



# Novel Framework for Generating Criminals Images Based on Textual Data Using Identity GANs

Mohamed Fathallah<sup>1,\*</sup>, Mohamed Sakr<sup>2</sup> and Sherif Eletriby<sup>2</sup>

<sup>1</sup>Department of Computer Science, Faculty of Computers and Information, Kafrelsheikh University, Kafrelsheikh, Egypt

<sup>2</sup>Department of Computer Science, Faculty of Computers and Information, Menoufia University, Menoufia, Egypt

\*Corresponding Author: Mohamed Fathallah. Email: mohamed.f.salem777@gmail.com

Received: 19 February 2023; Accepted: 14 April 2023; Published: 09 June 2023

**Abstract:** Text-to-image generation is a vital task in different fields, such as combating crime and terrorism and quickly arresting lawbreakers. For several years, due to a lack of deep learning and machine learning resources, police officials required artists to draw the face of a criminal. Traditional methods of identifying criminals are inefficient and time-consuming. This paper presented a new proposed hybrid model for converting the text into the nearest images, then ranking the produced images according to the available data. The framework contains two main steps: generation of the image using an Identity Generative Adversarial Network (IGAN) and ranking of the images according to the available data using multi-criteria decision-making based on neutrosophic theory. The IGAN has the same architecture as the classical Generative Adversarial Networks (GANs), but with different modifications, such as adding a non-linear identity block, smoothing the standard GAN loss function by using a modified loss function and label smoothing, and using mini-batch training. The model achieves efficient results in Inception Distance (FID) and inception score (IS) when compared with other architectures of GANs for generating images from text. The IGAN achieves 42.16 as FID and 14.96 as IS. When it comes to ranking the generated images using Neutrosophic, the framework also performs well in the case of missing information and missing data.

**Keywords:** GAN; deep learning; text-to-image; identity GAN

## 1 Introduction

Combating crime and terrorism and quickly arresting lawbreakers are among the most important issues that concern the police, governments, and public opinion [1]. Governments and concerned authorities are constantly working to improve their performance and the resources they use to combat crime and quickly apprehend criminals to spread safety throughout countries [2]. In the recent period, some modern technology and new methods have appeared that help countries in this regard, such as machine learning and deep learning [3], which are more effective than the traditional classical methods.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the most important duties in preventing crimes is creating visuals from textual information and other available data, especially when information about criminals is accessible. GAN was initially proposed in 2014 by Goodfellow et al. [4]. The generator and discriminator are the two components that make up the GAN model. The discriminator is used to distinguish between produced images and false images, while the generator is used to create fake images. The primary issue is how to handle the noise and redundant data once the GAN has generated the pictures. Many versions of GANs that have been introduced to generate images from text, such as attention GAN, high-definition GAN, stack GAN, etc. [5,6]. The main challenge for the entire GAN architecture is mode collapse [7].

After examining a sizable sample of generated images, a mode of collapse can be found. Little variety will be evident in the photographs, with repetitions of the same image or a small subset of identical images. Analyzing the line plot of model loss might also help to spot a mode collapse. As the generator model is updated and switches from generating one mode to another with a different loss, the line plot will display oscillations in the loss over time, most noticeably in the generator model. For more realistic individual samples, the mode of collapse is typically compromised. This compromise between mode collapse and high-quality, realistic samples might result in a biased model that portrays an unfair representation of either gender or race. Giving up on preventing the mode collapse in order to obtain a small, high-quality sample has the additional drawback of making the model unstable and labor-intensive to train. Reducing mode collapse affects the overall process of efficiently generating images from text [8,9].

After generating the images based on the textual data using the IGAN, the process of ranking the generated images according to the available information is one of the most important steps. Using multi-criteria decision-making techniques, these characteristics can be used to develop an effective ranking strategy for available images. Numerous methods have been presented in this area, including neutrosophic [10,11]. Many methods have been introduced to rank the alternatives according to different criteria, such as fuzzy and neutrosophic. By taking into account truth, intermediacy, and falsity membership degrees as well as its capacity to discriminate between relative and absolute truth as well as between relative and absolute falsity, the neutrosophic set can overcome the limits of the fuzzy set. Hence, several studies inspired the presentation of multi-criteria decision-making methods in the neutrosophic environment [12,13].

The contributions of this paper can be summarized as follows:

1. The model is considered the first model to arrange the criminals and the wanted, even in the event of conflicting information and incomplete or unclear information, as it relies on the neutrosophic approach in the process of arranging criminals.
2. The paper introduces a promising version of the GAN architecture that can be expanded to generate a high-resolution image with high accuracy compared with other available versions of GANs.
3. The proposed model avoids entanglement by encouraging the generator network to generate samples that are distributed more uniformly over the data distribution rather than clustering around certain points or features using the modified loss function.

The rest of the paper is organized as follows: Section 2 produces the related work about the previous research in image generation from the text. Section 3 methodology describes each stage in the novel framework for generating images using GAN, removing noise using hybrid models, and performing the ranking process using the neutrosophic approach. Section 4 discusses the results of different sections of the framework. Section 5's conclusion introduces the limitation, discussion, and future work.

