Journal of Information Systems and Informatics
Vol. 6, No. 4, December 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i4.900 Published By DRPM-UBD

Advanced 3D Artistic Image Generation with VAE-SDFCycleGAN

Dorcas Oladayo Esan[1], Pius Adewale Owolawi[2], Chunling Tu[3]

[1,2,3]Tshwane University of Technology South Africa
[1]oladayojadesola10@gmail.com, [2]owolawi@tut.ac.za, [3]du@tut.ac.za

Abstract

Generation of a 3-dimensional (3D)-based artistic image from a 2-dimensional (2D) image using a generative adversarial network (GAN) framework is challenging. Most existing artistic GAN-based frameworks lack robust algorithms lack suitable 3D data representations that can fit into GAN to produce high-quality 3D artistic images. To produce 3D artistic images from 2D image that considerably improves scalability and visual quality, this research integrates innovative variational autoencoder signed distance function, cycle generative adversarial network (VAE-SDFCycleGAN). The proposed method feeds a single 2D image into the network to produce a mesh-based 3D shape. The network encodes a 2D image of the 3D object into latent representations, and implicit surface representations of 3D images corresponding to those of 2D images are subsequently generated. VAE extracts feature from the two-dimensional input image and reconstructs a voxel-type grid using a signed distance function. Cycle GAN produces improved and high-quality 3D artistic images from 2D images. The publicly available COCO dataset was used to evaluate the proposed advanced 3D-VAE-SDFCycleGAN. The model produced a peak signal noise ratio (PSNR) of 31.35, mean square error (MSE) of 65.32, and structural similarity index measure (SSIM) of 0.772 which indicates the improved quality of the generated images. The results are compared with other traditional GAN methods and the results obtained show that the proposed method outperforms the others in terms of quantitative and qualitative evaluation metrics.

Keywords: advanced 3D image, VAE-SDFCycle GAN, artistic image, signed distance function.

1. INTRODUCTION

Art has always been a vital part of human civilization and is a manifestation of human creativity [1]. Man's ability to express himself artistically has enabled us to understand our historical background and track our progress over time. Humans have emphasized creating for centuries to communicate their ideas, imaginations, memories, and thoughts. People can view, interact with, and engage with powerful artworks that can elicit cultural empathy by critically examining historical and contemporary societal problems through historical-artistic images.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Generative adversarial networks, or GANs, have made a significant contribution to the creation of artistic images due to their ability to learn deep representations without requiring a large amount of training data [2]. However, in the current GANs technology, the dataset used to train the model often contains images of two-dimensional (2D) image structure which leads to artistic image view geometry constraints and consequently, fails to generate accurate multi-view image consistency with high resolution and shape quality due to problem of entangled viewpoint and content geometric ambiguity.

Several works have attempted to address the issue of generating three-dimensional artistic images from two-dimensional images such as dimensional augmenter GAN (DiAGAN)[3], X-dimensional GAN(XDGAN) [4], geometry-aware 3D GAN [5], 3D deep convolution generative adversarial network (DCGAN) [6] etc. For instance, the research in [3], [7], [8], [9] and [10] represents 2D images only a projection of the 3D world, with information on depth and volume effectively compressed or lost. Generative 3D methods can map latent spatial codes onto a three-dimensional volumetric scene to "remix" the scene and solve image rendering coherence issues. During training, a combination of 2D and 3D generative adversarial network (GAN) losses derived from differentiable volume traces are applied at multiple scales to improve realism in both 2D scenes and 3D structures, the model failed to reproduce the realistic and high-quality advanced 3D image.

To solve the problem of unrealistic and imperfect three-dimensional image generation from two-dimensional artistic images, this study proposes to integrate signed distance function, variational autoencoder, and cycle GAN (VAE-SDFCYCLEGAN) to address this challenge. The proposed method feeds a single 2D image into the network to produce a mesh-based 3D shape. The network encodes a 2D image of the 3D object into latent representations, and implicit surface representations of 3D images corresponding to those of 2D images are subsequently generated. VAE extracts feature from the two-dimensional input image and reconstructs a voxel-type grid using a signed distance function. CycleGAN then produces improved and high-quality 3D artistic images from 2D images. The main contributions of this research are as follows:

1) Generation of a three-dimensional image from a two-dimensional image: learn encoding, enhancement, and the creation of high-quality three-dimensional images from corresponding two-dimensional images with proposed VAE-SDFCYCLEGAN.
2) Evaluation Metrics: performance evaluation in terms of qualitative and quantitative metrics to compare the proposed model's image generation with other existing artistic image GANs techniques.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3) Comparative Analysis: comparing and benchmarking the performance of the proposed model with other artistic GAN methods on both the COCO Africa mask dataset and the publicly available CelebFace dataset.

The structure of this paper is as follows: related works are given in Section II. The proposed method is covered in Section III, while Section IV describes the experiment's findings and Section V contains the conclusion of the paper.

2. METHODS

This section describes the proposed method as shown in the system architecture in Figure 1. The proposed system architecture is divided into four steps: (a) image acquisition stage, (b) image preprocessing step, and (c) image extraction stage image. and (e) image generation stage. The proposed method is further detailed in the following sections.

