Image Generation Using StyleVGG19-NST Generative Adversarial Networks

Dorcas Oladayo Esan, Pius Adewale Owolawi, Chunling Tu
Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa

***Abstract*—Creating new image styles from the content of existing images is challenging to conventional Generative Adversarial Networks (GANs), due to their inability to generate high-quality image resolutions. The study aims to create top-notch images that seamlessly blend the style of one image with another without losing its style to artefacts. This research integrates Style Generative Adversarial Networks with Visual Geometry Group 19 (VGG19) and Neural Style Transfer (NST) to address this challenging issue. The styleGAN is employed to generate high-quality images, the VGG19 model is used to extract features from the image and NST is used for style transfer. Experiments were conducted on curated COCO masks and publicly available CelebFace art image datasets. The outcomes of the proposed approach when contrasted with alternative simulation techniques, indicated that the CelebFace dataset results produced an Inception Score (IS) of 16.57, Frecher Inception Distance (FID) of 18.33, Peak Signal-to-Noise Ratio (PSNR) of 28.33, Structural Similarity Index Measure (SSIM) of 0.93. While the curated dataset yields high IS scores of 11.67, low FID scores of 21.49, PSNR of 29.98, and SSIM of 0.98. This result indicates that artists can generate a variety of artistic styles with less effort without losing the key features of artefacts with the proposed method.**

***Keywords*—Artworks; VGG19; Neural Style Transfer; Generative Adversarial Network; inception score; StyleGAN**

I. INTRODUCTION

Drawings, paintings, and carvings with pencils, brushes, and cardboard were traditional tools used by artists to express their unique creativity and ideas [1]. This infers that the production of artwork requires the craftsman who expects to integrate unique and special imaginative artistic styles to invest much time and energy to show their innovative skills, which can be tiring and overwhelming.

The introduction of Artificial Intelligence (AI) into computer technology applications in recent years has made it possible for artists to enhance their original and creative artwork styles incrementally and effortlessly [2]. GANs have received impressive consideration in artistic image generation due to their ability to learn deep representations without extensive training data. GANs utilize their generator engine to reconstruct the input image and discriminator engine to differentiate between the generated images and input images [3].

Traditional GANs are facing challenges of model inability to generate high-resolution images, poor choice of parameter optimisation methods, and difficulty in the generation of another image style from the content of existing images [4].

The lack of in-depth capability to capture intricate image features makes effective transferring of complex artistic styles to images, resulting in unappealing visual image-generated outputs [5]. Also, most of the existing techniques lack a robust starting point for training, which significantly affects the efficiency and stability of the GANs during the training process [6]. Furthermore, the inability of conventional image generation techniques to separate content and style representations effectively is challenging, thereby affecting the content structure of the original image and the desired style [7]. These issues have significantly affected many artists in the generation of closely related artworks from existing works, thereby creating hindrances for the artist to mitigate their innovative and creative styles in their artworks.

Several studies have made efforts to tackle these issues by using GAN models for artistic image style generation and manipulation [8, 9]. These models include Convolutional Neural Networks (CNNs) [10], Cycle GAN [11], Conditional GAN [12], Genetic Algorithm (GA) [13], etc., which have performed to varying degrees of success, but none have given conclusive solutions to address the challenging gap of creating perfect and realistic artwork due to difficulty in many of the methods to adjust the content structure in images which consequently results in the missing of some important features, distortion, and ambiguity creating local features. Hence, it is important to address the issue of style loss and the generation of imperfect and unrealistic images that most existing GAN models exhibit to assist artists in the improvement and enhancement of their artwork creativity.

To better generate perfect and realistic styled artwork based on existing artwork, this study introduces an innovative method to enhance the creation of artistic images without losing the art image contents. Leveraging StyleGAN, VGG19Net and Neural Style Transfer, StyleGAN generates top-notch images, while the VGG19Net model extracts features and NST is utilized to retain artistic image characteristics for the generation of artistic artefacts having high perceptual and realistic art images [14]. Utilizing this approach can serve as a mechanism for artists to manipulate and generate different artistic styles from existing artworks with minimal effort.

The following are the primary contributions of this study:

1) Fusion of models: The development of a new architecture that integrates multiple inputs and outputs, employing a StyleGAN with the application of VGG19Net for feature extraction, and NST for the preservation of image features, and generation of artistic artefacts having high perceptual and realistic art images. This offers an innovative approach to enhance the performance of art image generation.

2) Parameter optimization: Tweaking different parameters to enhance the aesthetic images generated for visual quantitative visual evaluation.

3) Evaluation metrics: Utilization of different performance evaluation metrics to determine the quantitative enhancement and generation of the newly generated images on the proposed model and other state-of-the-art models.

4) Computational time: Evaluation of the proposed model and other selected recent GAN models on the datasets used to determine their execution time on both CPU and GPU systems.

5) Comparative analysis: Benchmarking and comparing the proposed model performance on curated Coco Mask African and publicly available CelebFace datasets against other Art GAN models.

The remainder of this article follows this structure: Section II outlines related works and the theories of the proposed method. Section III provides an in-depth explanation of the proposed method. Section IV deliberates on the experimental outcomes and assesses the proposed model. Concluding remarks are presented in Section V.

II. Related Works

An extension to GANs, called ARTGAN, is proposed in a study in [15] to artificially generate more difficult and complex images, like abstract art. The suggested model can create artwork that looks natural because it learns more quickly and produces high-quality, realistic images based on the CIFAR-10 dataset, as demonstrated by the results the authors obtained. The authors measure the log-likelihood of the generated artwork using the trained GAN models. One of the limitations of this approach is that the generator works with a limited number of image samples and with hyper-parameter choices and generates imperfect images.

The instability of GAN training was addressed with the proposal of StackGAN [16]. By stacking several generators that can produce images with varying resolutions, the authors used a hierarchical structure. The outcomes showed that 256 by 256 resolution images produced by StackGAN can be visually appealing. In contrast to the more advanced technologies used for statistical data conception, visualizing textual data, particularly for creative text is still in its early stages of development. The limitation of this method is that it cannot handle more extreme and varied image transformations.

In study [17], the concept of image super-resolution using Super Resolved Generative Adversarial Networks (SRGAN) was introduced. To produce photo-realistic natural images, the authors incorporated a perceptual loss function that combines both adversarial and content losses. The adversarial loss incentivizes the solution to conform to natural image attributes by employing a discriminator network that has been trained to differentiate between super-resolved images and authentic photo-realistic images. The model was tested on the BSD100 dataset, where the deep residual network successfully reconstructed photo-realistic textures from significantly down-sampled images within the public benchmarks' dataset. Extensive Mean Opinion Score (MOS) testing highlighted substantial enhancements in perceptual quality with SRGAN. The MOS scores obtained from SRGAN were notably closer to those of the original high-resolution images compared to other prominent methods.

