Generating Synthetic Images Using Generative Adversarial Network

S. Dhene[1], P. Bhosale[2], M. Avhad[3], J. Rawal[4], R. Dalvi[5]

[1,2,3,4,5] *Department of Computer Engineering(MMCOE), SPPU, India*

***Abstract*—Text to image technology with generative[3] adversarial networks (GAN) is a deep learning model technology. For many years due to lack of knowledge of technological resources the police department has to call the sketch artist to get the face of the criminal. The traditional method takes lots of time hence proposed approach uses text as an input of human facial traits and gives almost a similar image of criminal. Artificial intelligence (AI) has a challenge with converting information between text and image (NLP) that links natural language processing and image processing. As a result, we present a new efficient system based on the generative adversarial network (GAN) that will enhance performance by simple procedures. Our work primarily focuses on criminal face generation with minimum time consumption that extracts face traits from text descriptions and creates a realistic human face.**

***Index Terms*—Adversarial, Artificial Intelligence, Criminal, Deep learning, Face generation, Generative, GAN, NLP.**

I. INTRODUCTION

Beyond the model capabilities and evaluations presented above, there are broader issues to consider with large-scale models for text-to-image generation. Some of these issues pertain to the development process itself, including the use of large, mostly[1] uncurated training datasets of images obtained from the web with little oversight, or conceptual vagueness around constructs in the task formulation. Since large text-to-image models are foundation models — enabling both a range of system applications as well as fine tuning for specific image generation tasks — they act as a form of infrastructure which shapes our conceptions of what is both possible and desirable. Predicting all possible uses and consequences of infrastructure is difficult if not impossible, and so responsible AI practices which emphasize transparently documenting and sharing information about datasets and models are crucial.

Although applications are beyond scope of this paper, we discuss here some likely opportunities that can be anticipated Synthetic Image generation can solve the business problem of scarcity or unavailability of image data by creating synthetic image data. Synthetic Image generation can solve the criminal Investigation[5] image retrieval algorithm based on deep learning Text to image generation can be used in conversational chatbots to generate contextual images based on user input. Synthetic images can be employed to train ML models where the real image data doesn't have an important variety. Synthetic images can be generated to add further variation to the being image dataset before training the model. Text to image generation can be used for existing machines to produce images on the go as well as produce synthetic images on the go as well as produce synthetic images when there are limited hunt results.

II PROPOSED METHOD

In our model we are using GAN(Generative adversarial Network) for realistic image generation. For the text embedding we are using word2vec embedder[2].At first the input text will get converted into an embedding vector, this vector will then be used to give input to the GAN network. At the beginning the generator will produce a low-resolution image from the embedding vector then the generated image will be given input to the discriminator. Discriminator is a classification model hence it will help to classify the real and fake image as the real face dataset is given to the discriminator. The model will ensure to produce a high-resolution face image after each epoch the face image will improve and will produce quality image. Stack GAN[2] is the GAN we are choosing in our project to ensure a better quality image.

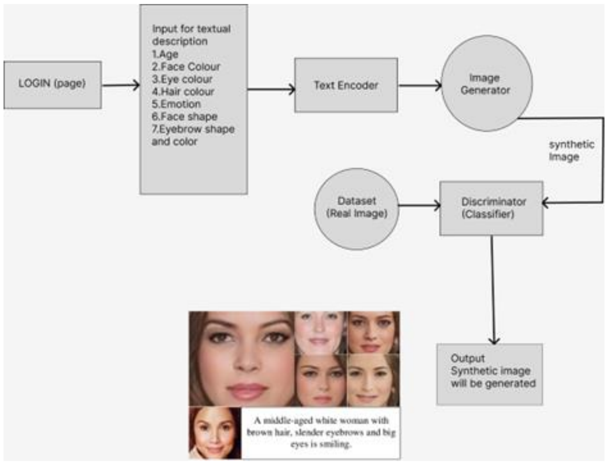

Fig1. Proposed model for image generation

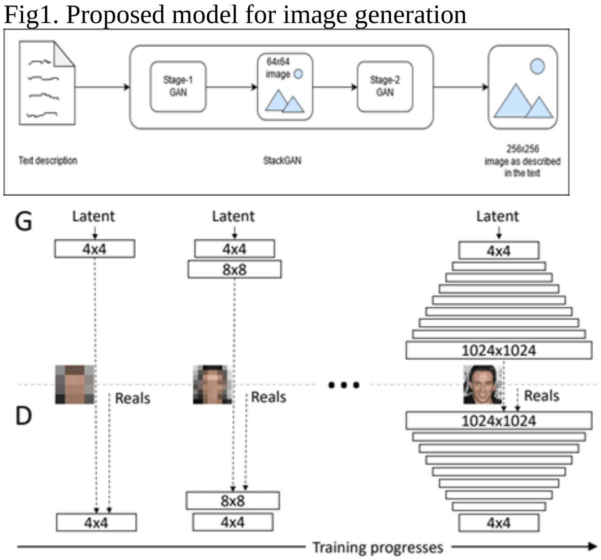

Fig2: Architecture for stack GAN

StackGAN works in two stages[7]:
1.In stage 1, main working will be to produce a low-resolution image that will have rough shape and correct colors of the object. Converted vectors will be given input to the generator that will help in initial image generation.
2.In stage2, the low-resolution image generated in the first image doesn't have vivid object parts and might contain shape deformations[3].
Some details in the text might also be neglected in the first stage, which is vital for generating photo-realistic images. Our Stage-II GAN is built upon Stage-I GAN results to induce high-resolution images. It is adjusted on low-resolution images and also the text embedding again to correct defects in Stage-I results[9]. The Stage-II GAN completes preliminarily ignored text information to produce more realistic details.