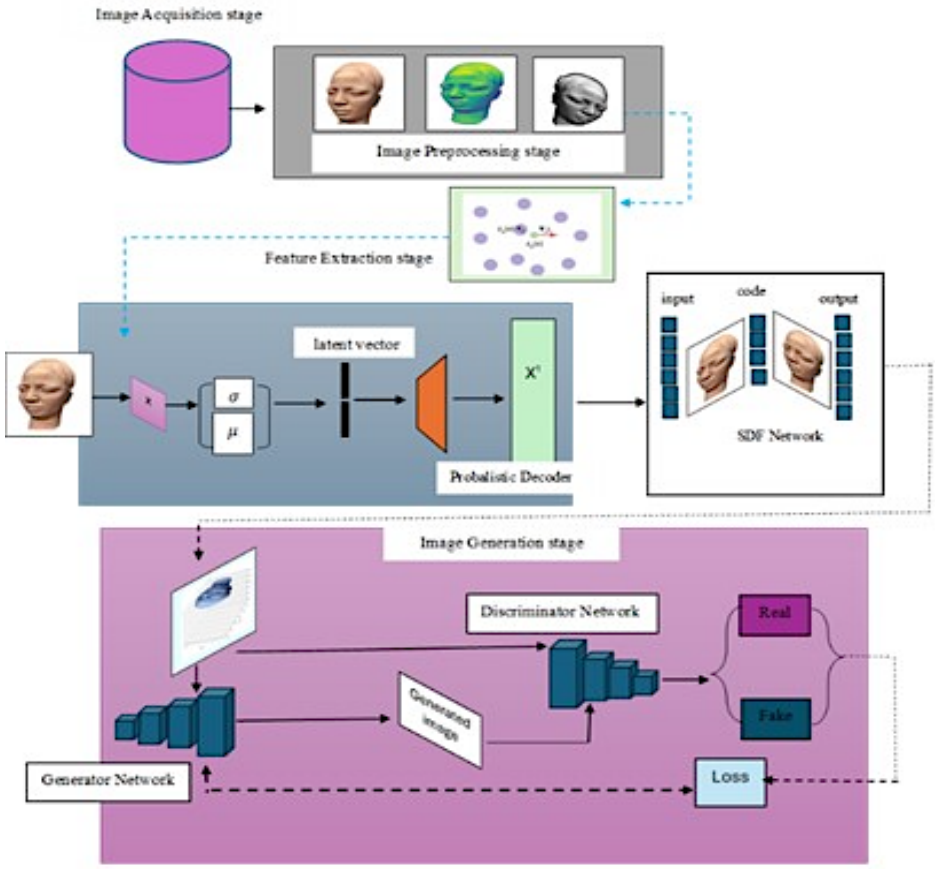

Figure 1. Proposed VAE-SDFCycleGAN model for artistic image generation.

2.1. Image acquisition stage

The experiments in this study used COCO Africa Mask art images [11]. The dataset is images of African masks that will help you experience the pinnacle of African art. This dataset's samples are shown in Figure 2.

Figure 2. Coco Africa mask dataset [11].

2.2. Image pre-processing stage

The image preprocessing step is an important step in computer vision to remove unnecessary noise and reduce the computational power of Cycle GAN. The image preprocessing used in this study is normalization and noise removal. The resulting image is sent to image normalization for image preprocessing. Here, all images are rescaled to a standard image range (typically between 0 and 1) to facilitate processing by machine learning algorithms as shown in Equation 1.

$$I^1 = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

Where, max(x) and min(x) represent the maximum and minimum values of the features, respectively, and x is the data point within the features.
Noise removal The image normalization output is fed into the denoising input. Here, unwanted noise and artefacts are removed from the image, improving image quality and model performance [12]. In this study, we replace the pixels of the

noisy image with the average value of the neighbouring pixels (mask) and use median filtering to remove the noise, sorted by grey value as shown in Equation 2.

$$I'(x',y')=median\{g'(x'+i),(y'+j),\ i,j\in w'\} \quad 2)$$

The input is represented by $g'(x',y')$, the output is $I'(x',y')$, and the 2-D image mask is *w*. The feature extraction step receives the output of the improved image for additional processing.

2.3. 3D point image feature extraction stage

The output of the normalized image is fed as input to the VAE-SDFCycle GAN to train the VAE encoder function [13]. The input image is passed through several layers in VAE to reduce the image size and obtain the z-compressed hidden vector. The encoder extracts the mean and standard deviation of each latent variable and then feeds them to the decoder to reconstruct the input image. The output of the trained image is sent to the SDF, where the 3D spatial points $P \in R^3$ are extracted using point distances k(p)∈R that are closer to the surface points [14],[15]. Here, the P value can be positive (+) or negative (-) for the object. The sign of point x concerning the boundary $\delta\Omega$ is defined as shown in Equation 3.

$$sign(x,\delta\Omega)=\begin{cases} 1 & if\ x\in\Omega \\ 0 & if\ x\in\delta\Omega \\ -1 & if\ x\in\Omega \end{cases} \quad (3)$$

Where the signed distance function is given as shown in Equation 4.

$$k(x)=dist(x,\delta\Omega)\cdot sign(x,\delta\Omega) \quad (4)$$

Where the boundary is indicated as δΩ, *x* is a point in 3D Euclidean space. The output of the three-dimensional point extracted from the image is fed as input to the CycleGAN to generate reconstructed artistic images in the next stage.