The recovering and restoring of artwork that has been damaged over time due to several factors was introduced in study [18]. The authors utilized a conditional Generative Adversarial Network that involves the generator combining adversarial loss and a discriminator that uses binary cross-entropy loss for optimization. The result of the experiment conducted shows that the method completely removes damage in most of the image and perfectly estimates the damage region. Although the method might be unstable causing the generator to output only a certain type of image and the discriminator to be unable to distinguish between the input image and generated image.

One can see that the existing literature regarding artistic image generation and style techniques has become apparent from the limitations. Previous research has contributed immensely but often lacked comprehensive and advanced approaches to deal with the intricacies of creating perfect and realistic artwork, style loss representation and in-depth capability in capturing intricate image features make effective transferring of complex artistic styles to images. Also, a robust starting point for training optimization of the GANs during the training process is hampered by the computing efficiency of real-time artistic image generation systems. Through a variety of noteworthy contributions like those listed above, this research greatly strengthens solutions to these weaknesses. It develops a novel approach integration of the GAN model which integrates principles of Style Transfer to generate high-quality image styles without losing any art image features.

III. Proposed Method

This section discusses the suggested approach. There are four steps in the suggested method: (a) image acquisition (b) the image preprocessing stage (c) the image feature extraction stage (d) the image generation stage. A detailed explanation is given in the following subsections.

A. Image Acquisition Stage

The experiments in this study used COCO African Mask art images [19]. The dataset is images of African masks that will help readers experience the pinnacle of African art. This dataset consists of 9,300 images of African art as in Fig. 1.

Fig. 1. Samples of the Coco African mask dataset [19].

The detailed proposed framework is shown below in Fig. 2.

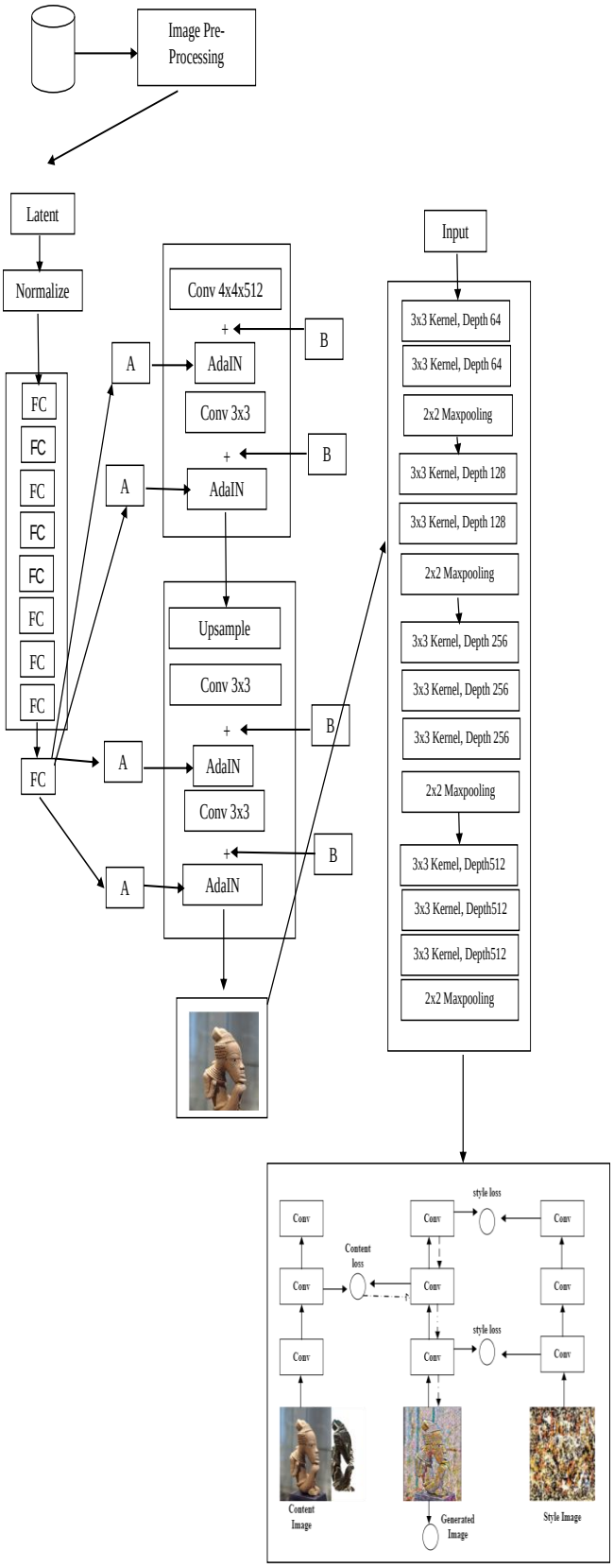

Fig. 2. Proposed styleVGG19-NST framework for artistic image generation.

B. Image Pre-Processing Stage

The input data is pre-processed to improve feature transfer. The pre-processing includes image noise removal and image segmentation. The image pre-processing is done to remove unnecessary and unwanted artefacts that can affect the performance of the GAN models used in this research. Fig. 3 illustrates all the steps in the pre-processing stage used for the implementation of this research.

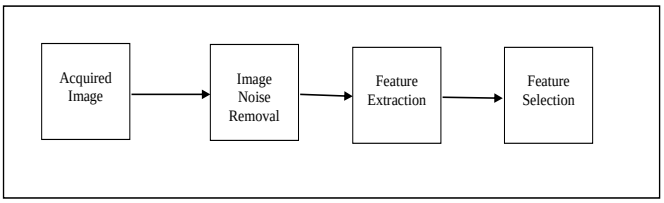

Fig. 3. Flowchart of pre-processing stage.

1) Image segmentation (background removal): In this study, an Otsu segmentation technique is used to model the image from a series of image frames to perform image segmentation. The foreground image's pixels are separated from the background using this technique. To accomplish this, subtract one image at a time (t-1) from the image generated at the time(t). Then the background subtraction is calculated as in Eq. (1).

$$B(x,y,t) = \left(I_{(t-1)}(x_{t-1}, y_{t-1}) - I_t(x_t, y_t)\right) > Thr \quad (1)$$

The selected threshold denoted by *Thr* is dynamically determined to adjust to the changes in frame surroundings. The background image is updated as in Eq. (2).

$$I_{t+1} = \begin{cases} I_t(x_t, y_t) > B(x,y,t), & foreground \\ otherwise & \\ I_t(x_t, y_t) < B(x,y,t), & background \end{cases} \quad (2)$$

2) Image noise removal: After extracting the image foreground, some environmental noises are still present in the foreground image, such as illumination, shadow, light intensity etc. This research adopted the median filtering technique where the noisy image pixels are replaced by the average value of their neighbouring pixels (mask) as in Eq. (3).

$$I'(x', y') = median\{g'(x'+i), (y'+j), i, j \in w'\} \quad (3)$$

Where, $I'(x', y')$ is the image median $g'(x', y')$ is the input image, and $j \in w'$denotes a 2-D image mask. The output of the enhanced image is passed to the feature extraction stage for further processing.

