closely match the user's purpose. By highlighting the significance of specific textual parts during the image-generating process, the attention mechanism serves as a guiding hand and improves the quality and interpretability of the generated information.

Our image synthesis method takes into account both high-level and low-level features. By using a multi-GAN system, we refine the generated images and produce realistic and varied faces. This approach provides a strong solution for a variety of applications, including character design and content creation in creative fields.

We have successfully validated the efficiency of our methodology through various evaluations and tests. Qualitative assessments based on user feedback and quantitative metrics like the Inception Score highlight the quality and user satisfaction of the generated faces. We consider this research a significant advancement for generating multi-modal faces, with implications for applications requiring tailored high-quality content.

2 Literature Review

In 2014 [1], Gauthier, Jon proposed a method "Conditional generative adversarial nets for convolutional face generation." Generative adversarial networks (GANs) can be used in a conditional setting as an extension. In the GAN framework, the "generator" network's objective is to deceive the "discriminator" network into believing that its samples are actual data. With the enhancement of each network's ability to condition operations on any external data that provides context for the discriminated or produced image, their effectiveness has been further improved.

Lu, Yongyi, Yu-Wing Tai, and Chi-Keung Tang [2] in 2018, proposed a method for attribute-guided face generation. This method takes a low-resolution face image and an attribute vector extracted from a high-resolution attribute image as inputs. Then, it generates a high-resolution face image that satisfies the given attributes for the low-resolution input image.

In 2018 Shen, Yujun, et al. [3] Researchers have published a new method called "Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis." This method has significantly reduced the training difficulty of GAN by using the identity classifier to extract identity features from both the generator's input (real) and output (synthesized) face images.

In 2021 [4] Liu, Mingcong, et al proposed BlendGAN for a flexible blending strategy, and a self-supervised style encoder is used to generate arbitrary stylized faces from a generic artistic dataset.
In 2022 [5] Wang, Xin, et al. proposed a method "Gan-generated faces detection: A survey and new perspectives." The text concentrates on techniques for identifying faces in produced or synthesized images using GAN models, presenting a thorough analysis of current developments in GAN-face detection.

In 2019, O. R. Nasir, S. K. Jha, M. S. Grover, Yi Yu, A. Kumar and R. R. Shah [6] worked on a related project where the discriminator was given noise and the model was used to flip the labels for real and fake images. The varied textual description was displayed in the generated image.

In 2021, Mohana, D. M. Shariff, A. H, and A. D. [7] provided a model. A generative model that can create high-quality images of human faces at scale was trained using a Convolutional Architecture based on GAN, named Deep Convolutional Generative Adversarial Networks (DCGAN). The DCGAN model was trained on the CelebFaces Attributes Dataset (CelebA) to produce these images.

"Using a conditional GAN for face generation: moving from attribute-labels to faces" [8], Yaohui Wang, Antitza Dantcheva, and Francois Bremond (2018) investigated the inverse problem, which is the creation of attribute-associated faces given attribute labels. Two fields that have shown interest in the topic are entertainment and law enforcement. This study proposes two models that utilize 2D and 3D deep conditional generative adversarial networks (DCGAN) to generate facial images and videos based on attribute labels. Attribute labels are a tool used to identify the specific attributes of the generated images and videos.

3 Methodology

It involves a series of structured steps to guide the research or implementation process. This structured methodology given below guides the process of implementing face generation using GANs, facilitating a systematic approach from data preparation to model training and evaluation, ensuring a thorough and methodical workflow.

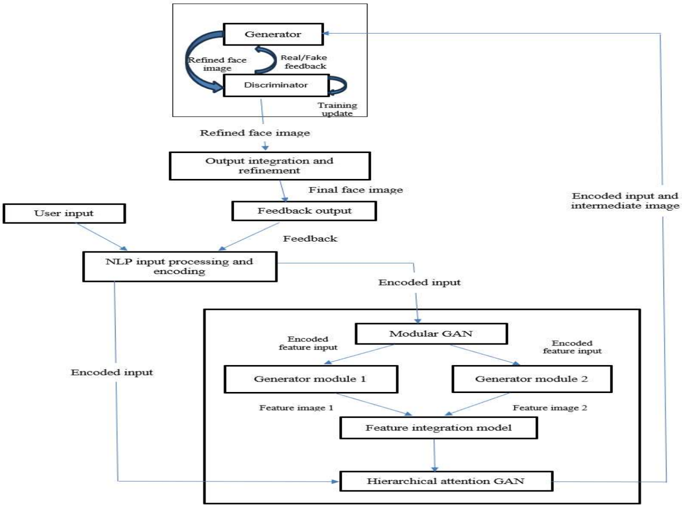

Figure1. Execution of Approach.

3.1 Gathering and preprocessing data

Define the dataset for training, so that the model will be able to understand the further processing steps involved with the dataset. This process includes the application of pixel value normalization and resizing every image to the same resolution.