## 2 Related Work

A deep learning model called “Text-to-Image Generation Using Generative Adversarial Networks” can generate real or nearly real pictures from text descriptions. For many years, police officers had to use a sketch artist to obtain a criminal’s face due to limited technological means. So, the suggested technique employs textual inputs of human face attributes to circumvent this time-consuming procedure and outputs the matching image of a criminal. Text-to-image conversion is difficult for artificial intelligence (AI) that combines natural language processing (NLP) and image processing (IP). To increase performance using simple techniques, this section presents the latest research on converting text to images using different types of GANs [14,15].

By adding a training variable to each generator and discriminator, Mirza et al. [16] introduced a conditional GANs. The conditional GAN was devised to generate the necessary Modified National Institute of Standards and Technology (MNIST) Data. Identify and choose one of the three designs in which tests were fed into the multi-layer perceptron (MLP) network through a single secreted film, producing a combined unseen instance for the originator. Images and tags are utilized similarly for the discriminator and MLP associations.

Reed et al. [17] introduced a modified version of GANs for converting text to images. The approach is taken by embedding the whole phrase, which is obtained after precoding the text cipher, into a discriminator capable of differentiating between the real and regular image text-driven material sets. As a result, the original Text to Image model is a generic variation of Conditional GAN in which the text content and location of the elegant Y label are both variables. Three unexpected pairings are sent to GAN. Real images with matching text content, produced images with matching text content, real images without matching text content, and real images with non-matching text content are all used as discriminators.

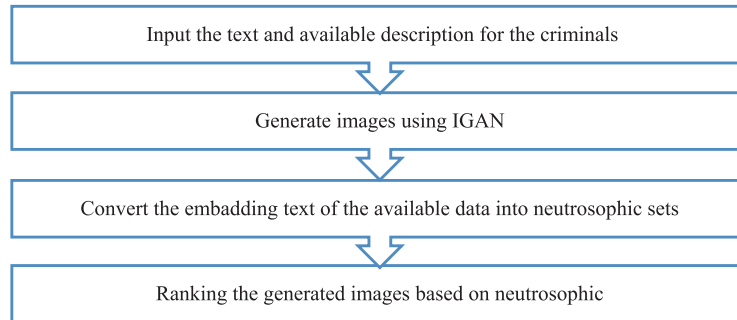
There is a modified version of the GAN called “stacked GANs.” The GAN stacked version can generate images with  $64 \times 64$  pixel. The use of multi-stacked generators has been proposed in several subsequent publications to enable the creation of high-dimensional pictures via text-to-image its w representations. The basic step of Stack GAN uses the provided random noise route and the text processing direction to produce a  $64 \times 64$  pixel coarse picture. A second generator receives this original image and the contained text and produces a  $256 \times 256$  pixel image. The discriminator can distinguish between similar and unrelated pairings of pictures and text in any range. By giving more weight to significant components than to unimportant ones, attention methods enable the network to concentrate on some attention-specific characteristics of an input. By merging into a multi-level refining process, Attention GAN expands the focus of Stacked GAN. The network can synthesize high-quality, fine-grained features of the key phrases derived from the global sentence vectors thanks to the attention processes. The network is urged to choose the phrases that best describe each area of the image.

The Attention GAN (AttnGAN) model generates high-quality images from textual descriptions. While it has numerous advantages, such as the ability to produce diverse images based on textual descriptions and the use of attention processes to focus on different regions of the images, it also has several limitations, such as the fact that it requires significant computational resources to train and generate images and that it is difficult to train as it can suffer from mode collapse [18].

Although there are many different GAN architectures for generating images of criminals from the text, there are some limitations in the previous architectures, such as the fact that all of the architectures cannot deal with mode collapse and don’t contain any methods for ranking the generated images according to the available text.

### 3 Methodology

This section of the paper presents the two different stages of the framework: the first stage is about generating the images based on the text and available data using IGAN, and the second stage is about ranking the generated images based on the available data. Fig. 1 shows the main stages of the proposed framework. The framework begins by reading the textual data and converting it to images using the IGAN, then ranking the generated images based on the degree of the textual data and other available information using the neutrosophic's multi-criteria decision-making framework.



**Figure 1:** Framework of the block diagram

#### 3.1 Faces Dataset

The Faces dataset is a library of face images called Labeled Faces in the Wild that was created to examine the issue of unrestricted face production and recognition. More than 13,000 facial photos gathered from the internet are included in the data collection. The names of the people shown have been written on their faces. 1680 of the individuals shown had two or more different images included in the data set [19]. Fig. 2 shows some of the images in the dataset.

