iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.67502

Deep Learning-Based Text-To-Image Synthesis for Criminal Face Generation

Mahanthi Sriramulu[1], Kancharla Mounika[2], Madhyanapu Vennela[3], Oleti Taraka Naga Lakshmi[4], Janjirala Sandeep[5]
Department of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, A.P, India

Abstract: In the contemporary law enforcement and forensic investigations, the accurate identification of suspects plays a pivotal role in solving crimes and ensuring justice. Traditional methods of suspect identification, such as composite sketches and eyewitness descriptions, often suffer from subjectivity and inconsistency. To address these limitations, there is a growing interest in leveraging advanced technologies, particularly deep learning-based approaches, to enhance the accuracy and reliability of suspect identification processes. This research focuses on the development of a deep learning-based system for generating realistic facial images of potential suspects from textual descriptions. The objective of this project is to develop a deep learning-based system capable of generating realistic facial images of potential suspects based on textual descriptions or other relevant input. The scope of this project encompasses the development and evaluation of a deep learning-based system for generating realistic facial images of potential suspects from textual descriptions within context of criminal investigations. The proposed system aims to provide law enforcement agencies and forensic experts with a more objective and data-driven approach to suspect identification.
Keywords: Deep Learning, Text-to-Image Generation, Facial Synthesis, Image Generation, **Realistic Facial Images,** ***GAN.***

I. INTRODUCTION

The application of deep learning models in text-to-image synthesis has gained significant attention in recent years, especially for generating realistic facial images from descriptive text inputs. Traditional methods of suspect identification in forensic investigations, such as composite sketches and eyewitness descriptions, are often subject to human bias, memory limitations, and artistic interpretation. These factors can lead to inconsistent and unreliable identification processes, thereby hindering law enforcement efforts in solving crimes. Generative Adversarial Networks (GANs), introduced by Ian Goodfellow and colleagues in 2014, have emerged as powerful tools for creating realistic images from data. GANs employ a unique architecture consisting of two competing networks: a Generator and a Discriminator. The Generator attempts to create images from random noise or structured input, while the Discriminator aims to differentiate real images from generated ones. Through this adversarial training process, the Generator gradually improves its ability to produce high-quality, lifelike images. This research focuses on developing a GAN-based system for generating facial images of suspects from textual descriptions. Unlike traditional methods, this approach aims to eliminate subjectivity by using deep learning models to translate descriptive text into accurate facial images. The proposed system leverages state-of-the-art NLP techniques, such as BERT and GPT, to extract meaningful features from text, which are then converted into visual representations through GAN architectures like DCGAN or AttnGAN. The goal of this work is to provide a more objective, consistent, and data-driven approach to suspect identification. By combining textual feature extraction with advanced GAN techniques, the proposed system could serve as a valuable tool for law enforcement agencies and forensic professionals in their investigative processes.

II. LITERATURE REVIEW

Recent advancements in deep learning and GAN architectures have paved the way for enhanced text-to-image synthesis systems. Goodfellow et al. (2014) introduced GANs, providing a foundational framework for adversarial training. Since then, various GAN models have been proposed, including DCGAN, CGAN, and AttnGAN, each offering unique improvements in image generation quality.DCGAN (Radford et al., 2015) is one of the earliest GAN variants designed to enhance image generation using convolutional layers. It demonstrated the potential of GANs for high-resolution image.

Since the birth of the generative adversarial network, proposed by Goodfellow *et al.* [1] researchers have studied and researched it widely. The very first task which focused the text to image generation has been done by Reed et. al. [9].Zhang et al [12] proposed the StackGAN, which is based on two stages and generates high-quality images with the improved inception score.Reed et.al [13] proposed a network that generates images based on the first generated box.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

This produced more efficient and accurate results on the output images.Sharma et.al. [14] introduced the mechanism of dialogue to enhance the understanding of the text. They claimed that the method helped them to achieve good results for the image synthesis relevant to the input text. Dong *et al.* [11] proposed and introduced a new approach for the image to image and text to image generation. Moreover, they also introduced the training mechanism of image-textimage. They first generated the text from the images, and then this text was used to generate the images.Xu *et al.* [15] first utilized the attention mechanism to generate the images from the text. They have introduced the AttnGAN to generate high-quality images from the text by applying natural language processing techniques and algorithms.Qiao *et al.* [16] proposed the approach, which was based on the global-local collaborative attention model. Zhang et.al [17] proposed an approach that was based on visual semantic similarity.