2.4. Image generation stage

For creating images, the 3D pixel vector output is fed into the GAN cycle stage. Generator and discriminator are the two components that make up CycleGAN. The encoder and the decoder comprise the bulk of the generator's architecture. To extract high-dimensional features and minimize the size of the feature map, the encoder is composed of three convolutional layers and nine residual blocks. To create a pseudo-image the same size as the input image, the decoder is composed of two transposed convolutional layers and an output layer. To determine whether an image is real or fake, a discriminator is trained. Our research employs 70x70

PatchGAN. Its purpose is to act as a discriminator in the 70 x 70 format to separate the real from the fake overlapping image patches. A complete convolutional network with five convolutional layers makes up the discriminator.

The feature maps' size is preserved by setting the pitches of the other two convolutional layers to 1. A single-channel prediction map with values between 0 and 1 at each pixel location was produced when the final output layer's number of filters was set to 1. In the discriminator's convolutional layer, instance normalization is applied along with Leaky ReLU. An input image of size 256 x 256 x 3 that is real or fake. Ultimately, the prediction map is the output once the feature map has been processed through the discriminator and shrunk in size. The prediction map's 70 × 70 receptive fields for each pixel value indicate whether the overlap region's 70 × 70 designation is true (1) or false (0). All encoder-decoder architectures are collectively trained on a loss function that drives the reconstruction of the input from the output. A summary of the Cycle GAN architecture used in the implementation is shown in Table I.

Table 1. SDFCYCLEGAN Architecture For 3D Image Generation

Layer	Output/Prams
Input1(input layer)	[(256,256,3)],0
Conv3D (conv3D)	[(128,128,64)],3136
LeakyReLU	[(128,128,64)],0
Conv3D_1(conv3D)	[(64,64,128)],131200
Normalization	[(64,64,128)],256
Leaky_relu1	[(64,64,128)],0
Conv3D_2(conv3D)	[(32,32,256)],524544
Normalization 1	[(32,32,256)],512
Leaky_relu_2	[(32,32,256)],0
Conv3D_3(conv3D)	[(16,16,256)],1048832
Normalization 2	[(16,16,256)],512
Leaky_relu_3	[(16,16,256)],0
Conv3D_4(conv3D)	[(16,16,256)],1048832
Normalization 3	[(16,16,256)],512
Leaky_relu_4	[(16,16,256)],0
Conv3D_5(conv3D)	[(16,16,1)],4097

2.5. Pseudocode for 3D image generation using the proposed method

In the experiment, a generator (G) was used to generate a 512-dimensional hidden vector Z divided into eight convolutional layers, called a mapping network (MP). The hidden vector Z is transformed into a space w that defines the style of the resulting image.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Algorithm 1: VAE-SDFCycleGAN
Input:2Dinput image x
Output: 3D Image Generation

1. a batch sampling of real data distributions $p_{data}(x)$ obtain real data samples.$x = \{x^{(1)}, x^{(2)}, \ldots . x^{(m)}\}$
2. encode real data samples and obtain latent variables.,mean vector $\mu = \{\mu_1, \mu_2, \ldots\ldots, \mu_m\}$, and standard deviation $\sigma = \{\sigma_1, \sigma_2, \ldots\ldots, \sigma_m\}$
3. compute $z_e^{'} = \mu + \varepsilon \cdot \sigma, \varepsilon \backsim \aleph(0,1)$, get the final coding vector $z_e^{'} = \{z_{e_1}^{'}, z_{e_2}^{'}, \ldots\ldots, z_{e_m}^{'}\}$
4. Updates encode 1 parameter on the appropriate optimization algorithms:
$$L_{VAE} = E_{z\sim q(z|x)}[\log(x|z] - D_{KL}(q(z|x) \parallel p(z))$$
5. Update discriminator parameters with appropriate optimization algorithm:
$$L_D = L(x|c) - k_t L(G(z_e|C)$$
6. Update generator parameters with appropriate optimization algorithm:$L_G = L(G(z_e|C)$
7. Initialize the signed distance U^0 with:
$$U^n(x) = min\left(U^n(x)U^{n-1}\left(x - \frac{\nabla(U^{n-1}|\Gamma)dt}{\|\nabla(U^{n-1}|\Gamma)\|}\right)\right) + dt$$
8. **for** n=1 until convergence **do**
9. $U^n(x) = U^{n-1}(x)$
10. **for** each node x to Γ
11. **for** each node x of Γ which does not belong to a simplex **do**
12. **if** $x \notin \Omega$ **then**
13. $$U^n(x) = min\left(U^n(x)U^{n-1}\left(x - \frac{\nabla(U^{n-1}|\Gamma)dt}{\|\nabla(U^{n-1}|\Gamma)\|}\right)\right) + dt$$
14. else
15. $$U^n(x) = max\left(U^n(x)U^{n-1}\left(x + \frac{\nabla(U^{n-1}|\Gamma)dt}{\|\nabla(U^{n-1}|\Gamma)\|}\right)\right) - dt$$
16. endif
17. end for
18. return U^n
19. extract a minibatch of samples $\{x, y^{(1)}, x, y^{(2)}, \ldots . x, y^{(m)}\}$ from domain X, Y
20. compute the discriminator loss on a real image.
$$\Im_{real}^{(D)} = \frac{1}{m}\sum_{i=1}^{m}\left(D_X\left(x^{(1)}\right) - 1\right)^2 + \frac{1}{n}\sum_{j=1}^{n}\left(D_X\left(G_{Y\to X}\left(y^{(j)}\right)\right)\right)^2$$
21. Compute the discriminator loss on fake images:
$$\Im_{fake}^{(D)} = \frac{1}{n}\sum_{i=1}^{n}\left(D_X\left(G_{X\to Y}\left(x^{(i)}\right)\right)\right)^2 + \frac{1}{n}\sum_{j=1}^{n}(D_X\left(G_{Y\to X}\left(y^{(j)}\right)\right))^2$$
22. Update the discriminator.
23. compute the $Y \to X$ generator loss
$$\Im^{(G_{Y\to X})} = \frac{1}{n}\sum_{j=1}^{n}(D_X\left(G_{Y\to X(y^{(j)})}\right) - 1)^2 + \Im_{cycle}(X \to Y \to X)$$
24. *Compute the $X \to Y$ generator loss*
25. $\Im^{(G_{X\to Y})} = \frac{1}{m}\sum_{i=1}(D_Y\left(G_{X\to Y}\left(x^{(i)}\right)\right) - 1)^2 + \Im_{cycle}(X \to Y \to X)$
26. *Update the generator*
27. endif
28. endif