3) Feature extraction stage: At this stage, the objects are recognized objects based on certain characteristics they possess and the HSV colour feature extraction is utilized as in Eq. (4).

$$\mu_{HSV} = \frac{1}{N}\sum_{i,j}^{N} = 1\, P_{HSV}(i,j) \quad (4)$$

Where μ_{HSV} is the image HSV colour mean value, N is the pixel number, and $P_{HSV}(i,j)$ *the* colour component in image shape. The output of colour extraction is fed to feature selection to remove any redundant features.

4) Feature selection: Feature selection is the selection of a subset of relevant features with short dimensionality, short training time, and low overfitting. The extracted features are then spatially related to each other, but there are some semantic inconsistencies between them which can lead to overfitting. In this study, feature selection is performed using the correlation-based features as in Eq. (5).

$$correlation = \frac{\sum(F_{1i}-\bar{F}_1)\sum(F_{2i}-\bar{F}_2)}{\sqrt{\sum(F_{1i}-\bar{F}_1)^2\sum(F_{2i}-\bar{F}_2)^2}} \quad (5)$$

Where, (F_1, F_2) represents the cross-correlation between space F_1 and F_2. The correlations(F_1, F_2) of the two features are in the range of -1 to 1. If two features F_1 and F_2 are independent of each other, then the correlation is $(F_1, F_2) = 0$. The image generation stage receives the feature extraction output after it has been processed.

C. Image Generation Stage

Here, the output of the image extracted is fed to the image generation stage, where the generator (G) is used to generate a 512-dimensional latent vector Z, which is fed into 8 convolutional layers of the Mapping Network (MP). The latent vector Z is converted to a space w that defines the style of the resulting image. The latent code of the input image is continuously optimized for the parameters to achieve the differences between the input image and the generated. The latent code Z, often referred to as the latent spatial mapping of the image, is applied to reduce the painterly style. The vector Z is sampled from a predefined distribution (uniform Gaussian distribution) in the latent space Z which is mapped to the latent space N to produce w which is passed to the AdaIN module. After the model has been trained, the generator is applied to gradually increase the resolution of generated images with 8 convolutional layers from 512x512 to 1024x1024, and AdaIN to add noise to each layer. AdaIN (Adaptive Instance Normalization) converts the latent vector into two scalars (scaling and bias) to control the style of the image generated at each resolution level. In this module, the encoded information w obtained from the mapping network to the generated image. The latent code 'w' generated by the mapping network is passed to the affine transform and AdaIN layer for training. Affine transformations are implemented using two linear planes to create a style using the AdaIN in Eq. (6).

$$AdaIN(x, y) = y_{s,i}\frac{x_i-\mu(x_i)}{\sigma(x_i)} + y_{b,i} \quad (6)$$

Where x is the output feature map to the previous level. AdaIN first normalizes the "zero mean" and "uniform variance" of each channel x_i and then applies the scales y, s andy, b. This means that style y controls the stats of the next convolutional layer's feature map. Where y, s is the standard deviation and y, b denotes the mean.

The discriminator (D) receives the generated image afterwards. A backpropagation algorithm is used to modify the weights of the three networks to enhance the quality of the final image. Furthermore, the generated image from the StyleGAN model was fed into the VGG-19 model alongside the content image. StyleGAN uses a combination of Progressive Generative Adversarial Networks (PGGAN) and neurotransmission techniques [20]. StyleGAN has gained prominence due to its ability to transform low-resolution images into enhanced images [21]. The mean and variance of the feature map x_i generated by the layers in the synthetic network are altered by StyleGAN using reference style bias $y_{b,i}$ and scale $y_{s,i}$in equation (6). As shown in Fig. 4, the generator grows incrementally, adding new constants, scaling the image, and applying style and noise to each block as illustrated in Fig. 4.

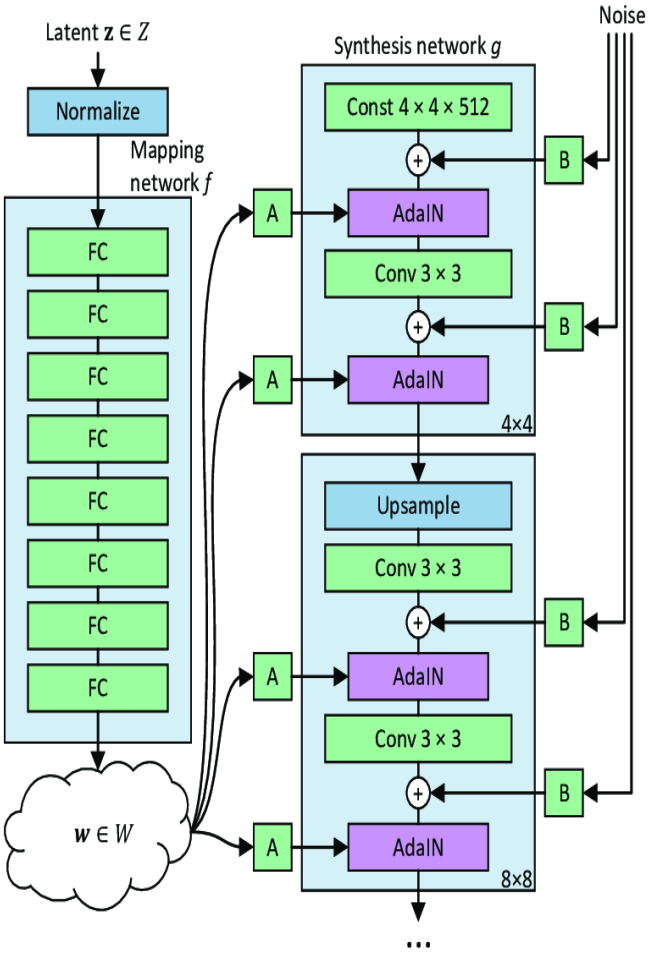

Fig. 4. Architecture of StyleGAN adopted from study [22].

D. Stochastic Variation with Noise Injection

Stochastic variations in an image are small details that do not change the context of the image. There are generators embedded in StyleGAN that try to learn how to generate image styles and content. Noise injection into the StyleGAN created before the AdaIN layer helps create such variations. The noise added to the feature map has zero mean and low variance compared to the feature map. Therefore, the overall context of the image is preserved as the feature map statistics remain the same. From the network, the latent W vector is employed to control the image style of the generated images.