#### 3.2 Generation of Images-Based Text Using IGAN

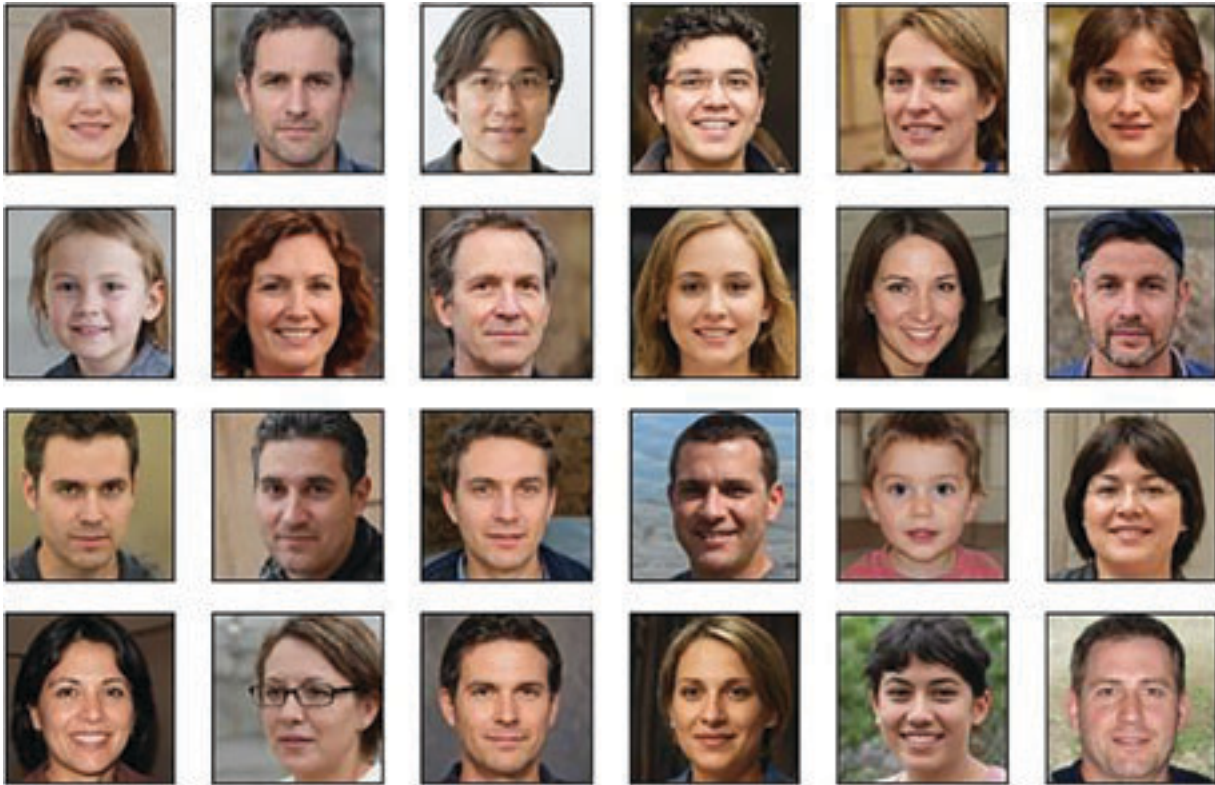
This subsection presents the first main part of the methodology, which involves converting the text to images. This part takes what witnesses say and turns it into clear images that can be used a lot to find criminals and people who are wanted because they are a threat to society. This part presents the IGAN architecture for converting the textual data into images.

The IGAN architecture contains three different parts. The first part is called the text encoder; the second part is called the generator; and the last part is called the discriminator, which is used to distinguish between the generated images and the real images.

##### 3.2.1 Text Encoder

The skip-thought model is an unsupervised encoder-decoder paradigm for encoding sizable portions of text. This method is innovative since it abandons compositional semantics-based approaches while keeping the same level of quality. An input tuple of sentences is used to train this model. At each

time step, the encoder creates a state vector matching the sentence's words. From the current state vector, one of the two decoders predicts the word in the subsequent phrase, while the other decoder predicts the word in the preceding sentence. The objective function, given the encoder representation, is the sum of the log probabilities of the forward and backward phrases.