Figure 3. Pseudocode for 3D image generation using the proposed method.

To realize the difference between the input image and the generated image, the parameters are continuously optimized against the hidden code of the input image. After training the model, we use the generator to incrementally increase the

resolution of the images produced by eight convolutional layers from 512x512 to 1024x1024 and add noise to each layer using AdaIN. Adaptive Instance Normalization (AdaIN) transforms the latent vectors into two scalars (scale and slope) to determine the style of the image generated at each resolution level. The generated image is then fed to the discriminator (D). Finally, the weights of the three networks are adjusted using a backpropagation algorithm to improve the quality of the resulting images. The training algorithm of the proposed model is in Figure 3.

2.6. Evaluation metrics

This study employed both qualitative and quantitative evaluation indicators to assess the effectiveness of the proposed model fréchet inception distance (FID), peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), inception score (IS), chamfer distance (CD) are among the quantitative evaluation metrics. These quantitative evaluation metrics are essential for assessing the quality, similarity, and realism of generated images. Each metric captures different aspects of image quality and comparison, offering a more comprehensive understanding of a model's performance [16], [17]. For qualitative evaluation, the image enhancement is visually inspected to illustrate how well the employed model performs in terms of image sharpness. Quantitative measurements are detailed in Equation 5-9.

1) Inception score (IS)
The inception score is a measure of the quality of images produced by GANs [18], as shown in Equation 5.

$$Exp^{(}E_x[KL(r(m/n)\ //r(m))]= exp(H_y-E_x[H(m/n)]) \quad (5)$$

Where, $r(m|n)$ is the probability of marginal image distribution.

2) Peak-signal-to-noise-ratio (PSNR)
This assesses how well the generated image compares to the corresponding real image utilizing two black-and-white images, I and k [19]. The resulting image quality increases with increasing PSNR (dB). This is computed shown in Equation 6.

$$PSNR(I;K)=10log_{10}\left(\frac{max_i^2}{MSE}\right)=20log_{10}(max^2I)-20log_{10}(MSE_{I,K}) \quad (6)$$

Where, $MSE_{I,K} = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{i=0}^{n-1}\left(I(m,n)-K(m,n)\right)^2$ and Max_i is the minimum possible pixel value.

3) Structural similarity index measure (SSIM)
This serves as a benchmark for calculating how similar two images are to one another [20]. Three key factors are considered when calculating quality measures: contrast, brightness, and structural or correlation. One way to express the SSIM is as shown in Equation 7.

$$SSIM_{(x,y)}=[l(x,y)]^{\alpha}.[c(x,y)^{\beta}.[s(x,y)^{\gamma} \quad (7)$$

Where, α, β, and γ are positive constants, and *l*, *s*, and *c* are the luminance, contrast, and texture, respectively, used to compare the brightness, similarities, and differences between two image patterns, and lightness, darkest regions, and range between them, respectively.

4) Mean square error (MSE)
MSE calculates the regression line's proximity to a set of data points [21]. This is expressed mathematically as shown in Equation 8.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(Y_I-\widehat{Y_i})^2 \quad (8)$$

Where MSE=mean squared error, n is the data point, Y_I is the encoder input, $\hat{Y}_i$ is the decoder output.

5) Chamfer distance (CD)
To measure each point's distance from the closest surface point and add up the squares of those distances, this evaluation metric can be used to determine how similar two sets of points are [22]. It is optimal to measure object accuracy when CD readings are low to compute the score. The CD is defined as shown in Equation 9.

$$d_{CD}(A_1,A_2)=\sum_{x\in A_1}\min_{y\in A_2}\|x-y\|_2^2+\sum_{x\in A_2}\min_{y\in A_1}\|x-y\|_2^2 \quad (9)$$

Where the generated and original images are denoted by $\mathbf{A_1}, \mathbf{A_2}$, and their vertices are represented by *x* and *y*, respectively.