E. VGG19-Network

The VGG-19 is a fully connected model with nineteen deep trainable convolutional layers that include dropout and max pooling layers. The convolutional layer is trained to extract features produced by the StyleGAN in this paper. model output

with a regularized dropout layer and a densely connected classifier [23]. To extend the depth, VGG-19 employs a 3 × 3 convNet configuration. To reduce dimensionality, max-pooling layers are utilized as handlers. The two FCN layers contain 4096 neurons each. To minimize the false positives, while testing, all lesions are considered, as VGG is trained on individual lesions. Convolution layers execute convolution operations across pixel by pixel, enabling the output to progress through the next layers. Filters within the convolution layer are typically 3 × 3 in size and are trained to extract features. After every series of convolutional layers, a Rectified Linear Unit (ReLU) layer and a max-pooling layer are included. ReLU is recognized as an effective non-linear activation function, permitting only the positive values from the input. The architecture of VGG-19 is illustrated in Fig. 5.

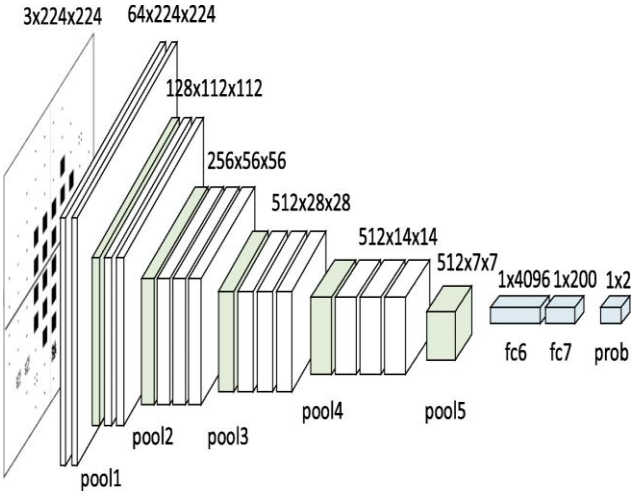

Fig. 5. VGG-19Net adopted from study [23].

F. Neural Style Transfer Model

Blending two images, one with content and the other with style can result in new artwork through Neural Style Transfer. The process of transferring an image's style while preserving its content is known as Neural Style Transfer. To add an artistic touch to your image, all that needs to be changed are the style configurations. The two sets of images that Neural Style Transfer works with are the content image and the Style image.

The content image can be replicated using this technique in the reference image's style. The creative style is applied from one image to another using Neural Networks. To synthesize features and transfer style from one image to another, NST uses a pre-trained Convolutional Neural Network with additional loss functions. NST specifies the following inputs:

- an input (generated) image (g) that contains the final result.
- a content image (c) that is the image to which a style is to be transferred.
- an input style image (s) that is the image from which the style is to be transferred.

The Neural Style Transfer architecture is depicted in Fig. 6.

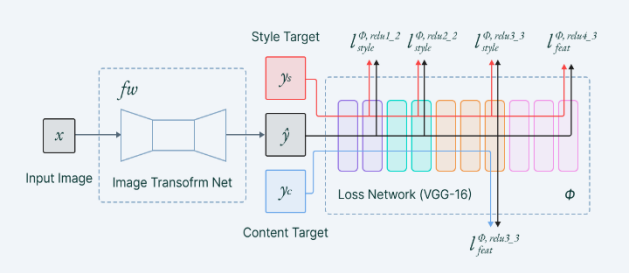

Fig. 6. Neural Style Transfer architecture adopted from study [24].

1) Content loss: Comparisons between the content image and the generated image are made easier by the content loss. The model's upper layers, intuitively, concentrate more on the characteristics seen in the image (the picture's general content). The equation for content loss computes the Euclidean distance between the input image (x) and the content image (p) at layer l, which correspond to the respective intermediate higher-level feature representations. The content loss is shown in Eq. (7).

$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2}\sum_{i,j}(F_{ij}^{l} - C_{ij}^{l})^2 \quad (7)$$

In this context, with the content image denoted as "C," the target image as *"x,"* and the processed layer as *"l," "F"* and *"P"* represent the feature representations of the content image and the target image, respectively, on layer "*l.*"

2) Style loss: Style loss, similar to content loss, uses the squared loss function to measure the difference in style between the synthesized image and the style image. The style loss involves calculating the Maximum Mean Discrepancy between two images, and it is determined using Eq. (8).

$$\mathcal{L}_{style}(\vec{p}, \vec{x}) = \sum_{i=0}^{L} w_l E_l \quad (8)$$

Where, w_l is a weight given to each layer during loss computation, the content image is *p*, and *x* is the target image.

Three key components are essential for generating a style transfer image: content image, style image and generated image. The content image and the style image are modified together to generate new artistic images. The style is the variation added to the content image that produces an entirely new image. The NST model produced stylized images resembling a blend of the content and style images.

Maintaining the generated image's proximity to the local textures of the style reference image was achieved by utilizing the style loss function. However, the generated image's high-level representation is maintained close to the base image by the content loss function. To ensure that the generated locally coherent, the total loss function is used.

G. Evaluation Mechanism

Quantitative and qualitative evaluation metrics are used in this research to analyse the performance of a proposed model. For the quantitative evaluation, the FID, PSNR, SSIM, and IS. For the qualitative evaluation, the enhancement of the image is visually inspected to show the performance of the models used in terms of image clarity. The quantitative metrics are further explained in the following sub-section.

1) Frecher Inception Distance (FID): This metric is used to quantitatively evaluate the quality of created images using the proposed model [25] as in Eq. (9).

$$FID(r,g) = \|\mu_r - \mu_g\|_2^2 + Tr(\sum_r + \sum_g + 2(\sum_r \sum_g)^{1/2} \quad (9)$$

Where the mean and covariance of the real and generated data are represented as $(\mu_g, \sum g)$ and$(\mu_r, \sum r)$.

2) Inception Score (IS): The quality of images generated by GANs is measured by the inception score [25], as in Eq. (10).

$$\exp(E_x[KL(p(m|n)\,||p(m))] = \exp(H_y - E_x[H(m|n)] \quad (10)$$

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

3) Peak Signal-to-Noise Ratio (PSNR): This compares the peak signal-to-noise ratio of two monochrome images, I and k, to determine how good a generating image is compared to a set of real images. As the PSNR (measured in dB) rises, the generated image's quality increases. It is computed as in Eq. (11).

$$PSNR(I;K) = 10log_{10}\left(\frac{max_i^2}{MSE}\right) = 20log_{10}(max^2 I) - 20log_{10}(MSE_{I,K}) \quad (11)$$

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

4) Structural Similarity Index Measure (SSIM): This is an indicator of how similar two images are to one another. The SSIM is expressed as in Eq. (12).

$$SSIM_{(x,y)} = [l(x,y)]^{\alpha}.[c(x,y)^{\beta}.[s(x,y)^{\gamma} \quad (12)$$

Where, α, β, and γ are the positive constants, and *l*, *s*, and *c* are the luminance, brightness, and contrast ranges, respectively, that are used to compare two images. On the other hand, the structure *s* is utilized to analyze the local luminance pattern of two images to determine their level of similarity or dissimilarity.