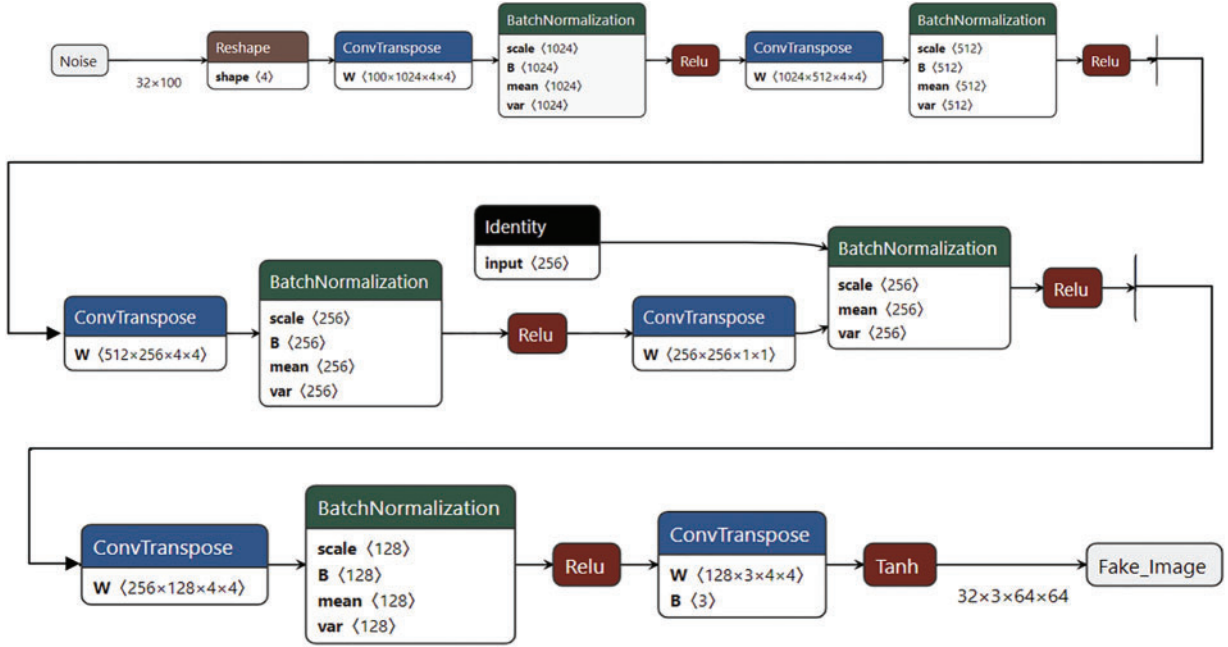


**Figure 2:** Faces dataset images

### 3.2.2 The Generator Architecture

The generator parts are used to generate the images from the textual vector. As shown in Fig. 3, the generator architecture contains six different layers. The six layers depend on six different convolutional transpose functions in each layer. The first layer takes the input vector as 100 dimensions and reproduces 1024 dimensions. Two fully connected layers are used to create a 128-dimensional conditioning latent variable, which is then combined with a 100-dimensional noise latent variable using an epsilon sample from the normal distribution. Map the attached vector into a  $4 \times 4 \times 1024$  tensor using a fully connected layer, then reshape it. A layer of deconvolution using a  $3 \times 3$  filter. An output of  $8 \times 8 \times 512$  is obtained by using a 2-stride length. Leaky ReLU activation with a slope of 0.3 is utilized. Batch normalization has been done on the output. a layer of deconvolution using a  $3 \times 3$  filter. An output of  $16 \times 16 \times 256$  is produced using a 2-stride length. In conjunction with leaky ReLU activation, the slope of 0.3 is used.





**Figure 3:** IGAN generator architecture

### 3.2.3 The Discriminator Architecture

The discriminator part was used to differentiate between the generated images and the real images in the dataset. As shown in Fig. 4, the discriminator contains five different convolution layers. Each convolution layer uses a kernel of size 4. The five layers use a leaky rectifier activation function in each convolution layer. The discriminator uses batch normalization 2d (128, momentum = 0.1). Therefore, the framework modifies the loss function of GAN by adding label smoothing. Instead of providing 1 and 0 labels for real and fake data while training the discriminator, we used softened values of 0.9 and 0.1, respectively. The loss function used in this paper is expressed in Eqs. (1)–(3).

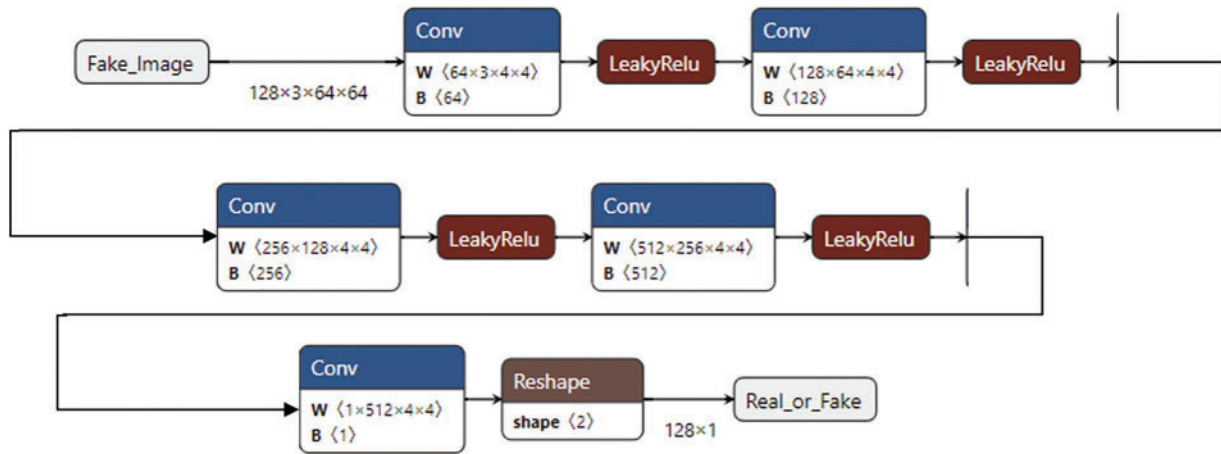
$$L(G, D) = \min(L_G) + \max(L_D) \quad (1)$$