3. RESULTS AND DISCUSSION

The training and testing experiments were performed on a computer with a GPU frequency of 2.5 GHz and a Python Google Collaboratory computer with the Tensor Flow library installed independently. Experiments were performed quantitatively and qualitatively on the publicly available African Coco mask and CelebFace datasets, each validated with selected methods in Section 2.2. A total of 10,000 frames were sampled during the simulation. The image resolution is

512*512, the learning rate is 0.0001, the serial number is 250, and the training is repeated 1500 times. The values chosen for training rate and batch iterations improve stability and speed during training. Detailed experiments are described in the next section.

3.1. Experiment 1: A Qualitative evaluation of the proposed model and the existing GANS on Coco Africa Mask dataset

This experiment demonstrates the qualitative impact of different image generation methods on the COCO African Mask dataset. The generated images using VAE-SDFCycleGAN and other selected source methods using the same input images are shown on the right side of Figure 4. Figures 4(a), (e), (i) show the original historical image, Figures 4(b), (f), and (j) show the image disparity map of the original image and Figures 4(c), (g), and (k) represent a three-dimensional artistic image, Figures 4(d), (h), and (l) represent the results of generating a three-dimensional generative image using the proposed method. Visual inspection of the results generated by the proposed results shows that VAE-SDFCycleGAN produced clearer three-dimensional images compared to other methods used in the implementation.

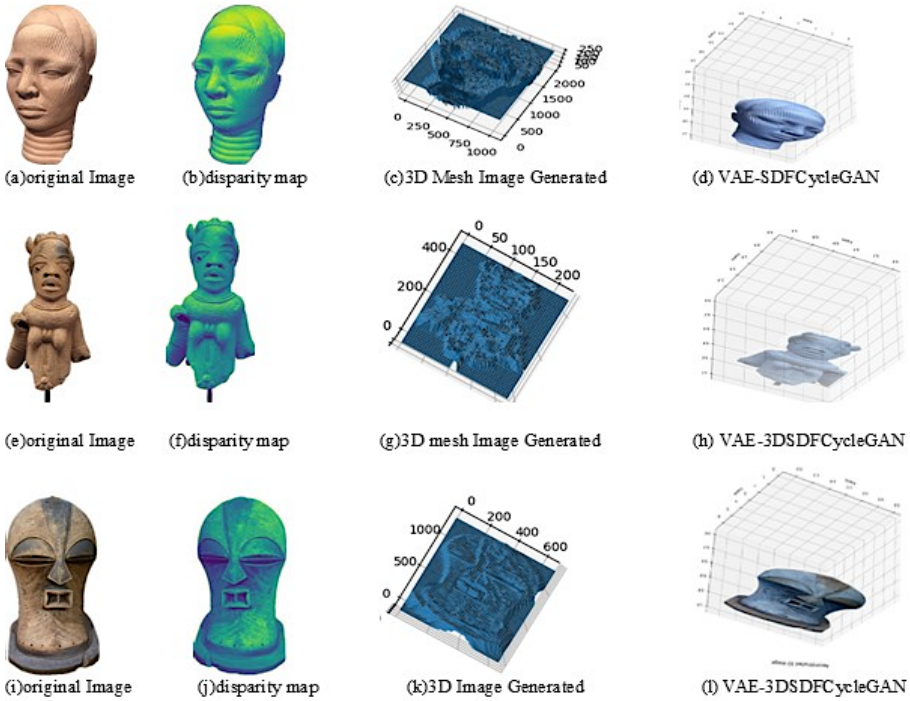

Figure 4. Qualitative evaluation of 3D Artistic image generation on the proposed method and other related GAN methods.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In addition, the performance of each method is evaluated by quantitative evaluation in terms of the FID, IS, SSIM, PSNR, MSE, and CD scores. The summarized results generated images are shown in Table 2.

Table 2. Performance Evaluations of the Proposed Model with other Related GAN models

Models	FID	IS	SSIM	PSNR	MSE	CD
Pix2pix	26.33	10.11	0.613	24.13	84.78	0.733
DCGAN	24.32	13.22	0.524	22.32	88.01	0.932
Proposed Model	12.09	26.37	0.772	31.35	65.32	0.243

Table 2 further analyzes the performance of the generated images in terms of FID, IS, SSIM, PSNR, MSE, and CD. The proposed model's FID score is 12.09, and the IS score of 26.37, based on the result shown in Table III. Images produced by the proposed method are more similar to the original input image when the FID and IS scores are lower. When comparing the similarity of the generated image to the input image, the proposed method yields an SSIM score of 0.772, a CD of 0.243 and a PSNR score of 31.35, the generated image is similar to the original input image. In addition, the proposed method achieved an MSE of 65.32 and IS of 26.37, which is lower than other methods used in implementation. It's convincing that the proposed model produces better, more complete, and more accurate 3D image quality than the DCGAN and Pix2PixGAN methods used in our implementation. The generator and discriminator loss values of the corresponding pre-trained selected models with the proposed model and baseline modes are shown in Table 3.