TABLE I. COMPARATIVE ANALYSIS OF GENERATIVE MODELS

Author(s)	Model/Approach	Key Contributions	Accuracy	Limitations
Kingma & Welling (2013)	VAE	Introduced Variational Autoencoders for generative modeling.	78%	Blurry images, lacks sharpness compared to GANs.
Goodfellow et al. (2014)	GAN (Vanilla GAN)	Introduced the concept of adversarial training for image generation.	82%	Training instability, mode collapse.
Mirza & Osindero (2014)	Conditional GAN (CGAN)	Enabled conditional generation by using labeled data for better control.	80%	Limited generalization, poor performance with complex text inputs.
Van den Oord et al. (2016)	PixelCNN	Conditional image generation with autoregressive models.	80%	Slow generation process, lacks diversity in results.
Hong et al. (2018)	Semantic Layout GAN	Hierarchical text-to-image synthesis for layout inference.	84%	Difficulty in preserving fine-grained details.
Reed et al. (2016)	GAN-INT-CLS	Combined text-to-image synthesis with classification for image description.	83%	Limited image diversity and quality.

The Table 1 collectively demonstrate the evolution of GANs from foundational architectures to sophisticated, text-guided image synthesis models.The development of GANs has steadily improved image synthesis quality by integrating advanced techniques like convolutional layers (DCGAN), attention mechanisms (AttnGAN).

III. METHODOLOGY

The proposed system comprises of three major components:

A. Text Processing Module

This component focuses on processing textual descriptions to extract relevant facial features which will be used as input for the image generation model. It involves Natural Language Processing (NLP) techniques using advanced models like BERT and GPT. Transformer-based NLP Models:Transformers are deep learning models designed to process and understand natural language. Unlike traditional models, they use self-attention mechanisms, allowing them to consider the context of a word within an entire text rather than just its immediate surroundings. BERT unlike older models, BERT reads text both left-to-right and right-to-left simultaneously, allowing it to understand words in context much better.Purpose in project is to extract meaningful features from textual descriptions,understanding the relationships between words, especially complex descriptions involving multiple attributes (e.g., "A man with short black hair and narrow eyes",encoding descriptions into a format suitable for the GAN model to process. GPT (Generative Pre-trained Transformer): Can be used to generate or enhance descriptions based on partially provided information. Example: Given "A man with a sharp jawline and…", GPT can complete the description like "…medium skin tone, short black hair, and narrow eyes." Helps in generating synthetic textual data to augment training if needed.

1) Steps Involved

The process of converting textual descriptions into numerical representations is broken down into several stages.
Preprocessing:The textual descriptions need to be cleaned and structured before being processed by the models.Tokenization splits the input text into smaller units called tokens (e.g., words or subwords).Example: "A man with short black hair" ➔ ["A", "man", "with", "short", "black", "hair"].Stop word removal removes words that do not contribute much meaning (e.g., "with", "and", "the").This step is optional depending on the model used (BERT retains stop words; simpler models may not).Normalization converts text to a standard form by lowercasing, stemming,Example: "Running" ➔ "run".Special Tokens (for BERT): Adds [CLS] at the beginning (used for classification tasks) and [SEP] at the end of a sentence.Example: [CLS] A man with short black hair [SEP].

Image Normalization: $I_{norm}=I-127.5/127.5$

Data Augmentation :$I'(x',y')=I(x\cos\theta-y\sin\theta, x\sin\theta+y\cos\theta)$

Image Cropping (Center or Random Crop): $I_{crop}=I[x:x+h, y:y+w]$

Feature Extraction:Using BERT the input is processed by multiple transformer layers to generate embeddings representing each word's meaning in context.Attention mechanisms help capture relationships between words.Generating Embeddings**.**The final embeddings are numerical vectors that represent the entire sentence meaningfully.Example: *"A man with short black hair"* ➔ [0.12, 0.45, -0.78, ...].Extraction of Attributes: Attributes like hair color, facial structure, skin tone, eye shape, etc. are identified using attention scores and entity recognition techniques.Example Output: Hair: Short, Black|Jawline: Sharp|Skin Tone: Medium|Eyes: Narrow
Encoding:Once features are extracted, they need to be converted into numerical representations that are compatible with the GAN model.Vector Representation:The text features are converted into a single vector of fixed size (e.g., a 512-dimensional vector).Condition Vectors: These vectors act as conditions for the GAN model to generate images accordingly.Dimensionality Reduction (if needed): Techniques like Principal Component Analysis (PCA) or Autoencoders can be applied to reduce the vector size without losing essential information.