IV. Result and Discussion

This section presents different experiments carried out to achieve the generation of historical artistic images using the proposed method. The configuration and parameter settings and the experimental simulation are discussed in the subsequent section.

A. Configuration Experimentation and Parameter Setting

The experiment was done on a Central Processing Unit (CPU) and Graphic Processing Unit (GPU) computer using the Google Collaboratory with the Tensor Flow library installed independently. Experiments were quantitatively and qualitatively conducted on both the COCO Mask African dataset and publicly available CelebFace datasets which were validated on selected techniques in sections B - D. A total of 1,500 frames were selected during the simulation. This test data can assist in effective observations of the test performance of the pre-trained proposed model on the trained model. The image resolution is 512*512 and the training has been iterated 1500 times with 0.0001 learning rates and 250 batch numbers. The values chosen for learning rate and batch iterations improve stability and speed during training. Detailed experiments are described in the next section.

B. Hyperparameter Selection for the Proposed Model

During the implementation, the proposed SyleGANVGG19-NST model parameter values were set to the following as shown in Table I.

TABLE I. Hyperparameters Used in the Implementation of Training Stylegan Models

Parameters	Values
Learning rate	0.0001
Batch size	250
Epochs	1500
Beta	0.5
Adversarial loss mode	lsgan
loss weight	10
Identity loss weight	0
Pool size to store fake samples	60

C. Experiment 1.1: A Qualitative Evaluation of the Proposed Model and the Existing Recent Used GANS on the COCO African Mask Dataset

The objective of this study is to address the issue of image style transfer and also generate images with high resolution using the style transfer. Different image pre-processing methods were used, such as image enhancement and feature extraction.

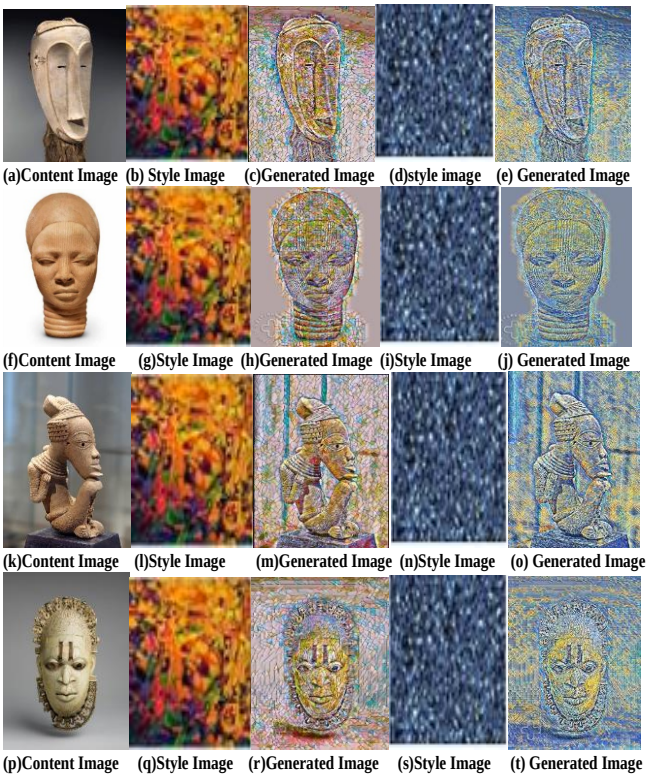

Fig. 7. Qualitative evaluation of SyleGANVGG19-NST model on COCO African mask dataset.

Fig. 7(a), 6(f), 7(k), and 7(p) consist of the original/content image, Fig. 7(b), 7(g), 7(l), and 7(q) is the style image to be matched with the content image, Fig. 7(c),7(h), 7(m), and 7(r) are the generated image and Fig. 7(d), 7(j), 7(n), and 7(s) is the second style image to be matched with the content image, Fig. 7(e), 7(j), 7(o), and 7(t) are the second generated images. This result shows that the proposed model was able to generate an artistic image that is different from the content image. Furthermore, the proposed method is compared with other recently used baseline methods as shown in Fig. 8.

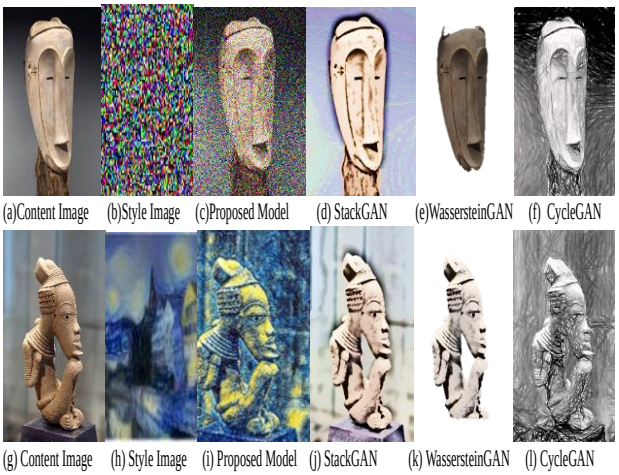

Fig. 8. Qualitative evaluation of the proposed model with other recent baseline art GAN models.

Fig. 8(a) and 8(g) show the content of the original image, Fig. 8(b) and 6(h) shows the style, Fig. 8(c) and (i) are the images generated by the proposed StyleVGG19-NST, Fig. 8(d) and (j) represent the image generated by the Stack GAN model, Fig. 8(e) and (k) and Fig. 8(f) and (l) are the images generated by the Wasserstein GAN and CycleGAN models respectively. One can see that the images generated by the Style method do not transfer the style image into the generated images, the Wasserstein GAN-generated images in Fig. 8(e) have a style transfer issue as some parts of the image generated have already been cut off. The image obtained in Fig. 8(c) and 8(i) a clearer image generated by the proposed StyleVG19-NST method which contains the style image compared with other methods. Loss values of the corresponding training of the selected models with the proposed model as in Table II.

Table III 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. The findings from this experiment show that the optimization used in this experiment was able to generate a perfect image with epoch 1500 which is better in comparison to the image generated with optimization parameters in study [15].

D. Experiment 1.2: Quantitative Evaluation of the Proposed Model and the Existing Recent Used GANS on Curated Dataset

This section aims to quantitatively compute the generated images because of the difficulty in evaluating the model objectively using only subjective visual assessment of the synthetic image. The summary of the result generated in terms of FID, IS, PSNR and SSIM score is used in Table III.