Table 3. Generator and Discriminator Loss Values with Different Epochs

Epoch	Pix2Pix		DCGAN		Proposed	
	Dis	Gen	Dis	Gen	Dis	Gen
200	0.28	0.47	0.40	0.59	0.20	0.33
500	0.25	0.39	0.34	0.34	0.18	0.38
800	0.38	0.34	0.39	0.47	0.13	0.40
100	0.34	0.44	0.30	0.38	0.15	0.35
1500	0.33	0.38	0.28	0.34	0.11	0.38

Table 3 shows the iteration of training loss for all the models, it is observed that the proposed technique has lesser content loss values compared to other baseline models, and this shows the consistency of the model in terms of generated image content with the original image. The lower the discriminator loss and the higher the generator loss the proposed model generates images that are highly similar to the input image.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.2. Experiment 2: Benchmarking the proposed methodology on the publicly available CelebFace dataset

This section aims to confirm that the proposed model's performance on the publicly accessible CelebFace dataset is consistent. Figure 5 displays the main patterns of the created images as well as the qualitative performance of the model. Figures 5(a), (e), and (i) are the original two-dimensional image, Figures 5(b), (f), and (j) show the image disparity map of the original image, and Figures 5(c), (g), and (k) display the 3D image of the input image. Figures 5(d), (h) and (l) show the results of images generated using the proposed method. The results of the images generated by the proposed method show that the images are sharper compared to other methods.

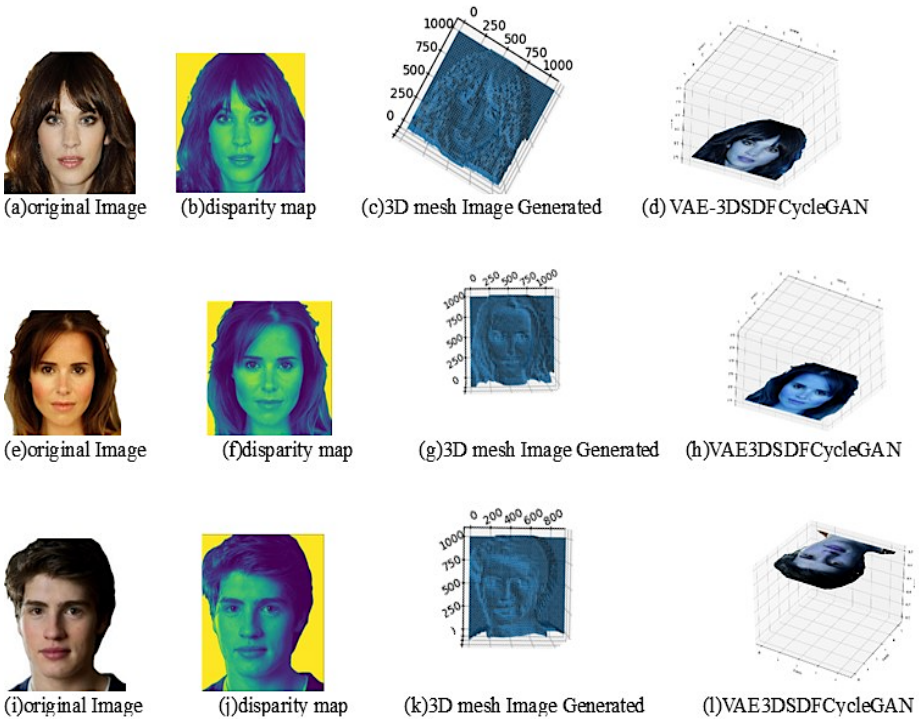

Figure 5. Qualitative evaluation of 3D artistic image generation on the proposed method and other related GAN methods.

To evaluate how close the images generated using each selected method are to the original input images, we used various performance evaluation metrics such as FID, IS, SSIM, PSNR, MSE, and CD scores and a summary of the obtained results. Performance. Results were obtained for each method. The methods are presented in Table 4.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Table 4. Performance Evaluations of the Proposed Model with other Related GAN models

Methods	FID	IS	SSIM	PSNR	MSE	CD
Pix2pix	20.11	16.27	0.652	238.03	75.32	1.340
DCGAN	18.32	21.33	0.701	35.34	72.76	0.721
Proposed method	6.22	33.96	0.823	45.77	60.11	0.232

Table 4 shows that the FID of the proposed model is low at 6.22, the MSE score is 60.11, and the CD is 0.232. The PSNR of the proposed method was 45.77, SSIM was 0.823, and IS was 33.96, which were higher than other basic methods. It is crucial to realize that the generated image is better when the SSIM and IS values are higher, and the generated image is closer to the original image when the FID is lower. For CD, the generated image is closer to the original input image. However, the ten evaluation metrics used in the implementation show that the proposed method produces very similar artistic 3D images compared to the state-of-the-art GAN image generation method used in the artistic image generation implementation. The loss values of the corresponding training of the selected models with the proposed model as in Table 5.