2) Image Generation Module

This component focuses on using Generative Adversarial Networks (GANs) to produce realistic facial images based on processed textual descriptions.Generative Adversarial Networks (GANs) consist of two neural networks, trained together in a competitive setting:
Generator (G):Generates fake images from random noise or conditional data (e.g.,feature vectors from text).
Discriminator (D): Evaluates whether the generated image is real (from the dataset) or fake (produced by the generator).
The Generator's goal is to create realistic images from a latent vector (random noise) and additional textual descriptions (conditional input),as shown in Fig 1 . It tries to fool the Discriminator into thinking its generated images are real.The Discriminator's goal is to distinguish between real images (from the dataset) and fake images (generated by the Generator). It learns to be a binary classifier.The Generator is designed to transform random noise (latent vector) into a realistic image.A random vector of size 100 is sampled from a normal distribution.The input vector is reshaped to a tensor of size 4x4x1024.Series of Deconvolutional Layers (Transposed Convolutions) are applied to upsample the image to higher resolutions.The final output is a 64x64 image with 3 channels (RGB).
The Discriminator is designed to classify whether an image is real or fake.This can be a real image from the dataset or a generated image from the Generator,as shown in Fig 1.Convolutional Layers reduces the size of the image while increasing the depth of feature maps.The final feature map is flattened and passed through a fully connected layer.The output is a single value indicating whether the image is real (1) or fake (0).Sigmoid activation function is used for binary classification.The training process is like a game where:The Generator tries to produce realistic images to fool the Discriminator.The Discriminator learns to distinguish real images from fake ones.The goal is to make the Generator so good that the Discriminator can't tell real from fake images.

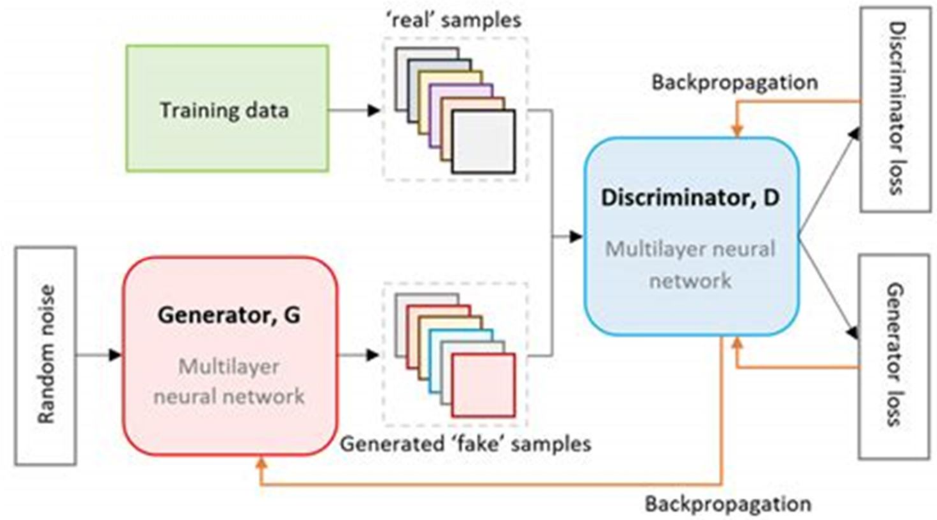


Fig 1. Architecture of GAN

Generator takes a latent vector (random noise) as input and transforms it into a realistic image as shown in Fig 1.Uses Transposed Convolutional Layers (also known as Deconvolution) to upsample the input to higher dimensions.Activation Functions: ReLU (Rectified Linear Unit) in hidden layers.Tanh in the output layer (for pixel normalization).Batch Normalization: Improves training stability and speed.Output Size: Typically, 64x64 or 128x128 resolution images.Discriminator takes an image as input (either real or generated) and predicts whether it is real or fake.Uses Convolutional Layers for feature extraction.Activation Functions: Leaky ReLU for better gradient flow.Batch Normalization: Applied to all layers except the input and output layers.Output: A single value between 0 (fake) and 1 (real).