$$L_G = -\frac{1}{m} \sum_{i=1}^m [\log D(G(z^i))] \quad (2)$$

$$L_D = \frac{1}{m} \sum_{i=1}^m [\log D(x) + \log(1 - D(G(z^i)))] \quad (3)$$

### 3.3 Neutrosophic Decision-Making Method

This section presents the second main part of the proposed framework, which involves ranking the generated images using the IGAN based on the available data. The hybrid framework uses multi-criteria based on neutrosophic sets for ranking the criminals according to different characteristics of the available data. The neutrosophic-based method contains three degrees: true degrees, false degrees, and intermediate degrees. A value between 0 and 1 is assigned to each neutrosophic set. The framework uses Table 1 to convert linguistic data such as high, low, medium, etc. into neutrosophic sets.



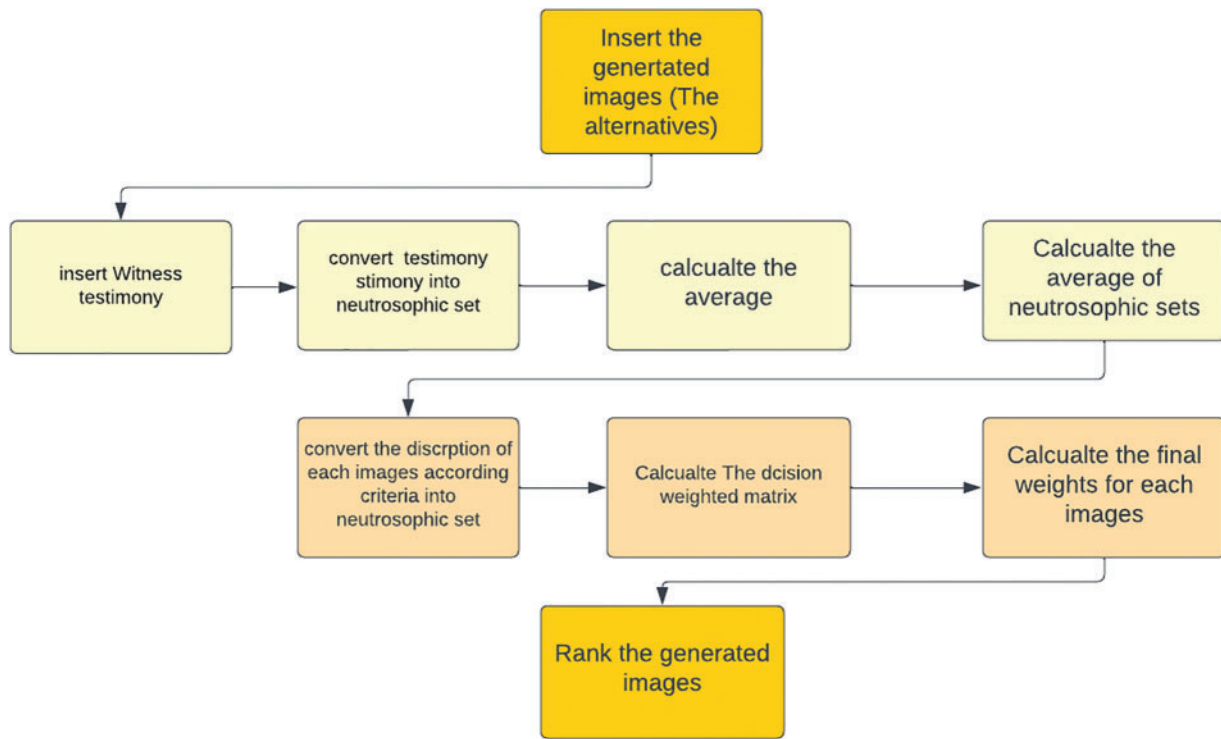
**Figure 4:** IGAN discriminator architecture

**Table 1:** Linguistics terms and neutrosophic set

Linguistic terms	Neutrosophic set
Absolutely low (AL)	(0, 1, 1)
Very low (VL)	(0.2, 0.8, 0.8)
Low (L)	(0.3, 0.7, 0.7)
Medium (M)	(0.5, 0.5, 0.5)
High (H)	(0.7, 0.4, 0.4)
Very high (VH)	(0.8, 0.2, 0.2)
Absolutely high (AH)	(1, 0, 0)

The neutrosophic-based methods build different tables for description based on the available data for the crime and the criminals, both for the generated images and for the real description of the criminal during the crime. The next and most important step is building the decision table, which combines the characteristics of the generated images and the available data during the crime.