Table 5. Generator and Discriminator Loss Values with Different Epochs for CELEB FACE Dataset

Epoch	Pix2Pix		DCGAN		Proposed	
	Dis	Gen	Dis	Gen	Dis	Gen
200	0.26	0.40	0.33	0.52	0.21	0.31
500	0.25	0.30	0.26	0.38	0.22	0.35
800	0.19	0.32	0.30	0.44	0.20	0.39
100	0.24	0.41	0.22	0.33	0.24	0.34
1500	0.26	0.36	0.18	0.31	0.16	0.36

Table 5 shows the iteration of training loss for all the models, one can observe that the proposed model has lower content loss values compared to other models, and this shows the consistency of the model in terms of generated image content with the original image.

3.3. Discussion

To better evaluate the regenerated performance of the VAE-SDFCycleGAN, we compare it with other traditional and artistic regenerated algorithms, including MSGAN [9], PROGAN[23], SNGAN [24], DAGAN [34], as well as SGAN [25] on CelebFace dataset. According to the comparison study results, the proposed method shows significant improvement for all test indicators compared with other methods. The FID of VAE-SDFCycleGAN is significantly lower compared to

other methods. The six quantitative methods, including FID score, IS score, SSIM, PSNR, and MSE, were used as in Table 6.

Table 6. Comparative of Proposed Model with Other Existing Methods

Methods and Ref	FID	IS	SSIM	PSNR	MSE
MSGAN [9]	28.44	17.78	0.73	15.84	84.33
PROGAN [23]		17.8	0.57	23.11	59.75
SNGAN [24]	22.81	13.58	0.65	27.23	65.12
LSGAN [25]	26.5	17.25	0.82	22.45	76.10
OURS	11.24	28.43	0.96	32.66	54.32

Table 6 indicates that, when compared to other state-of-the-art GAN image-to-image generating methods used in this study, the proposed method performs significantly better on the CelebFace dataset. This method is appropriate for real-time applications because of the high quality of the image generated and the consistent three-dimensional image geometry view generated by the proposed model.

Based on the findings, the performance of the generated images on real-life and publicly available CelebFace image datasets in terms of FID and IS. The proposed model has a lower FID score, and the IS score as shown in Tables 2 and 4. This implies that the images produced by the proposed method are more similar to the original input image. To compare the similarity of the generated image to the input image, the proposed method yields better SSIM score, CD and PSNR score on both real-life and publicly available CelebFace image datasets. In addition, the proposed method achieved lower MSE in comparison with other methods used in implementation. It's convincing that the proposed model produces better, more complete, and more accurate 3D image quality than the DCGAN and Pix2PixGAN methods used in the implementation. The generator and discriminator loss values of the corresponding pre-trained selected models with the proposed model and baseline modes are shown in Table 5.

Furthermore, from the iteration of training loss for all the models in Tables 3 and 5, it is observed that the proposed technique has lesser content loss values compared to other baseline models, and this shows the consistency of the model in terms of generated image content with the original image. The lower the discriminator loss and the higher the generator loss the proposed model generates images that are highly similar to the input image. One can observe that the proposed model has lower content loss values compared to other models, and this shows the consistency of the model in terms of generated image content with the original image.

4. CONCLUSION

The present study elucidated the theoretical framework and practical uses of VAE-SDFCycleGAN in generating high-quality three-dimensional artistic images from two-dimensional input images. This helps to create effective artistic designs and relieves artists of the challenge of creating precise and consistent three-dimensional (3D) images with excellent resolution and shape quality. This study demonstrates that the proposed VAE-SDFCycleGAN has the potential to develop into a potent tool that artists can use to design their creative works. It helps to optimize the creation of artistic three-dimensional images. For the generation of the artistic image implementation, various baseline image generation methods were employed. However, due to the effective performance of the proposed model in advanced 3D artistic image generation, this study takes advantage of this model for the style and generation of three-dimensional artistic images. When compared to other chosen baseline methods for artist image generation in this research, the qualitative simulation results of the proposed model validate its superior performance. The proposed method has a better quantitative and qualitative performance. For broader implications, the VAE-SDFCycleGAN model may be applied in other artistic or design domains. Future work can apply the proposed model to generate images from different art styles or experiment with the proposed mode with different GAN variants. Also, future researchers could look into optimizing computational efficiency or testing the model with larger datasets.

REFERENCES

[1] P. Salehi, A. Chalechale, and M. Taghizadeh, "Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments," *IEEE Transactions on Visualization and Computer Graphics,* vol. 24, no. 6, pp. 216-221, 2018.

[2] D. O. Esan, P. A. Owolawi, and C. Tu, "Image Generation Using StyleVGG19-NST Generative Adversarial Networks " *International Journal of Advanced Computer Science and Applications,* vol. 15, 2024, doi: 10.14569/IJACSA.2024.0150808.

[3] G. Coiffier, P. Renard, and S. Lefebvre, "3D Geological Image Synthesis From 2D Examples Using Generative Adversarial Networks," *Frontier Water,* vol. 2, 2020, doi: 10.3389/frwa.2020.560598.

[4] H. A. Alhaija, A. Dirik, A. Knorig, S. Fidler, and M. Shugrina, "XDGAN: Multi-Modal 3D Shape Generation in 2D Space," *NNN,* pp. 1-14, 2022, doi: doi:10.5244/C.36.NNN.