TABLE II. GENERATOR AND DISCRIMINATOR LOSS VALUES WITH DIFFERENT EPOCHS FOR COCO AFRICAN MASK DATASET

Epoch	StackGAN		Wasserstein GAN		CycleGAN		Proposed model	
	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss
200	0.3095	0.8521	0.3721	0.4176	0.3216	0.3987	0.2112	0.2021
500	0.3728	0.5711	0.3364	0.3792	0.3411	0.3769	0.3462	0.3291
800	0.3111	0.4277	0.3516	0.3618	0.3423	0.3591	0.3063	0.2537
1000	0.3693	0.3631	0.3433	0.3536	0.3482	0.3419	0.3146	0.3093
1500	0.4331	0.2911	0.4036	0.3240	0.3855	0.3892	0.3021	0.2187

TABLE III. METRICS RESULT OF THE PROPOSED MODEL WITH OTHER RECENTLY USED ART GAN MODELS

Models	FID	IS	PSNR	SSIM
StackGAN	24.83	6.13	21.85	0.61
Wasserstein GAN	25.56	7.76	24.19	0.68
CycleGAN	29.32	5.72	25.59	0.78
Proposed Model	21.49	11.67	29.98	0.98

From Table III, the art image generated using the proposed StyleVGG19-NST model has improvements in FID, IS, PSNR, and SSIM compared with the Stack GAN, Wasserstein GAN, and CycleGAN models. The proposed model has a higher IS of 11.67, a lower FID of 21.49, an SSIM of 0.98 and a higher PSNR of 29.98 The higher IS, SSIM, and PSNR signifies that the better image quality produced by the proposed StyleVGG19-NST model, and the lower FID indicates that the proposed model generated images that have more structural features than the original image. The findings here show that the proposed method has a better image generation in terms of FID in comparison to the method used in study [16].

E. Experiment 2: Benchmarking the Proposed Methodology on Publicly Available Celebface Dataset

Since there is no common set of image data of similar artistic for the different existing techniques, the validation performance of the proposed technique is tested on a publicly available CelebFace dataset. The face image is randomly selected. The result obtained from the proposed model is compared with some existing recent state-of-the-art Art GAN techniques in terms of image style generation and the use of qualitative evaluation

metrics such as FID, IS, PSNR, and SSIM. The details of the experiments are presented in the following sections.

F. Experiment 2.1: A Qualitative Assessment of the Proposed Model and the Existing Recent Art GANs on the Celebface Dataset

The objective is to assess the dependability and strength of the proposed model using the CelebFace dataset, which is openly accessible. The performance of the image pre-processing stages in image enhancement and feature extraction on curated datasets is qualitatively evaluated as shown in Fig. 9(a)-(o).

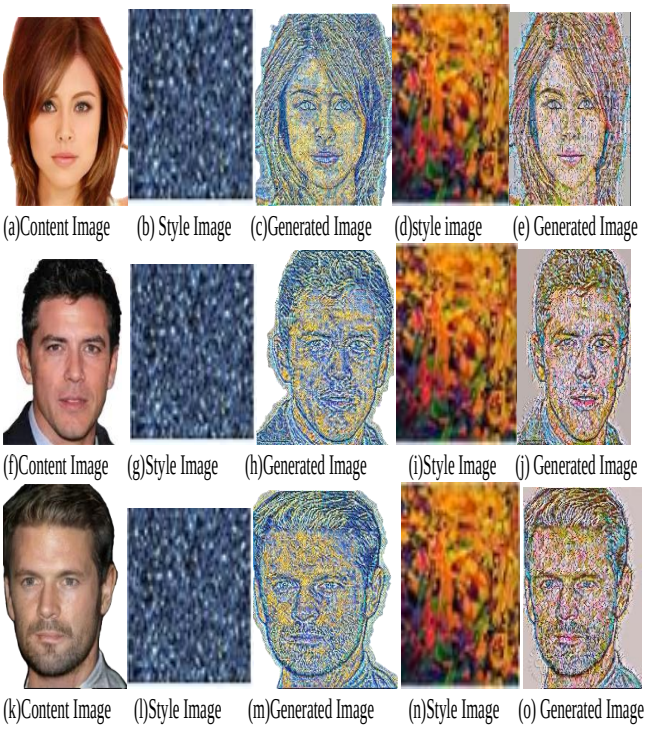

Fig. 9. Qualitative evaluation of the proposed model on the CelebFace dataset.

Fig. 9(a), 9(f), and 9(k) consists of the original/content image, Fig. 9(b), 9(g), 9(l) is the style image to be matched with the content image, Fig. 9(c), 9(h), and 9(m) are the generated image and Fig. 9(d), 9(i), and 9(n) are second style images to be matched with the content image, Fig. 9(e), 9(j), and 9(o) are the second generated images. This result shows that the proposed model was able to generate a stylistic artistic image that is different from the content image. Furthermore, the proposed model and the other existing recent Art GAN models as shown in Fig. 10(a) - (l).

Fig. 10(a) and 10(g) show the images Content image, Fig. 10(b) and 10(h) shows the style, Fig. 10(c) and 10(i) are the images generated by the proposed model, Fig. 10(d) and 10(j) represent the image generated by the Stack GAN model, Fig. 10(e) and 10(k) and Fig. 10(f) and 10(l) are the images generated by the Wasserstein GAN and CycleGAN models respectively. The image obtained image in Fig. 10(c) and 10(i) a clearer image generated by the proposed image which contains the style image compared with other methods.

Fig. 10. Qualitative evaluation of the proposed model with other recent baseline art GAN models.

The generator and discriminator loss values of the corresponding pre-trained selected models with the proposed model are shown in Table IV.

TABLE IV. GENERATOR AND DISCRIMINATOR LOSS VALUES WITH DIFFERENT EPOCHS

Epoch	StackGAN		Wasserstein GAN		CycleGAN		Proposed model	
	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss	Discriminator loss	Generator loss
200	0.3811	0.4473	0.3931	0.5211	0.2652	0.2827	0.3667	0.3117
500	0.3693	0.3652	0.3672	0.3633	0.2977	0.3081	0.3406	0.3513
800	0.3511	0.3630	0.4021	0.4213	0.3211	0.2953	0.3542	0.3911
1000	0.3271	0.4019	0.3163	0.3177	0.3123	0.3078	0.3183	0.3485
1500	0.3522	0.3586	0.3011	0.3033	0.3433	0.3988	0.3496	0.3698

Table IV shows the iteration of training loss for all the art GAN models, it is observed that the proposed model has lesser content loss values compared to other existing recent Art GAN models, and this shows the consistency of the model in terms of generated image content with the original image.