The final step in the neutrosophic decision-making algorithm is building the final weights of each degree by calculating the summation of each degree for each alternative and then calculating the final ratio of the true summation to the sum of the accumulative three degrees. Fig. 5 shows the block diagram of the criteria-based neutrosophic for ranking the generated image-based textual data using the IGAN model.



**Figure 5:** Pseudocode of multi-criteria decision-making based on neutrosophic

### 3.4 Hardware and Software Specifications

Table 2 shows the hardware and software specifications for the frameworks, which have been done using the identity GANs and the other versions of GANs for producing images from textual data.

**Table 2:** Hardware and software specification

Device	Description
Processors	Intel(R) Core (TM) i7-10750H CPU @ 2.60 GHz
Random access memory (RAM)	16.0 GB
Graphical processing unit (GPU)	NVIDIA GeForce RTX 3050Ti
Programming language	Python

### 3.5 Evaluation Metrics

The framework uses two different evaluation metrics for comparing the results of the identity GANs and other versions of GANs such as stacked GANs, attention GANs, etc. The framework uses Freshet Inception Distance and Inception Score using Eqs. (4) and (5). The  $r$  in equations refers to the real image, and  $g$  refers to the generated image.

$$IS = \exp \left( E_{x \sim p_g} D_{KL} (p(y|x) || p(y)) \right) \tag{4}$$



$$FID = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}). \quad (5)$$

## 4 Results and Discussion

Section 4 contains two subsections. The first subsection shows the results of generating images from the text and the comparison between the results of the identity GANs and the other versions of GANs. The second subsection introduces the results of multiple criteria for ranking the generated images based on the available textual data.

### 4.1 Results of Generation Step

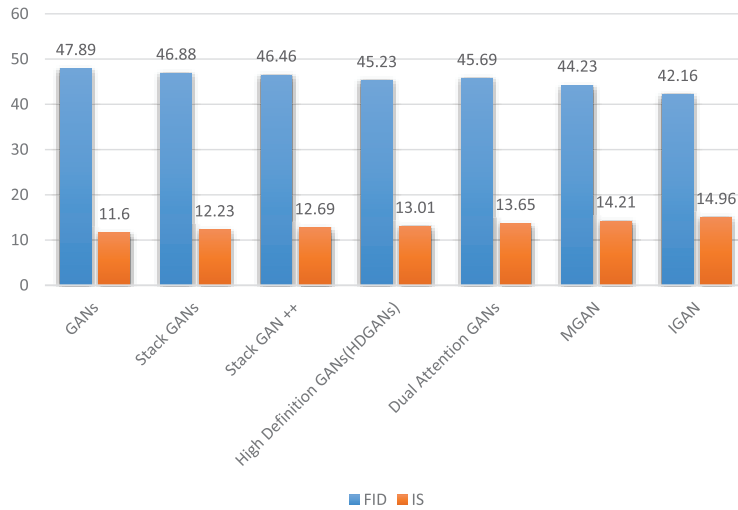
This sub-section shows the results of generating images from the identity GANs and also shows the comparison among the model and other GANs models for generating images from text. The IGAN generates an image for the text “blonde man, near brown hair, blue-gray eye, or green, and the face is near fat”.

Table 3 shows the comparison between IGAN and the other GAN architectures based on the values of the FID and IS. The results in the Table 3 show the efficiency of the modified model against the other architectures of the GANs based on the smaller value of the FID and the largest value of the IS. The results of the table after 200 epochs with Adam as an optimizer and a learning rate of 0.0001.

**Table 3:** Comparison between different models in images generation from text

Model	FID	IS
GANs [20]	47.89	11.6
Stack GANs [21]	46.88	12.23
Stack GAN ++ [22]	46.46	12.69
High definition GANs (HDGANs) [23]	45.23	13.01
Dual attention GANs [24]	45.69	13.65
Modified GAN (MGAN) [25]	44.23	14.21
IGAN	<b>42.16</b>	<b>14.96</b>

Fig. 6 shows the comparison charts between different architectures of GANs for generating images from textual data. Fig. 7 also shows the effectiveness of the modified model compared to other models in generating images, which makes the modified model more effective than other models as the used model generates more accurate images than other models, and this greatly helps the concerned authorities identify the criminals and outlaws. Fig. 7 shows the four generated images using IGAN after 200 epochs.



**Figure 6:** A comparison chart of generating images from the text between GAN architectures



**Figure 7:** Generated images from text using IGANs

#### 4.2 Results of Dynamic Ranking-Based Neutrosophic

This section shows the results of using multi-criteria decision-making based on the neutrosophic approach for ranking the generated images of the criminals using the IGAN according to the available

textual data. This subsection ranks the four nearest generated images according to different chosen criteria and the degree of importance for each criterion of the available data.

Table 4 shows the criteria and functional requirements in linguistic terms for the three witness testimonies before converting them into neutrosophic sets based on Table 1, which describes the method of converting the linguistic terms into neutrosophic sets. This table presents the testimonies of different people after their conversion into a degree of relation using a linguistic term.