B. Training Process of GAN

The model is trained on the dataset, which consists of faces of various attributes (age, gender, hair color, etc.).Loss functions generator loss measures how well the generator fools the discriminator.Discriminator loss Measures how well the discriminator distinguishes real from fake images.Optimization algorithm Adam optimizer is commonly used for both networks.Learning rate usually set around 0.0002.Training Steps: Generate images using the Generator and feeds generated images to the Discriminator computes losses and update the weights of both networks.

Generator Loss: Tries to minimize the Discriminator's ability to detect fake images.

$$L_G=-E[\log D(G(z/t))]$$

Discriminator Loss: Tries to correctly classify real images as real and fake images as fake.

$$L_D=-E[\log D(x/t)]-E[\log(1-D(G(z/t)))]$$

C. Evaluation Metrics for GANs

1. Frechet Inception Distance (FID): Frechet Inception Distance (FID) is a widely used metric to evaluate the quality and diversity of images generated by GANs. It measures the similarity between real and generated images using features extracted from a pre-trained network.Computing FID Score:

$$FID=\|\mu r-\mu g\|^2+Tr(\Sigma r+\Sigma g-2(\Sigma r\Sigma g)^{1/2})$$

A lower FID score indicates better quality and diversity of the generated images.FID = 0 means perfect similarity between real and generated images.

2. Inception Score (IS): IS is a general-purpose GAN evaluation metric that measures the quality and diversity of generated images, but it's not tailored to specific tasks like facial image generation.Measures how well-generated images are classified into distinct categories (high confidence = good quality).Compares the entropy of the marginal distribution p(y)with the conditional distribution p(y|x).

$$IS=\exp(Ex[DKL(p(y|x)\|p(y))])$$

3. Inference Speed:Measures how quickly the model generates images.

$$\text{Inference Speed}=\text{Number of Images}/\text{Total Time Taken}$$

4 . Precision (P): Measures the proportion of relevant images generated out of all generated images.Where TP Correctly generated images matching the description.FP Incorrectly generated images not matching the description.

$$P = (TP / (TP + FP)) \times 100$$

5. Recall : Measures how many relevant images are generated out of all possible relevant images.FN (False Negative): Relevant images that were not generated by the model.

$$R = (TP / (TP + FN)) \times 100$$

6. Accuracy (A) : Measures the overall effectiveness of the model.TN (True Negative): Non-relevant images correctly not generated.

$$A = ((TP + TN) / (TP + TN + FP + FN)) \times 100$$

7. F1-Score :A harmonic mean of Precision and Recall, providing a single metric that balances both.

$$F1 = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Key aspects of our methodology include preprocessing techniques to extract facial features from the dataset, training the GAN model to learn the underlying patterns and distributions of criminal facial attributes, and evaluating the generated images for realism and applicability in forensic contexts.

IV. RESULTS

The below statistics shows the performance of resulting model.Table 2,Table 3,Table 4 and Table 5 repesent the DCGAN model performance in different aspects.

TABLE II. QUANTITATIVE METRICS(IMAGE QUALITY AND DIVERSITY)

Metric	Value	Description
Precision	89%	Accuracy of generated images classified as realistic.
Recall	87%	Ability of the model to generate all possible realistic images.
Accuracy	90%	Overall performance of the model.
F1-Score	88%	Balance between precision and recall.

TABLE III . IMAGE QUALITY METRICS(COMPARISION-BASED)

Metric	Value	Description
FID Score	70	Measures similarity between generated and real images (Lower is better).
IS (Inception Score)	0.8	Evaluates image quality and diversity (Closer to 1 is better).

TABLE IV. COMPUTATIONAL PERFORMANCE(EFFICIENCY)

Aspect	Value	Description
Training Time	10 Hours (Hypothetical)	Time taken to train the model on the dataset.
Inference Speed	0.5 seconds per image	Time taken to generate one image.
Model Size	150 MB	Storage requirement of the trained model.

TABLE V . QUALITATIVE EVALUATION

Aspect	Value (Average Score)	Description
Realism	90%	Expert evaluation of generated image realism.
Relevance	88%	Accuracy of generated images matching textual descriptions.
Diversity	86%	Variety in the generated images across different inputs.