G. Experiment 2.2: Quantitative Evaluation of the Proposed Technique and the Existing Recent Art GANs on the Celebface Dataset

This section aims to quantitatively evaluate the generated images because of the difficulty in evaluating the model objectively using only subjective visual assessment of the synthetic image. The summary of the result generated in terms of FID, IS, PSNR and SSIM score is used as shown in Table V.

TABLE V. METRIC RESULT OF THE PROPOSED TECHNIQUE WITH OTHER EXISTING RECENT ART GAN MODELS

Models	FID	IS	PSNR	SSIM
StackGAN	45.32	9.31	21..32	0.78
Wasserstein GAN	31.45	10.29	24.25	0.83
CycleGAN	35.92	8.11	22.73	0.89
Proposed StyleVGG19-NST method	18.33	16.54	28.33	0.93

From Table VI, one can observe that the proposed StyleVGG19-NST method used in this research on the input image has improvements in terms of FID, IS, PSNR and SSIM compared with the Cycle GAN, DC GAN, and C-GAN models. The proposed model has a higher IS of 16.54, a lower FID of 18.33, an SSIM of 0.93 and a higher PSNR of 28.33 The higher IS, SSIM, and PSNR indicate that the proposed model generates better image quality, and the low FID score signifies that the proposed StyleVGG19-NST model produced images that have more structural features than the original image. Furthermore, the FID of the proposed method is better than the author in study [18].

H. Computational Time

Since most of the image-generated models evaluate the computational time of the model, the computational time of the selected recent Art GAN models is also measured in this research using the same image resolution, the epoch of 1500, and GPU processor on both curated Coco African Mask and publicly available CelebFace datasets. The computational time for each model is shown in Fig. 10(a) and (b) respectively.

From Fig. 11(a), the proposed model has a lower computational time on the same dataset with the same GPU when compared to other techniques. Also, Fig. 11(b) exhibits that the proposed model has a reduced computational time of 45 hours in comparison to other models.

I. Benchmarking Proposed Model with Other Existing Recent Art GAN Techniques in Literature

Validating the performance of the proposed model on CelebFace datasets by comparing it to widely used GAN techniques for artistic image generation is one of the study's goals. The comparison of the proposed StyleVGG19-NST model with recent Art GAN approaches in terms of the dataset, IS score, FID, PSNR, and SSIM score is shown in Table VI.

From Table VI, the proposed model shows significantly better performance on the celeb datasets with FID of 21.49, IS of 11.69, PSNR of 29.98 and SSIM of 0.98 compared to other models. It's worth noting that the higher the IS, PSNR and SSIM of the model the better image quality generated, and the lower the FID the better the structural features that the proposed methodology exhibits when compared with selected baseline recent Art GAN techniques.

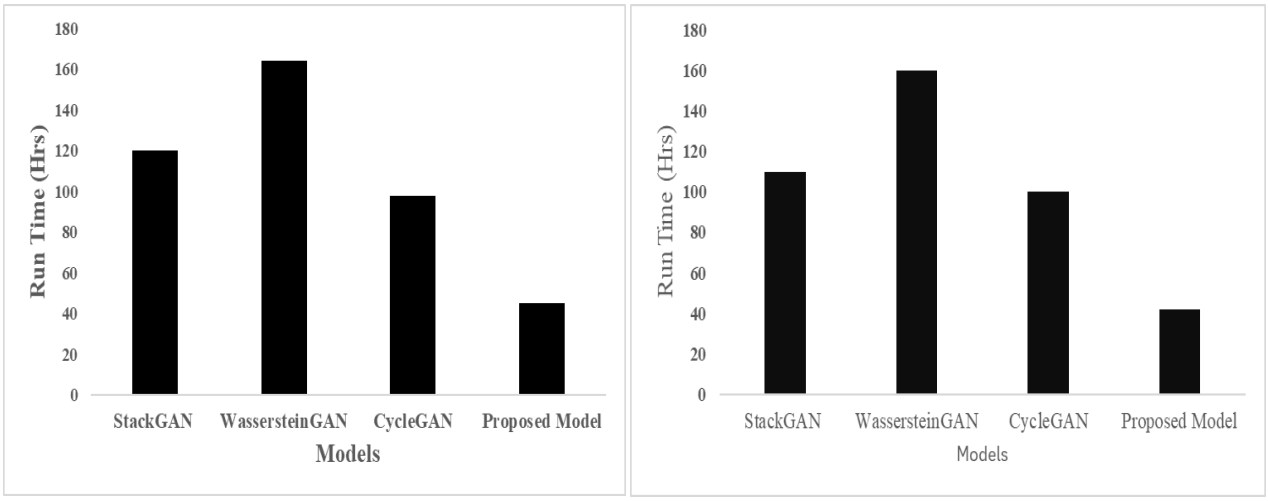

Fig. 11. Computational time of the proposed model with other recent art GAN models; (a) Coco african mask dataset and (b) publicly available CelebFace dataset.

TABLE VI. COMPARATIVE OF PROPOSED MODEL WITH OTHER EXISTING RECENT ART GAN MODELS

Ref	Model	Dataset	FID	IS	PSNR	SSIM
[26]	Boundedness and Continuity GAN (BC-GAN).	Celeb Face	22.8	8.40	19.76	0.76
[27]	Layered Recursive GAN (LR-GAN)	Celeb Face	-	7.17	-	0.89
[28]	Orthonormal	Celeb Face	27.40	2.9	13.56	-
[29]	Denoising Feature	Celeb Face	37.72	-	-	-
[30]	MSGAN	Celeb Face	28.44	-	17.78	-
Proposed Model	StyleVGG19-NST	Celeb Face	21.49	11.67	29.98	0.98

V. CONCLUSION

In this study, the proposed StyleVGG19-NST model was applied to the Coco African Mask artistic dataset and the publicly available CelebFace dataset to generate realistic artistic. The trained model network generated convincing artistic images compared to other baseline model images due to the ability of the proposed model to learn the rich and varied distribution of images. The use of a generator and discriminator network allows the proposed method to capture the spatial structure of an image, which is essential for many artistic image generation tasks. The application of this proposed model on curated artistic dataset images can transform works of art into different styles. Also, both generative loss and adversarial loss values are presented to apply constraints on brightness, colour contrast, and structure of the generated image. This allows the network to converge faster and retain more image detail as a result.

Qualitative and quantitative simulations were performed on the publicly available CelebFace dataset and the curated artistic dataset using the proposed method and other selected baseline methods. The qualitative comparison results show that the proposed model produces better and higher image quality in terms of structural and texture features compared with the baseline models. From the quantitative analysis perspective, the results of the proposed technique on the curated dataset have a high IS score of 11.67, a low FID score of 21.49, PSNR of 29.98 and SSIM of 0.98 while on the CelebFace dataset, the IS of 16.57, FID of 18.33, PSNR of 28.33 and SSIM of 0.93 which is superior compared to other methods used in the simulation. Furthermore, the computational period of the proposed method and baseline models on both curated and publicly available CelebFace datasets with the same training iterations processes show that the proposed technique has a lower computational time than to other models used in the simulation with 48 hours. The overall results of this research exhibit the potential of the proposed methodology for artistic image generation and suggest that the proposed model can be used for extensive image-generation tasks.