**Table 4:** Criteria and functional requirements according to Witness testimony

Criteria	C1	C2	C3	C4
	L	H	VH	AH
Functional requirement	H	L	VH	AH
	VH	VH	VH	AH

Table 5 shows the witness testimony in the term “neutrosophic set” based on Table 1. The final row item in Table 5 shows the averages of the three witnesses based on the neutrosophic values of the three previous rows in the same table. This row is a merger of the importance ratings of the previous three rows into one row for use in the next tables. It greatly expresses the absolute importance of each criterion. Table 6 shows the degree ratio of each criterion for each generated image in linguistic terms and Table 7 shows the same data as Table 6, but with neutrosophic sets. Table 8 shows the final weighted decision matrix after multiplying the data in Table 5 with Table 7. Table 9 shows the final weight for each generated image, the percentage degree, and the final rank for each generated image.

**Table 5:** Criteria and functional requirements according to Witness testimony in neutrosophic

Criteria	C1	C2	C3	C4
	(0.3, 0.7, 0.7)	(0.7, 0.4, 0.4)	(0.8, 0.2, 0.2)	(1, 0, 0)
Functional requirement	(0.7, 0.4, 0.4)	(0.3, 0.7, 0.7)	(0.8, 0.2, 0.2)	(1, 0, 0)
	(0.8, 0.2, 0.2)	(0.8, 0.2, 0.2)	(0.8, 0.2, 0.2)	(1, 0, 0)
Average weighted value	(0.6, 13/30, 13/30)	(0.6, 13/30, 13/30)	(0.8, 0.2, 0.2)	(1, 0, 0)

**Table 6:** Generated images description in linguistic terms

Generated images	C1	C2	C3	C4
1	L	AL	AH	M
2	L	VH	M	VH
3	AH	AH	M	L
4	VH	AL	AL	AH

**Table 7:** Generated images description in neutrosophic

Generated images	C1	C2	C3	C4
1	(0.3, 0.7, 0.7)	(0, 1, 1)	(1, 0, 0)	(0.5, 0.5, 0.5)
2	(0.3, 0.7, 0.7)	(0.8, 0.2, 0.2)	(0.5, 0.5, 0.5)	(0.8, 0.2, 0.2)
3	(1, 0, 0)	(1, 0, 0)	(0.5, 0.5, 0.5)	(0.3, 0.7, 0.7)
4	(0.8, 0.2, 0.2)	(0, 1, 1)	(0, 1, 1)	(1, 0, 0)

**Table 8:** Decision-weighted matrix

Generated images	C1	C2	C3	C4
1	(0.18, 0.303333, 0.303333)	(0, 0.433333, 0.433333)	(0.8, 0, 0)	(0.5, 0, 0)
2	(0.18, 0.303333, 0.303333)	(0.48, 0.086667, 0.086667)	(0.4, 0.1, 0.1)	(0.8, 0, 0)
3	(0.18, 0, 0)	(0.6, 0, 0)	(0.4, 0.1, 0.1)	(0.3, 0, 0)
4	(0.48, 0.086667, 0.086667)	(0, 0.433333, 0.433333)	(0, 0.2, 0.2)	(1, 0, 0)

**Table 9:** Ranking table

Generated images	$\sum T$	$\sum I$	$\sum F$	Ratio	Rank
1	0.68	0.74	0.74	47%	4
2	1.9	0.48	0.5	79%	2
3	1.6	0.1	0.1	94%	1
4	1.48	0.7	0.7	67%	3

## 5 Conclusion and Future Work

Combating crime and terrorism and quickly arresting lawbreakers are among the most important issues that concern the police, governments, and public opinion. This paper has presented a new hybrid model based on IGAN and neutrosophics for generating high-quality images of wanted criminals according to the available data. The model was presented in two main stages: the first stage used IGAN to generate images based on the textual data, and the second stage used neutrosophic to rank the generated criminals. The framework has achieved efficient IS and FID values when compared with other architectures of the GAN model and has also achieved good computational time in the ranking process using the neutrosophic algorithm. Although the model achieved good results in removing noise and generating images, it still has some limitations in terms of the quality of the images and the number of criteria used to rank the criminals. We intend to improve image quality in the future by using transfer learning and various modified models, and we will also base the ranking process on a large number of criteria.