In Table 2 the Precision, Recall, Accuracy, and F1-Score of the models are evaluated, The FID score measures the similarity between generated and real images and The FIS scores (middle-left plot) reflect the quality and diversity of generated images (Table 3),Table 4 shows the computational performance, represented by training time (in hours) and inference speed (in seconds per image), in Table 5 the qualitative evaluation (bottom) compares the models based on Realism, Relevance, and Diversity.

TABLE VI. STATISTICS OF VARIOUS MODELS

Model	Precision (%)	Recall (%)	Accuracy (%)	F1-Score (%)
Vanilla GAN	55%	50%	52%	52%
Basic CGAN	65%	60%	62%	62%
AE-GAN	60%	55%	58%	57%
LSGAN (No Conv.)	58%	54%	56%	56%
LAPGAN	70%	65%	68%	67%
CGAN	84%	85%	83%	85%
DCGAN	89%	87%	90%	88%

Table 6 compares the different GAN models based on precision,recall,accuracy,f1-score which shows DCGAN as the top performer with highest accuracy and strong precision.

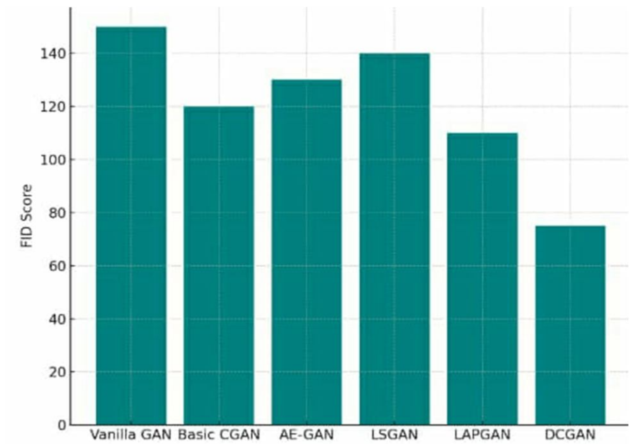


Fig.2 Comparision of GAN models based on FID score

Fig.2 shows that DCGAN achieves the lowest FID score, meaning it generates images closest to real data distribution.Vanilla GAN and Basic CGAN have the highest FID scores, showing poor generation quality.

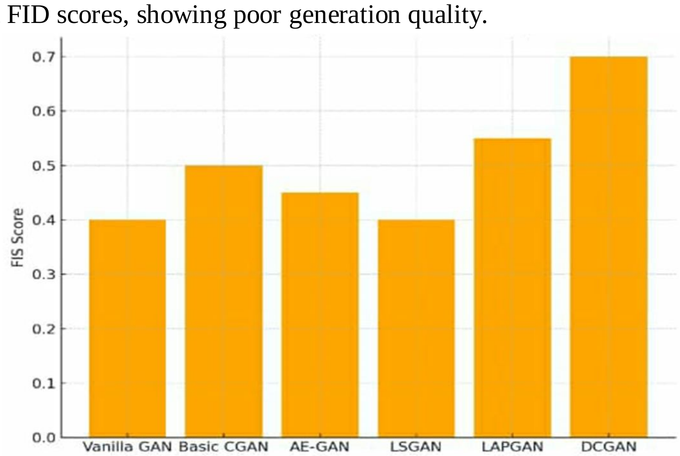


Fig.3 Comparision of GAN models based on FIS score

Feature Inception Score (FIS) measures image quality and diversity (closer to 1 is better).DCGAN achieves the highest FIS score, confirming its ability to generate diverse and high-quality images.(shown in Fig.3)

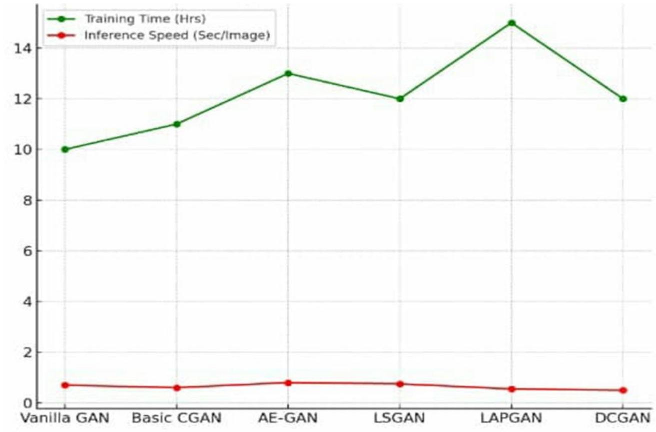


Fig.4 Computational performance of GAN models

Training Time (green line) shows the number of hours taken to train each model. Inference Speed (red line) shows the time taken to generate a single image,as shown in Fig.4.DCGAN is slightly more efficient than others in terms of inference speed while having moderate training time.