3.2 Integration of Natural Language Processing (NLP)

Text Descriptions: Permit users to enter written explanations for the chosen faces.
NLPI Tokenize and clean user input as part of preprocessing. Pretrained word embeddings can be used to transform text descriptions into dense vectors that capture semantic data

3.3 Architecture for Conditional GAN

Generator: Create a Conditional GAN architecture that takes two inputs: feature vectors based on natural language processing (c) and random noise (z). This input is mapped to a fake image by the generator.

$$\text{Generated Image}=G(z;\theta G) \tag{1}$$

Discriminator: Put in place a discriminator that assesses an image's veracity based on the accompanying text descriptions.

$$\text{Output}=D(x;\theta D) \tag{2}$$

3.4 Integration of Attention Mechanisms

Soft Attention: To improve the generation process, add a soft attention mechanism to the GAN architecture. Based on the text descriptions, the attention mechanism dynamically chooses which areas of the image to focus on.

Text-Image Alignment: Teach the attention mechanism to match relevant areas in the created images with words in the text descriptions. The GAN can create complex and well-organized images thanks to this alignment.

3.5 Mixed-Modal Instruction

To assemble a training dataset, we need to match text descriptions with actual images. Once we have paired data, we can use it to train the conditional GAN. The discriminator in the GAN will discern between real and synthetic images, while the generator will aim to create images that match the text descriptions as closely as possible.

4 Result

Based on an annotated dataset described in Section 1, the conditional GAN model can be combined with multiple GAN models to generate an output based on user requirements. This technology can create facial images with distinct characteristics, such as gender, race, eye color, hair color, facial hair, and facial expression. The model produces realistic and diverse face images.

The generator has been trained to produce lifelike face images that precisely match the given attributes. Any errors in spelling, grammar, or punctuation have been corrected. Given a noise vector and attribute vectors, it generates a facial image. The generator is trained using the following loss function:

Loss(G) = -E[D(G(z, a))] (3)

where:

- z = noise vector
- a = attribute vector
- D = discriminator

The role of the discriminator is to be trained in distinguishing between real and generated images of faces. Given a face image, the system predicts the likelihood of the image being authentic. To train the discriminator, the following loss function is utilized:

Loss(D) = -E[log(D(f)) + log(1 - D(G(z, a)))] (4)

where:

- f = real face image
- z = noise vector
- a = attribute vector
- G = generator

Figure 2 displays examples of images generated by the model, given a raw input from the user. The model generates a face that corresponds to the given attributes. While the generated images may appear blurry, it is still possible to notice that the model has learned the physical characteristics and appearance of the person, and it is capable of presenting various facial expressions based on the input.

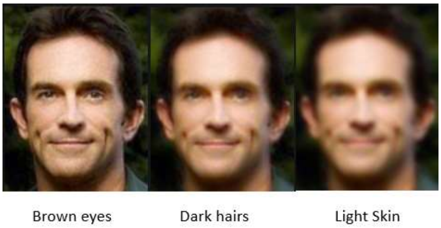

Figure 2 Generated Sample from the given conditions

The model learns to generate synthetic images by training on the dataset of images. Where specific attributes such as light skin color, brown eyes, and dark hair are developed using conditional GAN. The labels or the conditions introduce an additional input, typically referred to as a "condition vector," which encodes information about the attributes we want in the generated image. The generator takes random noise as well as the condition vector as input and generates the synthetic image.

5 Conclusion

A promising new method for producing realistic and varied face images is to use a GAN model for face generation with NLP and conditional GAN. It can be applied to many different things, like making new characters for movies or video games, making lifelike avatars for virtual reality apps, or making customized filters for social media apps.

Compared to other face-generation techniques, this strategy offers many benefits. Initially, it can produce facial images with particular characteristics like age, gender, race, and expression. Secondly, training and implementing it is not too difficult. Thirdly, it can produce excellent images that are hard to tell apart from photographs of real faces.

6 Acknowledgement

We are grateful to our advisor, Professor Neha Gautam, for her guidance and support throughout our research. We also appreciate the feedback and suggestions from my committee member, Professor Konika. I am grateful to Chandigarh University for providing us with the necessary resources and opportunities to carry out our research also like to extend my thanks for their financial support and the Research Computing Center for granting access to the required computing resources.

References

[1] Gauthier, Jon. "Conditional generative adversarial nets for convolutional face generation." Class project for Stanford CS231N: convolutional neural networks for visual recognition, Winter semester 2014.5 (2014): 2.

[2] Lu, Yongyi, Yu-Wing Tai, and Chi-Keung Tang. "Attribute-guided face generation using conditional cyclegan." Proceedings of the European conference on computer vision (ECCV). 2018.

[3] Shen, Yujun, et al. "Faceid-gan: Learning a symmetry three-player gan for identity-preserving face synthesis." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[4] Liu, Mingcong, et al. "Blendgan: Implicitly gan blending for arbitrary stylized face generation." Advances in Neural Information Processing Systems 34 (2021): 29710-29722.

[5] Wang, Xin, et al. "Gan-generated faces detection: A survey and new perspectives." arXiv preprintarXiv:2202.07145 (2022).

[6] Nasir, Osaid Rehman, et al. "Text2facegan: Face generation from fine grained textual descriptions." 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). IEEE, 2019.

[7] Shariff, Daanish Mohammed, H. Abhishek, and D. Akash. "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). IEEE, 2021.

[8] Wang, Yaohui, Antitza Dantcheva, and Francois Bremond. "From attribute-labels to faces: face generation using a conditional generative adversarial network." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.