Further research is needed to explore how the proposed model performs. Also, future work can be investigated on how inherent biases in the training data of the proposed model can translate to the generated images. Nevertheless, the findings presented in this study can help artists in the generation of different artistic styles with less effort.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Systems Engineering for their financial support.

REFERENCES

[1] S. Chakrabarty, R. F. Johnson, M. Rashmi, and R. Raha, "Generating Abstract Art from Hand-Drawn Sketches Using GAN Models," Proceedings of International Joint Conference on Advances in Computational Intelligence pp. 539–552, 2023, doi: https://doi.org/10.1007/978-981-99-1435-7_45

[2] H. Taherdoost and M. Madanchian, "AI Advancements: Comparison of Innovative Techniques," AI, vol. 5, no. 1, pp. 38-54, 2024. [Online]. Available: https://www.mdpi.com/2504-4990/5/1/3

[3] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy," Computer Science Review vol. 48, p. 100553, 2023, doi: https://doi.org/10.1016/

[4] O. N. Oyelade, A. E. Ezugwu, M. S. Almutairi, A. K. Saha, L. Abualigah, and H. Chiroma, "A generative adversarial network for synthetization of regions of interest based on digital mammograms," Scientific Reports, vol. 12, no. 6166, 2022.

[5] Y. Deng et al., "StyTr2: Image Style Transfer with Transformers," CVF, pp. 11326-11336, 2022.

[6] N. Singh and T. Sandhan, "Learnable GAN Regularization for Improving Training Stability in Limited Data Paradigm," In Kaur, H., Jakhetiya, V., Goyal, P., Khanna, P., Raman, B., Kumar, S. (eds) Computer Vision and Image Processing. CVIP 2023. Communications in Computer and Information Science, vol. 2010, 2024, doi https://doi.org/10.1007/978-3-031-58174-8_45.

[7] S. He et al., "Context-aware layout to image generation with enhanced object appearance," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15049–15058, 2021.

[8] H. Guo, Z. Ma, X. Chen, X. Wang, J. Xu, and Y. Zheng, "Generating Artistic Portraits from Face Photos with Feature Disentanglement and Reconstruction," Electronics vol. 13, no. 5, p. 955, 2024, doi: https://doi.org/10.3390/electronics13050955.

[9] D. O. Esan, P. A. Owolawi, and C. Tu, Generative Adversarial Networks: Applications, Challenges, and Open Issues (intechopen). 2023.

[10] J. Z. Laith Alzubaidi, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, Laith Farhan "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data vol. 8, no. 53, pp. 1-73, 2021, doi https://doi.org/10.1186/s40537-021-00444-8.

[11] C. Dewi, R.-C. Chen, Y.-T. Liu, and H. Yu, "Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation," Applied Sciences, vol. 11, no. 7, p. 2913, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/7/2913.

[12] R. T. A. Guna, R. Benitez, and O. K. Sikha, "Interpreting CNN Predictions using Conditional Generative Adversarial Networks," arXiv preprint arXiv:2301.08067, 2023, doi: https://arxiv.org/abs/2301.08067.

[13] R. Carbonne, S. Gauthier, and J. Leclerc, "Generative Model based on Genetic Algorithm for Artistic Image Generation," arXiv preprint arXiv:2301.08067, 2023. [Online]. Available: Retrieved from https://arxiv.org/abs/2301.08067.

[14] T. Zhu, J. Chen, R. Zhu, and G. Gupta, "StyleGAN3: Generative Networks for Improving the Equivariance of Translation and Rotation," arXiv preprint arXiv:2307.03898, 2023. [Online]. Available: Retrieved from https://arxiv.org/abs/2307.03898.

[15] W. R. Tan, C. S. Chan, H. a. E. Aguirre, and K. Tanaka, "ArtGAN: Artwork Synthesis With Conditional Categorical GANs," 2020.

[16] T. X. H. Zhang, and H. Li, "StackGAN: Text to Photo-Realistic Image Synthesis With Stacked Generative Adversarial Networks," IEEE International Conference on Computer Vision, pp. 5908–5916, 2020.

[17] Y. Jiang and J. Li, "Generative Adversarial Network for Image Super-Resolution Combining Texture Loss," Applied Sciences, vol. 10, no. 5, p. 1729, 2020, doi: https://doi.org/10.3390/app10051729.

[18] B. J. Sowmya, Meeradevi, and S. Shedole, "Generative adversarial networks with attentional multimodal for human face synthesis," Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 2, pp. 1205-1215, 2024, doi: 10.11591/ijeecs.v33.i2.pp1205-1215.

[19] 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.

[20] A. Bhattad, D. McKee, D. Hoiem, and D. Forsyth, "Examining Pathological Bias in a Generative Adversarial Network Discriminator: A Case Study on a StyleGAN3 Model," arXiv. 2023., 2023.

[21] S.-W. Park, J.-S. Ko, J.-H. Huh, and J.-C. Kim, "Review on Generative Adversarial Networks: Focusing on Computer Vision and its Applications," Electronics, vol. 10, 2021.

[22] T. Kramberger, "LSUN-Stanford Car Dataset: Enhancing Large-Scale Car Image Datasets Using Deep Learning for Usage in GAN Training," Applied Sciences, 2020, doi: 10. 10.3390/app10144913.

[23] X. Jia, S. Liu, and Y. Chen, " Enhanced Feature Extraction with VGG-19 for StyleGAN-based Image Synthesis," International Journal of Computer Vision and Machine Learning, vol. 12, no. 4, pp. 45-60, 2023, doi: https://doi.org/10.1016/j.ijcvml.2023.03.004.

[24] A. A. Justin Johnson, and Li Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution " 2016.

[25] 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: https://doi.org/10.3390/s21113726.

[26] K. Liu and G. Qiu, "Lipschitz constrained GANs via boundedness and continuity," Neural Computing and Applications, vol. 32, pp. 18271-18283, 2020.

[27] Y. J, K. A, and B. D, "Lr-Gan: Layered Recursive Generative Adversarial Networks for Image Generation," 2017.

[28] T. Miyato, T. Kataoka, and M. Koyama, "Spectral Normalization for Generative Adversarial Networks," 2018.

[29] D. Warde-Farley and Y. Bengio, "Improving Generative Adversarial Networks with Denoising Feature Matching," presented at the ICLR 2017, 2017.

[30] 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.