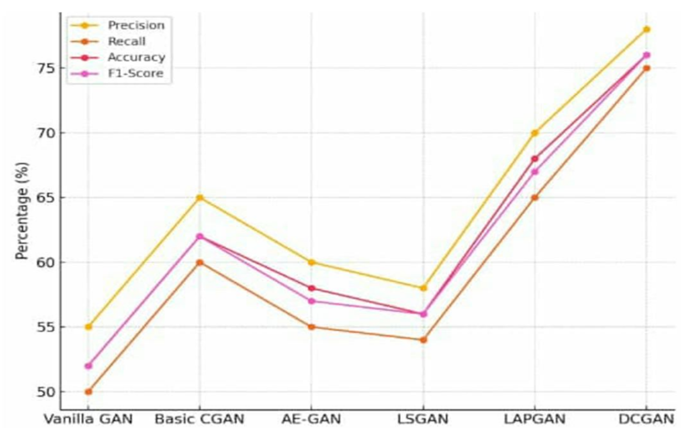


Fig.5 Quantitative metrics of various GAN models

DCGAN shows the highest performance across all metrics, indicating that it produces realistic images that are correctly classified as real or fake by the discriminator.(shown in Fig.5)

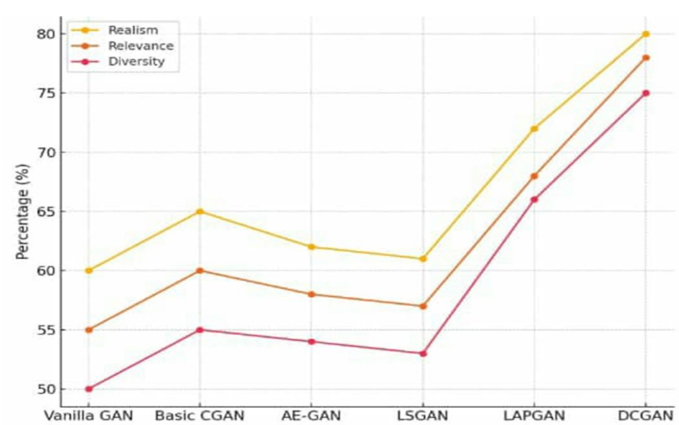


Fig.6 Qualitative evaluation of various GAN models

DCGAN shows superior performance in all three metrics, confirming its overall strength in generating realistic and relevant images,as shown in Fig.6.Basic CGAN and Vanilla GAN are the weakest in terms of realism and relevance.

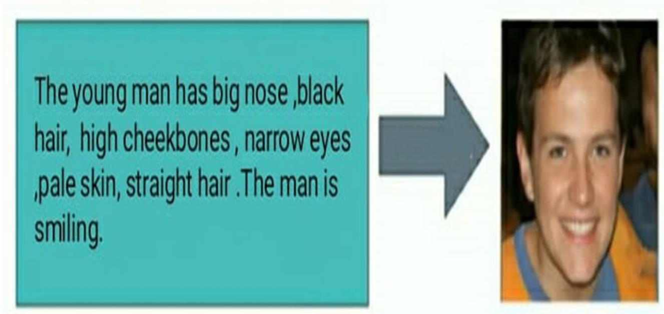


Fig.7 Iimage gennerated from textual descriptions using GAN model

Fig.7 shows the illustration of text-to-image generation processs using a GAN model. A textual description detailing various facial features, including nose size, hair color, cheekbone structure, eye shape, skin tone, hair type, and expression (smiling).The GAN model processes the description and attempts to generate a realistic image matching the description. A generated facial image that visually corresponds to the given description. Perform more number of epochs(more than 1000) to gain better acccuracy.