[5] E. R. Chan *et al.*, "Efficient Geometry-aware 3D Generative Adversarial Networks," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 16102-16112, 2022, doi: 10.1109/CVPR52688.2022.01565.

[6] R. J. Spick, S. Demediuk, and J. A. Walker, "Naive Mesh-to-Mesh Coloured Model Generation using 3D GANs," *In Proceedings of Australasian Computer Science Week (ACSW'20). ACM,* pp. 1-6, 2020, doi: 10.1145/1122445.1122456.

[7] H.-K. Ko, Gwanmo-Park, H. Jeon, J. Jo, J. Kim, and J. Seo, "Large-Scale Text-to-Image Generation Models for Visual Artists' Creative Work," *ACM Symposium on Neural Gaze Detection,* pp. 1-15, 1999.

[8] Z. Niu, S. Xiang, and M. Zhang, "Application of Artificial Intelligence Combined with Three-Dimensional Digital Technology in the Design of Complex Works of Art," *Hindawi Wireless Communication and Mobile Computing* vol. 2022, pp. 1-9, 2022.

[9] A. Karnewar and O. Wang, "MSG-GAN: Multi-Scale Gradients for Generative Adversarial Networks," *arXiv:1903.06048v4 [cs.CV] 12 Jun 2020,* pp. 1-18, 2020.

[10] R. J. Spick, P. Renard, S. Demediuk, and J. A. Walker, "Naive Mesh-to-Mesh Coloured Model Generation using 3D GANs," *Proceedings of Australasian Computer Science Week (ACSW'20),* 2019.

[11] D. Victor, "COCO-AFRICA: A Curation Tool and Dataset of Common Objects in the Context of Africa," *Conference on Neural Information Processing, 2nd Black in AI Workshop,* 2018.

[12] P. A. O. a. C. T. Dorcas. Esan, "Anomalous Detection in Noisy Image Frames using Cooperative Median Filtering and KNN," *IAENG International Journal of Computer Science,* vol. 49, no. 1, 2022.

[13] E. A. Ajayi, K. M. Lim, S.-C. Chong, and C. P. Lee, "Three-dimensional shape generation via variational autoencoder generative adversarial network with signed distance function," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 13, no. 4, pp. 4009-4019, 2023, doi: 10.11591/ijece.v13i4.pp4009-4019.

[14] C. Kingkan and K. Hashimoto, "Generating mesh-based shapes from learned latent spaces of point clouds with VAE-GAN," *in 2018 24th International Conference on Pattern Recognition (ICPR),* pp. 308–313, 2018, doi: 10.1109/ICPR.2018.8546232.

[15] C. M. Jiang and P. Marcus, "Hierarchical detail enhancing mesh-based shape generation with 3D generative adversarial network," *arXiv preprint arXiv:1709.07581,* 2017.

[16] M. Soloveitchik, E. M. Tzvi Diskin, and A. Wiesel, "Conditional Frechet Inception Distance," *arXiv:2103.11521v2 [cs.LG] 28 Feb 2022,* pp. 1-11, 2022.

[17] M. J. Chong and D. Forsyth, "Effectively Unbiased FID and Inception Score and Where to Find Them," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2020, doi: 10.1109/CVPR42600.2020.00611.

[18] I. Vaccari, V. Orani, A. Paglialonga, E. Cambiaso, and M. Mongelli, "A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data," *Sensors,* vol. 3726, no. 21, pp. 1-14, 2021, doi: 10.3390/s21113726.

[19] H. Zhang, H. Li, J. R. Dillman, N. A. Parikh, and L. He, "Multi-Contrast MRI Image Synthesis Using Switchable Cycle-Consistent Generative Adversarial Networks," *Diagnostics,* vol. 12, no. 816, 2022, doi: 10.3390/diagnostics12040816.

[20] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," *Journal of Computer and Communications* vol. 17, no. 3, 2019, doi: 10.4236/jcc.2019.73002.

[21] M. Wenzel, "Generative Adversarial Networks and Other Generative Models," *In Colliot O, editor. Machine Learning for Brain Disorders [Internet],* 2023, doi: 10.1007/978-1-0716-3195-9_5.

[22] Y. Z. N. Wang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2Mesh: generating 3D mesh models from single RGB images," *in Computer Vision ECCV 2018, Springer International Publishing,* pp. 55-71, 2018.

[23] L. Tran, X. Yin, and X. Liu, "Disentangled Representation Learning GAN for Pose-Invariant Face Recognition," in *Conference: IEEE Computer Vision and Pattern Recognition (CVPR 2017)*, Honolulu, Hawaii, 2017, vol. Tran, Luan & Yin, Xi & Liu, Xiaoming. (2017). Disentangled Representation Learning GAN for Pose-Invariant Face Recognition. 10.1109/CVPR.2017.141. , pp. 1415–1424, doi 10.1109/CVPR.2017.141.

[24] T. Miyato, Kataoka, T., Koyama, M., and Yoshida, Y, "Spectral Normalization for Generative Adversarial Networks," *International Conference on Learning Representations (ICLR),* 2020.

[25] X. Mao, L. Qing, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks," *IEEE International Conference on Computer Vision (ICCV),* pp. 2813-2821, 2017, doi: doi: 10.1109/ICCV.2017.304.