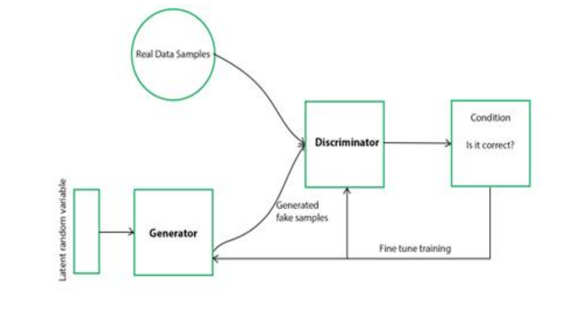

Fig3:Working of GAN

II. IMPLEMENTATION DETAILS

In the StackGAN[2] model, the up-sampling blocks consist of nearest-neighbor upsampling followed by a 3x3 convolution with a stride of 1. Batch normalization and ReLU activation are applied after each convolution, except for the last one. The residual blocks include 3x3 convolutions, batch normalization, and ReLU8 activation. Two residual blocks are used in the 128x128 StackGAN models, while four are used in the 256x256 models. The down-sampling blocks consist of 4x4 convolutions with a stride of 2, batch normalization, and LeakyReLU activation, with the exception of the first one, which does not have batch normalization. The training process involves iteratively training the D0 and G0 of the Stage-I GAN for 600 epochs while fixing the Stage-II GAN. Then, the D and G of the Stage-II GAN are trained for another 600 epochs while fixing the Stage-I GAN. All networks are trained using ADAM solvers with a batch size of 64 and an initial learning rate of 0.0002. The learning rate is halved every 100 epochs.
To further demonstrate the ability of our StackGAN to learn a smooth latent data manifold, we have used it to generate images from linearly interpolated sentence embeddings, as shown in Figure 2. In this demonstration, we fixed the noise vector z so that the generated image is inferred solely from the given text description. The images in the first row were generated using simple sentences that only contain basic color descriptions. These results show that the generated images from interpolated embeddings can accurately reflect color changes and produce plausible faces. In the second row, we have provided samples generated from more complex sentences that contain more detailed information on facial appearances.

Our StackGAN is designed to capture the complex relationship between language and images, rather than simply memorizing the training samples. To accomplish this, we will extract visual features from both the generated images and all of the training images using the Stage-II discriminator of our StackGAN. By analyzing these features, we will be able to identify the nearest neighbors of each generated image in the training set. By visually inspecting these retrieved images, we can see that while the generated images share some similarities with the training samples, they are ultimately distinct.

III. CONCLUSION

In this paper, we have a tendency to propose Stacked Generative Adversarial Networks (StackGAN)[6] with conditioning Augmentation for synthesizing photo-realistic pictures. The planned technique decomposes the text-to-face synthesis to a completely unique sketch-refinement method. Stage-I GAN sketches the article following basic color and form constraints from given text descriptions. Stage-II GAN corrects the defects in Stage-I results and adds additional details, yielding higher resolution pictures with higher image quality. In depth quantitative and qualitative results demonstrate the effectiveness of our planned technique. Compared to existing text-to-image generative models, our technique generates higher resolution pictures (e.g., 256*256) with additional photo-realistic details and variety.

REFERENCE

[1] Muhammad Zeeshan Khan, Saira Jabeen, Muhammad Usman Ghani Khan, Tanzila Saba, Asim Rehmat, Amjad Rehman , Usman Tariq, *“A Realistic Image Generation of Face from Text Description using the Fully Trained Generative Adversarial Networks”* India 2017.

[2] Yikun Wang; Liang Chang; Yuhua Cheng; Lihua Jin; Zhengxin Cheng; Xiaoming Deng; Fuqing Duan”Text2Sketch: Learning Face Sketch from Facial Attribute Text”Athens, Greece 2018

[3] Shailesh Kumar Jha,Manraj Singh Grover,Ajit Kumar, Rajiv Ratn Shah, MIDAS Lab,*”Text2FaceGAN: Face Generation from Fine Grained Textual Descriptions”* IIIT-Delhi Delhi,India 2019

[4] Tianren Wang, Teng Zhang, Brian C. Lovell, “Faces la Carte: Text-to-Face Generation via Attribute Disentanglement” Australia 2020

[5] Chintan Kotian, Samiksha Lokhande, Mohit Jain, Aruna Pavate, *“D2F: Description to Face Synthesis Using GAN”*, India 2020

[6] Rohan Wadhawan; Tanuj Drall; Shubham Singh; Shampa Chakraverty *“Multi-Attributed and Structured Text-to-Face Synthesis”* Bengaluru, India 2020

[7] Ting Zhang; Wen-hong Tian; Ting-ying Zheng; Zu-ning Li; Xue-mei Du; Fan LI, *”Realistic Face Image Generation Based on Generative Adversarial Network”*, Chengdu, China 2020

[8] Ms. Faseela Kathun. C,*” Generating faces From the Sketch Using GAN”*,2021

[9] Xing Qiao; Yanghong Han; Yan Wu; Zili Zhang, *“ Progressive Text-to-Face Synthesis with Generative Adversarial Network”* Jodhpur, India 2021.

[10] Anukriti Kumar; Anurag Mudgil; Nakul Dodeja; Dinesh Kumar Vishwakarma *“Realistic face generation using a textual description”*, Erode, India 2021