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] P. Mahesha, K. Royina, S. Lal, Y. Anoop and M. P. Thrupthi, "Crime scene analysis using deep learning," in *Proc. Int. Conf. on Signal Processing, Computing and Control (ISPCC 2021)*, Solan, India, pp. 760–764, 2021.
- [2] Y. Chu, X. Yue, L. Yu, M. Sergei and Z. Wang, "Automatic image captioning based on ResNet50 and LSTM with soft attention," *Wireless Communication Mobile Journal*, vol. 20, no. 1, pp. 1–7, 2020.
- [3] L. Ying, F. Jiulun and L. Zong, "Case study on content-based image retrieval for crime scene investigation image database," *Journal of Xi'an University of Posts and Telecommunications*, vol. 20, no. 3, pp. 11–20, 2015.
- [4] I. Goodfellow, J. Pouget, M. Mirza, B. Xu, D. Warde *et al.*, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, no. 1, pp. 2672–2680, 2014.
- [5] K. Safa and S. Chinchu, "Generating text to realistic image using generative adversarial network," in *Proc. Int. Conf. on Advances in Computing and Communications (ICACC 2021)*, Kochi, Kakkannad, India, pp. 1–6, 2021.
- [6] A. Patil and A. Venkatesh, "DCGAN: Deep convolutional GAN with attention module for remote view classification," in *Proc. Int. Conf. on Forensics, Analytics, Big Data, Security (FABS2022)*, Bengaluru, India, pp. 1–10, 2022.
- [7] J. Bian, X. Hui, S. Sun, X. Zhao and M. Tan, "A novel and efficient CVAE-GAN-based approach with informative manifold for semi-supervised anomaly detection," *IEEE Access*, vol. 7, pp. 88903–88916, 2019.
- [8] M. Abedi, L. Hempel, S. Sadeghi and T. Kirsten, "GAN based approaches for generating structured data in the medical domain," *Applied Sciences*, vol. 12, no. 14, pp. 1–16, 2022.
- [9] D. Bang and H. Shim, "MGGAN: Solving mode collapse using manifold-guided training," in *Proc. Int. Conf. on Computer Vision Workshops (ICCVW2021)*, Montreal, BC, Canada, pp. 2347–2356, 2021.
- [10] M. Abobala, "A study of AH-substructures in refined neutrosophic vector spaces," *International Journal of Neutrosophic Science*, vol. 9, no. 1, pp. 74–85, 2020.
- [11] M. Dhar, S. Broumi and F. Smarandache, "A note on square neutrosophic fuzzy matrices," *Neutrosophic Sets and Systems*, vol. 3, no. 1, pp. 37–41, 2014.
- [12] A. Mezirani, A. Bourouis and M. S. Chebout, "Neutrosophic data analytic hierarchy process for multi criteria decision making: Applied to supply chain risk management," *IEEE Access*, vol. 1, no. 2, pp. 1–6, 2022.
- [13] R. Gualdrón, F. Smarandache and C. Bohorquez, "Neutrosophic probabilistic expert system for decision-making support in supply chain risk management," *Neutrosophic Operational Research*, vol. 22, no. 2, pp. 343–366, 2021.
- [14] H. Lee, M. Ra and W. Kim, "Nighttime data augmentation using GAN for improving blind-spot detection," *IEEE Access*, vol. 8, pp. 48049–48059, 2020.
- [15] B. Yang, F. Feng and X. Wang, "GR-GAN: Gradual refinement text to image generation," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME2022)*, Taipei, Taiwan, pp. 1–6, 2022.
- [16] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv: 1411.1784, 2014.
- [17] S. Reed, Z. Akata, X. Yan, L. Logeswaran and H. Lee, "Generative adversarial text-to-image synthesis," *IEEE Access*, vol. 22, no. 1, pp. 1060–1069, 2016.
- [18] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft and S. Belongie, "Stacked generative adversarial networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR2017)*, Honolulu, HI, USA, pp. 1866–1875, 2017.
- [19] B. Bamba, "Faces lite dataset," 2021. [Online]. Available: <https://www.kaggle.com/code/theblackmamba31/generating-fake-faces-using-gan/notebook>

- [20] L. Hanne, R. Kundana, R. Thirukkumaran, Y. V. Parvatikar and K. Madhura, "Text to image synthesis using GANs," in *Proc. Int. Conf. on Advances in Computing, Communication and Applied Informatics (ACCAI 2022)*, Chennai, India, pp. 1–7, 2022.
- [21] H. Zhang, "StackGAN: Text to photo realistic image synthesis with stacked generative adversarial networks," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV 2017)*, Venice, Italy, pp. 5908–5916, 2017.
- [22] H. Zhang, T. Xu, S. Zhang, X. Wang, X. Huang *et al.*, "Stack GAN++: Realistic image synthesis with stacked generative adversarial networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1947–1962, 2019.
- [23] D. Parekh, A. Maiti and V. Jain, "Image super resolution using GAN-A study," in *Proc. Int. Conf. on Trends in Electronics and Informatics (ICOEI 2022)*, Tirunelveli, India, pp. 1539–1549, 2022.
- [24] Y. Cai, "Dualattn-GAN: Text to image synthesis with dual attentional generative adversarial network," *IEEE Access*, vol. 7, no. 2, pp. 183706–183716, 2019.
- [25] Z. Zhang, L. Zhang, X. Chen and Y. Xu, "Modified generative adversarial network for super resolution of terahertz image," in *Proc. Int. Conf. on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD 2020)*, Xi'an, China, pp. 602–605, 2020.