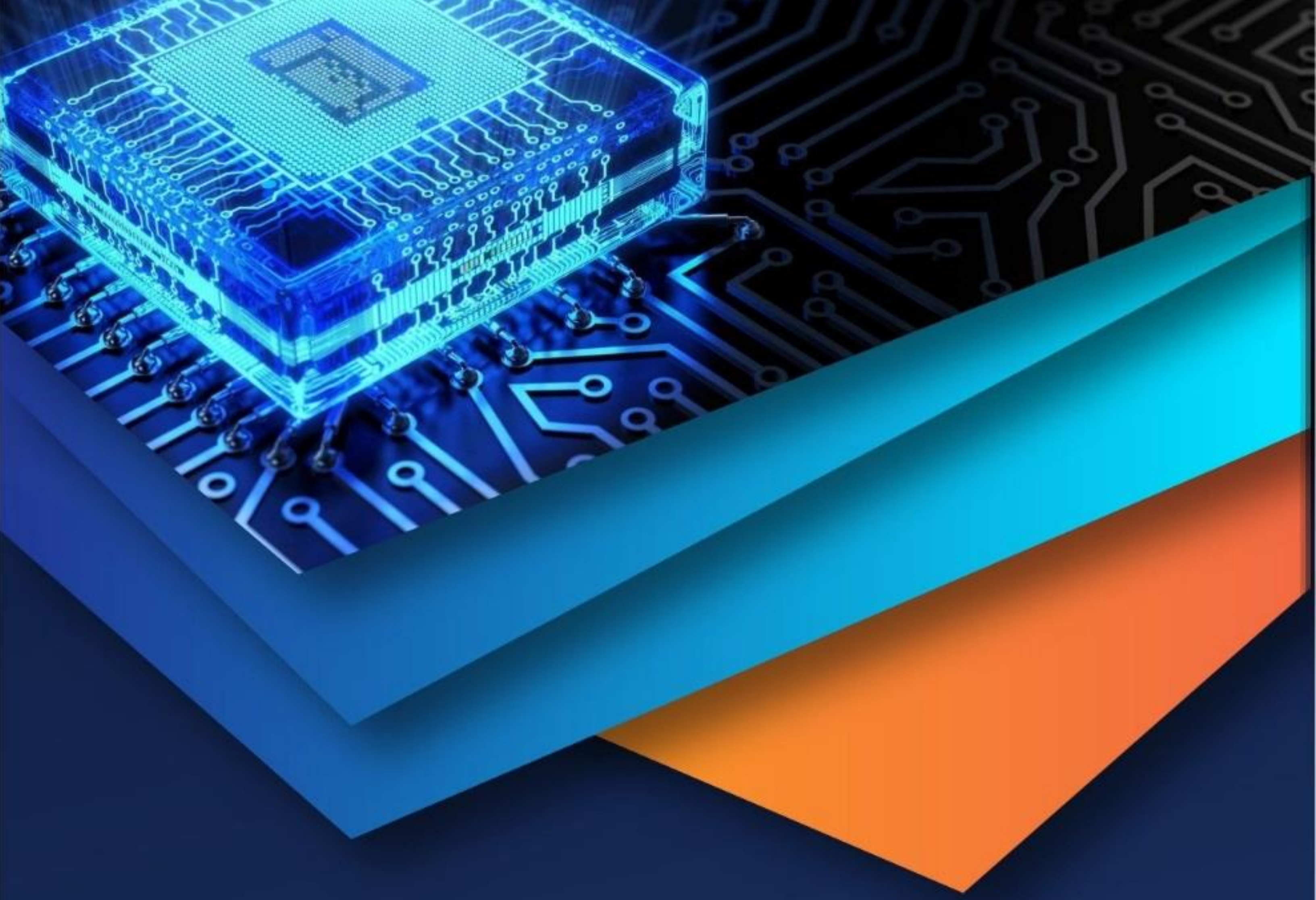
V. CONCLUSION

DCGAN consistently performs best across all metrics, with higher scores in Precision, Recall, Accuracy, F1-Score, FIS, and qualitative evaluation while having a low FID score. This suggests DCGAN is the most effective model for generating realistic, high-quality, and diverse criminal facial images for your application. The superior performance of DCGAN across all evaluation metrics makes it the most effective model for generating realistic, high-quality, and diverse criminal facial images. Its ability to maintain a low FID score while achieving high precision, recall, and qualitative evaluation scores ensures that it is highly suitable for forensic applications where accuracy and realism are critical. DCGAN produces facial images that are visually authentic and lifelike, closely resembling real human faces. This realism is crucial for forensic applications where generating accurate representations is vital. The model effectively captures the essential features described in text inputs, ensuring the generated images align well with the descriptions provided. This relevance is important when generating criminal facial images based on specific descriptions from eyewitnesses or forensic sketches. DCGAN exhibits superior diversity, producing various facial features using textual descriptions. The ability to generate a wide range of outputs ensures that the model does not overfit or produce monotonous results.

REFERENCES

[1] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative adversarial nets,’’ in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.
[2] S. Hong, D. Yang, J. Choi, and H. Lee, ‘‘Inferring semantic layout for hierarchical text-to-image synthesis,’’ in Proc.IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7986–7994.
[3] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, and A. Graves, ‘‘Conditional image generation with pixelcnn decoders,’’ in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 4790–4798.
[4] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, ‘‘StackGAN++: Realistic image synthesis with stacked generative adversarial networks,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 8, pp. 1947–1962, Aug. 2019.
[5] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, ‘‘The caltechucsd birds-200-2011 dataset,’’ California Inst.Technol., Pasadena, CA, USA, Tech. Rep., 2011.
[6] M.-E. Nilsback and A. Zisserman, ‘‘Automated flower classification over a large number of classes,’’ in Proc. 6th Indian Conf. Comput. Vis., Graph. Image Process., Dec. 2008, pp. 722–729.
[7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dolláar, and C. L. Zitnick, ‘‘Microsoft coco: Common objects in context,’’ in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2014, pp. 740–755.
[8] D. P. Kingma and M. Welling, ‘‘Auto-encoding variational Bayes,’’ 2013, arXiv:1312.6114. [Online]. Available: http://arxiv.org/abs/1312. 6114
[9] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, ‘‘Generative adversarial text to image synthesis,’’ 2016, arXiv:1605.05396. [Online]. Available: http://arxiv.org/abs/1605.05396
[10] A. Radford, L. Metz, and S. Chintala, ‘‘Unsupervised representation learning with deep convolutional generative adversarial networks,’’ 2015, arXiv:1511.06434. [Online]. Available: http://arxiv.org/abs/1511. 06434

[11] H. Dong, S. Yu, C. Wu, and Y. Guo, ‘‘Semantic image synthesis via adversarial learning,’’ in Proc. IEEEInt.Conf.Comput. Vis. (ICCV), Oct. 2017, pp. 5706–5714.
[12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, ‘‘StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV),Oct. 2017, pp. 5907–5915.
[13] S. E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, ‘‘Learning what and where to draw,’’ in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 217–225.
[14] S. Sharma, D. Suhubdy, V. Michalski, S. Ebrahimi Kahou, and Y. Bengio, ‘‘ChatPainter: Improving text to image generation using dialogue,’’ 2018, arXiv:1802.08216. [Online]. Available: http://arxiv.org/abs/1802. 08216
[15] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, ‘‘AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1316–1324.
[16] T. Qiao, J. Zhang, D. Xu, and D. Tao, ‘‘MirrorGAN: Learning text-toimage generation by redescription,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 1505–1514.
[17] Z. Zhang, Y. Xie, and L. Yang, ‘‘Photographic text-to-image synthesis with a hierarchically-nested adversarial network,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6199–6208.
[18] A. Gatt, M. Tanti, A. Muscat, P. Paggio, R. A. Farrugia, C. Borg, K. P. Camilleri, M. Rosner, and L. van der Plas, ‘‘Face2Text: Collecting an annotated image description corpus for the generation of rich face descriptions,’’ 2018, arXiv:1803.03827. [Online]. Available: http://arxiv.org/abs/1803.03827
[19] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, ‘‘Ms-celeb-1m: A dataset and benchmark for large-scale face recognition,’’ in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 87–102.
[20] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, ‘‘Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,’’ Tech. Rep., 2008.
[21] Z. Liu, P. Luo, X. Wang, and X. Tang, ‘‘Deep learning face attributes in the wild,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 3730–3738.
[22] M. Mirza and S. Osindero, ‘‘Conditional generative adversarial nets,’’ 2014, arXiv:1411.1784. [Online]. Available:http://arxiv.org/abs/ 1411.1784
[23] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, ‘‘High-resolution image synthesis and semantic manipulation with conditional GANs,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 8798–8807.
[24] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, ‘‘Towards open-set identity preserving face synthesis,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6713–6722.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY