International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 3, June 2025, pp. 3118~3126
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp3118-3126

SIGAN: A generative adversarial network architecture for sketch to photo synthesis

Buddannagari Latha, Athiyoor Kannan Velmurugan
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India

Article Info

Article history:

Received Jul 13, 2024
Revised Dec 22, 2024
Accepted Mar 4, 2025

Keywords:

Artificial intelligence
Deep learning
Generative adversarial network
Image generation
Sketch-photo synthesis

ABSTRACT

Of late, with the rise of artificial intelligence (AI) and deep learning (DL) models, image translation has become a very important phenomena which could produce realistic photographic results. Synthesizing new images is widely used in different applications including the ones used by investigation agencies. Image generation from hand-drawn sketch to realistic photos and vice versa is required in different computer vision applications. Generative adversarial network (GAN) architecture is extensively employed for generating images. However, there is need for investigating further on improvising GAN architecture and the underlying loss functions towards leveraging performance. In this paper, we put forth a GAN architecture known as sketch-image GAN (SIGAN) for synthesizing realistic photos from hand-drawn sketches. Both generator (G) and discriminator (D) components are designed based on DL models following a non-cooperative game theory towards improving image generation performance. SIGAN exploits improvised image representation and learning of data distribution. The algorithm we have proposed is known as learning-based sketch-image generation (LbSIG). This algorithm exploits SIGAN architecture for efficiently generating realistic photo from given hand-drawn sketch. SIGAN is assessed using a benchmark dataset called CUHK face sketch database (CUFS). From the empirical study, it is observed that the proposed SIGAN architecture with underlying deep learning models could outperform existing GAN models in terms of Fréchet inception distance (FID) with 38.2346%.

Corresponding Author:

Buddannagari Latha
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation
Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India
Email: latha.reddy5808@gmail.com

1. INTRODUCTION

The field of computer vision and pattern recognition has increasingly focused on face sketch synthesis in recent years. This is due to its significant usefulness in digital entertainment and law enforcement. When it comes to criminal prosecutions, the quantity and quality of surveillance cameras can leave much information about the suspects unreliable. In such cases, artists create sketches based on eyewitnesses memories are typically used as a stand-in for the defendants' identities. By obtaining the face databases from law enforcement agencies or combining the sketches with security camera video, the police can use the sketches to reduce the number of potential suspects on their list [1]. Additionally, the facial sketches are utilized in the creation of animations and as social media avatars. Matching face sketch pictures to photo images in face sketch to photo recognition poses a greater challenge compared to homogenous face recognition because of the significant modality disparity between digital photos and face drawings. Various methodologies have been employed to tackle the issue of modality disparity in face sketch recognition, including techniques including picture to sketch synthesis, common subspace projection, and modality-invariant feature extraction. The system utilizes photo-sketch image pairs as training data to transform facial representations from photographs to sketches, effectively minimizing the differences between these two facial modalities. Conventional homogenous face recognition techniques may be used to recognize faces sketched after they have been created from digital pictures [2].

There are many approaches found in the literature to solve the problems of image translation based on generative advertised network (GAN) architecture [3], [4]. Yan *et al.* [5] introduced the identification-sensitive generative adversarial network (ISGAN) as a solution to the problem of preserving identification information in face photo-sketch synthesis. Subsequent research endeavors to augment efficacy and broaden the methodology for creating caricatures. Bi *et al.* [6] examined the application of conditional generative adversarial networks (cGAN) for translating faces into sketches. Wan *et al.* [7] discussed the retention of facial information in face sketch synthesis and suggests a GAN-based method. Wan *et al.* [8] presented a structure for simultaneous face sketch generation and recognition using a residual dense U-Net generator, which is based on GANs. A multi-task discriminator extracts discriminative characteristics and directs synthesis. Zheng *et al.* [9] the proposed encoder guided generative adversarial network (EGGAN) model utilizes a cycle-consistent GAN architecture, employing two generators and two discriminators, for face photo-sketch synthesis. The literature indicates that because of their learning-based methodologies, the current GAN models used for image translation could perform better. However, the GAN models need to be improved further towards generating more realistic images from given hand-drawn sketches [10]–[13].

An extended U-Net, two discriminators, and an identity constraint are the components of the identity maintained adversarial model (IPAM), which tackles the problem of face sketch-photo synthesis [14]–[18]. Better facial recognition results are shown. Hajarolasvadi included an application [19] such as facial expression, voice, and cross-modal synthesis are covered in great detail by generative models, specifically GANs, in the field of human emotion synthesis. Khan *et al.* [20] highlights improved findings and presents a fully trained GAN for text-to-face synthesis. The technique that has been suggested integrates many datasets to conduct thorough assessments and yields encouraging results. Denser face-related information is the goal of future investigation. Li *et al.* [21] provided a face sketch synthesis technique called regularized broad learning system (RBLS) that uses an incremental learning methodology to preserve rich features. Despite several limits in complicated circumstances, experiments show its usefulness and efficiency. Subsequent work will focus on improving spatial correspondence, handling problems, and investigating new datasets with intricate sceneries and depth. Zhang [22] addressed the shortcomings of previous systems by introducing a cascaded face sketch generation method that is resistant to different lighting conditions. The outcomes of the experiments indicate a notable enhancement and possibility for useful optical systems.

The approach ensures aesthetically pleasing visuals by combining textual and visual aspects. KO et al suggested Superstar GAN, an enhanced StarGAN variant for expansive domains that uses ControlGAN to overcome drawbacks. Improved performance on a variety of datasets with reduced Fréchet inception distance (FID) and learned perceptual image patch similarity (LPIPS). From the literature, it is observed that the existing GAN models used for image translation could improve performance due to their learning-based approaches. However, the GAN models need to be improved further towards generating more realistic images from given hand-drawn sketches [23]–[29].

2. METHOD

This section presents the proposed methodology which includes our SIGAN architecture, details about generator and discriminator, proposed algorithm and evaluation methodology. A generative adversarial network is made up of two parts, the generator and the discriminator, that are trained together via adversarial training. The discriminator is used to improve its ability to discern between actual and fake data, and the generator takes random noise as input and generates synthetic data close to the real data.

2.1. Problem definition

Provided a hand-drawn sketch, developing a novel GAN architecture which exploits DL models for building generator and discriminator with a non-cooperative game between them for automatic generation of realistic photo from the given sketch is the challenging problem considered. Generative adversarial networks (GANs) have emerged as a powerful tool rapidly advancing the state-of-the-art in numerous domains. This paper conducts a comprehensive review to analyses the applications of GANs in the construction industry over the years, and the review aims to enrich the body of knowledge on this emerging deep learning (DL) algorithm in the construction sector.

2.2. Proposed GAN architecture

Recent research in semantic inpainting regards, inpainting as a task of limited picture generation issue [30] it is considered essential that the generated content maintains semantic coherence within the observed context and seamlessly integrates with the surrounding pixels. Similarly, we frame the problem of picture production as one of image completeness, with sketch acting as a weak contextual constraint. The GAN architecture suggested in [31] serves as the foundation for our deep model, with the following technological modifications.

2.2.1. Image representation

We suggest modeling sketch and picture in a combined input space, as opposed to the conventional methods of separating them. Specifically, we spatially combine genuine photographs (B) with their corresponding sketch styles (A) to form joint sketch-image pairs (AB) based on a corpus of samples. This joint image inherently captures the contextual relationship between the sketch and the photo components, aiding in understanding of their combined distribution through GAN. We commence the training process of a GAN model using these combined images, allowing it to leverage the contextual information provided by the sketch component to automatically predict and reconstruct the missing parts of the image. In contrast to earlier work [32]) where z was solely an image embedding, the generator produces a consolidated representation by mapping the merged sketch and image into a non-linear joint space called z. Instead of directly limiting the generated image with the complete z, we can indirectly impose restrictions on it by only using the sketch of the joint embedding z of the input. This allows us to maintain faithfulness while allowing for a certain level of flexibility in the visual presentation of the resulting image.

2.2.2. Objective function

Our purpose is to discover a generated joint image, $G(z\ \hat{})$, that closely resembles the input sketch, to accomplish the most accurate mapping between the distorted and recovered joint images. Our goal is to construct the loss function to incorporate two losses, utilizing the randomly selected input $z\ \sim p\_z$. To assess the contextual resemblance between the unaffected sections that is the input sketch along with the reconstructed drawing we utilize a contextual loss [20], which is "expressed as in (1).

$$L\_contextual\ (z) = D\_KL\ (M \odot y, M \odot G(z)) \tag{1}$$

where M represents the Hadamard production and is the binary mask of the damaged joint picture. Unlike [20], we employ the KL-divergence to measure the similarity between the distribution of two drawings. This approach enhances the alignment of sketches, taking into account that a sketch is a binary image" instead of a natural image. In an ideal scenario, each pixel in the sketch areas of both y and $G(z)$ would have been indistinguishable, resulting in $L\_contextual\ (z) = 0$. Consequently, we impose a penalty on $G(z)$ for its failure to generate a drawing that most closely matches the observed input sketch y. We used the adversarial loss of the G network; the perceptual loss preserves the semantic information of the anticipated "image as in (2).

$$L\_perceptual\ (z) = log(1 - D(G(z))) \tag{2}$$

The desired function for z^ is formulated by combining the two losses through a weighted summation, as elucidated in (3).

$$z\ \hat{} = arg\ (min) \top z\ (L\_contextual\ (z) + \lambda L\_perceptual\ (z)) \tag{3}$$

where λ is a hyperparameter" that uses the input to restrict the output picture. A modest λ will ensure that the input and output look the same.

2.2.3. Our sketch-image GAN (SIGAN)

The training and completion stages make up our GAN. Except for using joint pictures for our training samples, the phase of a training is identical to classic training of a GAN. After completing the training process, we select a "generative network G that effectively reproduces the combined distribution of image data by converting samples from the noise distribution p_z into the data distribution p_data. To achieve our desired outcomes, we must provide either the input image of the damaged joint or the masked-out image component. This representation will enable us to choose the image on the manifold of G that is closest in the latent space. We determine the $z\hat{}$ vector in (3) that minimizes our objective function rather than maximizing $D(y)$. This shows that the distorted input is being projected onto the generator's z space by repetitive back propagation. In particular, the input comprises a combined image featuring solely the left-side drawing, while the right-side image remains concealed, and a vector z started with evenly distributed random noise.

To modify the randomly chosen input z of network G, we apply the loss function described in (3). At present, just the input vector z is modified by using gradient descent, whereas the relative weights of networks G and D stay unchanged. Figure 1 illustrates the process of traversing the latent space during back-propagation, showcasing four iterations. Keep in mind that [33] uses an analogous optimization technique of gradient descent for inverse mapping. Following the back-propagation process, the corrupted input (y) is mapped to the closest vector $z\,\hat{}$ in the latent space. This vector is then fed into the G network to generate $G(z)$. The image is created by utilizing $G(z\,\hat{}\,)$ to complete the absent values of y, which represents the image section:

$$X\_generated = M\odot y + (1 - M)\odot G(z\,\hat{}\,) \tag{4}$$

We utilize a noise vector that is uniformly sampled as input. The influence of the initialization on the resultant image is a clear concern. If the initial sketch section of $G(z)$ perceptually deviates significantly from the input sketch, gradient descent will face challenges in mapping the damaged picture to the closest z in the latent space. Failure samples will arise from this, even if we establish a minuscule λ in (3). We improve the initialization as follows to solve this issue: We use a forward pass to sample N noise vectors uniformly at random and extract the initialized drawings for each one. Next, we determine the pairwise KL-divergence between these N initialized drawings and the input sketch. Out of all the N samples, the initial sketch will be the one with the least KL-divergence, signifying the best initialization. In Figure 1, the entire network is visible. Generator G receives a 100-D random noise vector which is uniformly sampled from −1 to 1. This is done in compliance with [8]. Afterward, the input is reshaped to 4×8×512 using an 8192×2 linear layer.

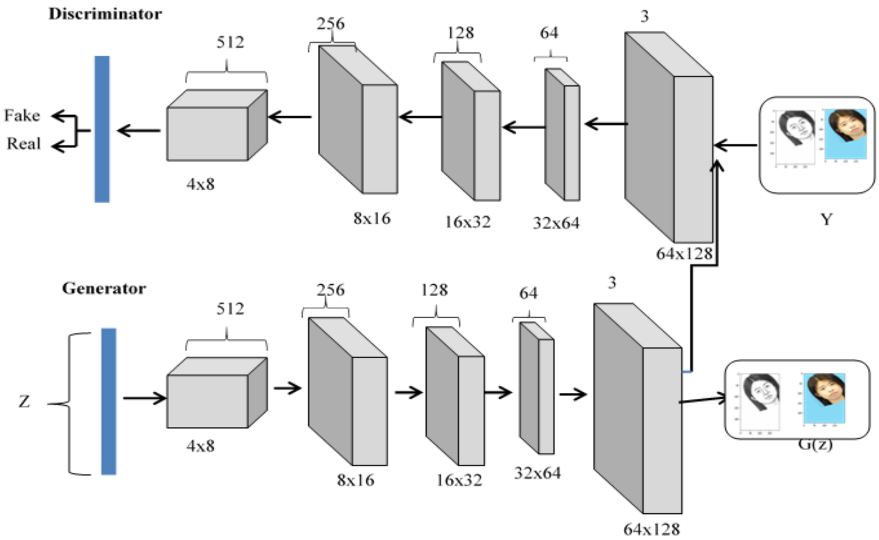

Figure 1. Architectural overview of SIGAN

We employ five up-convolutional layers with a stride of two and a kernel size of five. To speed up training and stabilize learning, after every up-convolutional layer, except for the final one, we add a batch normalization layer. Additionally, all layers utilize the leaky rectified linear unit (LReLU) activation. Tanh is applied at the output layer at the end. A higher resolution picture measuring 64×128 is produced via a nonlinear weighted up sampling of the latent space through a sequence of up-convolutions and non-linearities. A picture with dimensions of 64 by 128 by 3 is used as the discriminator's input. It is followed by four convolutional layers, each of which has twice as many channels as the layer before it and half the feature map's dimension. To generate a 4×8×512 output, we specifically add 4 convolutional layers with kernel size 5 and stride 2. After reshaping the output to one dimension with a fully connected layer, we compute loss using a SoftMax layer.

Genuine freehand drawings come in a wide range of styles and can differ significantly from synthesis drawings that are mechanically created from pictures. We enhance our training data by employing several kinds of drawings as the training set to prevent overfitting to a certain style of sketch image pairings and to increase the generality of the network. To make diverse types of drawings, we specifically employ the FDoG filter provided in [10], the Photocopy effect [1] in Photoshop, and the XDoG edge detector proposed in [11]. We also use [12] to simplify the edge pictures so that they more closely resemble hand-drawn drawings. To train distinct style models, we divide the data in each style into training and testing sets. Firstly, we retrieve the pre-trained XDoG style model, rather than training all style models from start. The networks are then refined using drawings in different styles, such as FDoG, the simplification, and the copier style. The rationale is that XDoG, in our opinion, is more detailed and more akin to the original photographic image. This ensures that the network is trained on high-quality local minima before incorporating additional sketch styles. During the experiment, we demonstrate how the augmenting styles allow for some degree of appearance flexibility while also improving the generalization of the sketch-image relationship.

2.3. Implementation details

Using SGAN, we pretrain the network for every category. Both the generator and discriminator networks employ the Adam optimizer [13] with a beta value of 0.5 and a learning rate of 0.0002. The training duration varies based on the dataset's magnitude, spanning from 6 to 48 hours. This is accomplished by utilizing a batch size of 64 and running the training process for 200 epochs. Once we have acquired a highly skilled version of the XDoG style, we utilize "the same network structure to train other drawing styles sequentially. This is achieved by employing a lower learning rate (e.g., 1e^(-5)) to produce models for these different styles. Throughout the completion process, the input z is modified through the incorporation of a contextual loss and a perceptual loss. The contextual loss and perceptual loss are assigned a λ value of 0.01 and a momentum value of 0.9, respectively". Back-propagation utilizes stochastic clipping. We optimize at test-time to maximize the sketch part in the produced picture that most closely resembles the original drawing by setting a relatively small λ, which makes contextual loss more significant''. During back-propagation, the discriminator and generator are fixed. This update may be completed in 500 iterations based on the experimental findings (the loss (3) converges quickly with our revised initialization, usually reaching stability within 100 iterations in less than 1 second). For each of the three categories, we employ the same network design.

2.4. Proposed algorithm

Our proposition entails an algorithm recognized as learning-based sketch-image generation (LbSIG). This algorithm exploits SIGAN architecture for efficiently generating realistic photo from given hand-drawn sketch. As presented in algorithm 1, it has a learning-based approach for generating images from sketches, using the CUHK face sketch database (CUFS) dataset as input. The output of the algorithm includes the generated image results (R) and performance statistics (P). The process begins with data augmentation of the CUFS dataset (D), which is then split into two subsets known as training set (T1) and test set (T2). The core of the algorithm is the creation and training of the SIGAN model. The architecture of SIGAN is configured as shown in Figure 1, and the model is then compiled. The training process involves using the first subset (T1) to train the SIGAN model (m), which is then persisted for future use. Once the model is trained, it is loaded, and the second subset (T2) is used to generate images with the trained model (m'). The performance of the generated images is evaluated against the ground truth, and both the results (R) and the performance statistics (P) are displayed. In summary, the algorithm follows a structured approach to train a neural network on augmented data, generate images from sketches, and evaluate the generated images based on their fidelity to the ground truth. The key components include data preparation, model configuration and training, image generation, and performance evaluation.

Algorithm 1. Learning based sketch-image generation

```
Input: CUFS dataset D
Output: Image generation results R, performance statistics P
1.  Begin
2.  D' = Data Augmentation(D)
3.  (T1, T2) = Split Data(D')
Building and Training SIGAN
4.  Configure SIGAN architecture (as shown in Figure 1)
5.  Compile SIGAN model m
6.  m' =    Train SIGAN(T1)
7.  Persist model m'
Image Generation
8.  Load m'
9.  R=Image Generation (T2, m')
10. P=Evaluation (R, ground truth)
11. Display R
12. Display P13.
End
```

2.5. Dataset details

CUFS dataset [14] is used for the empirical study in this paper. This dataset is widely used for image translation and computer vision applications. This dataset it has 606 faces and there is a corresponding sketch for each face. Another dataset named CUFSF [15]–[34] is also used in the experimental study of this paper. It has 1,194 samples.

2.6. Evaluation methodology

The realism and diversity of synthetic photographs and drawings are assessed in this study using the (FID)Fréchet inception distance. Given its great degree of agreement with human vision, FID has found widespread use in picture generating applications. Real and synthetic data distributions are closer when the FID value is lower.

3. EXPERIMENTAL RESULTS

In our experiments, shown in Figures 2 to 4 and Tables 1 and 2. we make use of all the test samples and calculate the FID using the 2048-dimensional features extracted from the Inception-v3 network. This network has been subjected to pre-training on ImageNet. To impartially evaluate the quality of the synthesized picture, we utilize the feature similarity index metric (FSIM) for contrasting the synthetic image with the corresponding ground-truth image. Interestingly, while FSIM has gained popularity in the realm of face photo-sketch synthesis and has proven effective in assessing the quality of real images, it falls short in terms of aligning with human perception when it comes to synthesized face photos and drawings. By utilizing the artificially generated photos/sketches as the images in the gallery and the authentic photos/sketches as the probe image, we ultimately conduct a statistical evaluation of the accuracy of face recognition. Null-space linear discriminant analysis (NLDA) is employed to conduct the face recognition tests. Before reporting the average accuracy, we conduct every single face recognition experiment 20 times, randomly partitioning the data during each iteration.

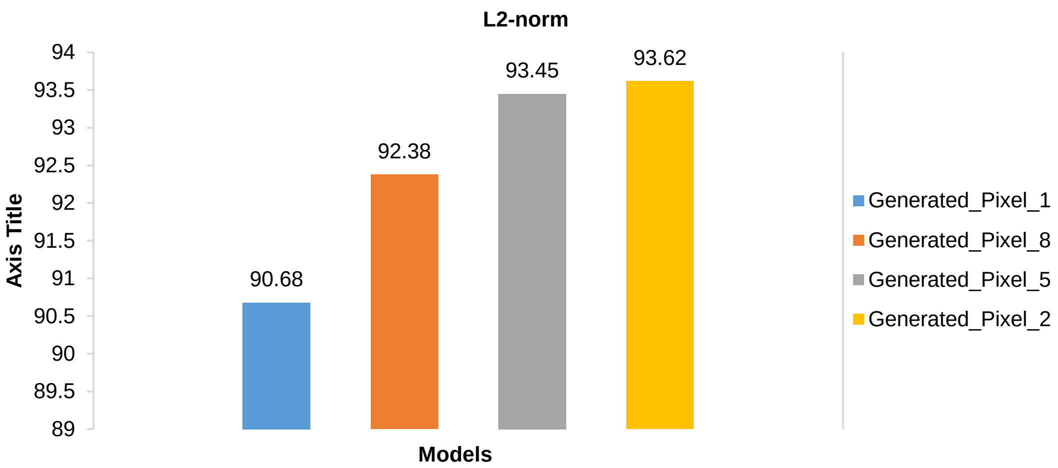

Figure 2. Compute L2-norm and SSIM

Table 1. Performance comparison with CUFS dataset

Sketch-image generation method	Performance (%)		
	FID	FSIM	NLDA
BP-GAN	86.1861	69.1691	93.1931
Conditional GAN	43.2432	71.1711	95.5955
SIGAN (Proposed)	38.2346	71.2531	95.6854

Table 2. Performance comparison with CUFSF dataset

Sketch-image generation method	Performance (%)		
	FID	FSIM	NLDA
BP-GAN	42.9429	68.2682	67.5675
Conditional GAN	29.2292	72.8728	80.9809
SIGAN (Proposed)	20.3142	72.8965	78.0143

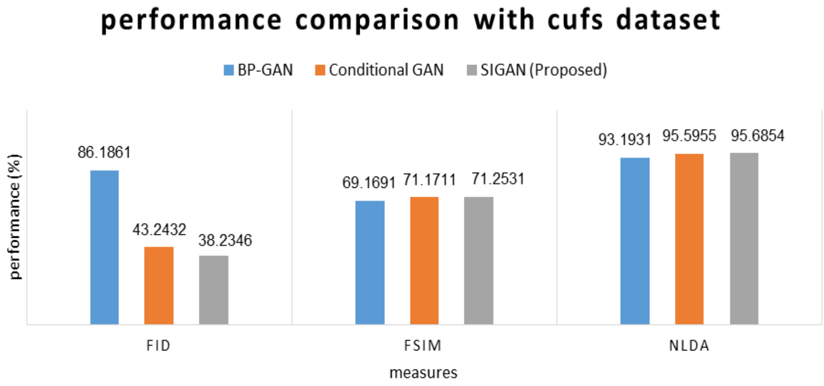

Figure 3. Performance comparison with CUFS dataset

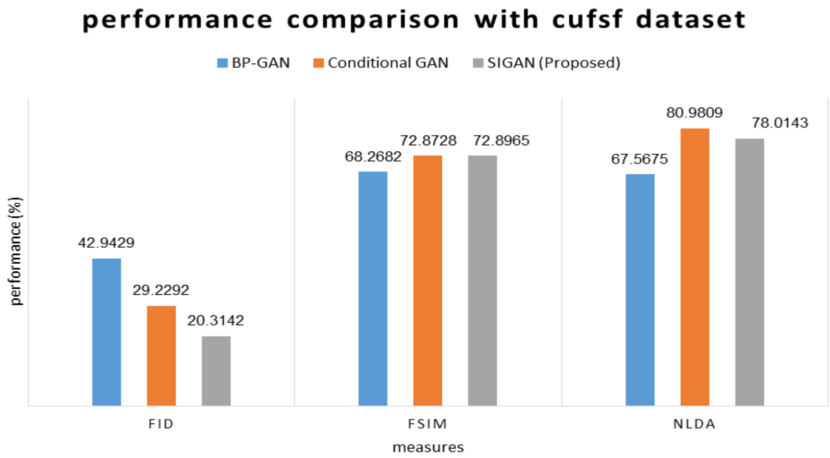

Figure 4. Performance comparison with CUFSF dataset

4. DISCUSSION

With the emergence of AI and DL models, there has been an increased success rate in solving problems in many real-world domains. Computer vision applications are widely used to simplify complex applications. Image translation is one of the important applications desired by many domains. The generation of a sketch image from a given real image and also generation of a real image from the given sketch image are two important activities of image synthesis. Traditional approaches for image synthesis could not provide the required performance. In other words, image processing and heuristic approaches showed limitations in producing accurate outcomes. Of late, learning based approaches came into existence to have more comprehensive approach in learning from input images and to generate desired outcomes with better accuracy. Particularly, GAN architectures were developed for data augmentation and synthesis of new images in the computer vision domain. This paper focuses on generation of real image from the given hand-drawn sketch. The GAN architecture provided in [8] serves as basis for our work in proposing a novel GAN architecture known as SIGAN. SIGAN is found to have better effectiveness compared to state-of-the-art architectures.

5. CONCLUSION AND FUTURE WORK

A GAN architecture known as sketch-image GAN (SIGAN) was proposed for synthesizing realistic photos from hand-drawn sketches. Both generator (G) and discriminator (D) components are designed based on DL models following a non-cooperative game theory towards improving image generation performance. SIGAN exploits improvised image representation and learning of data distribution. Our proposal encompasses an algorithm titled learning-based sketch-image generation (LbSIG). This algorithm exploits SIGAN architecture for efficiently generating realistic photo from given hand-drawn sketch. The contextual information, or the relationship between the sketch and picture components, is automatically captured by the joint image in our system, and this is useful for learning their collective distribution utilizing GAN. In particular, we use joint pictures to train a GAN model, which then uses the context of the matching sketch component to automatically predict the damaged image part. SIGAN is assessed using a benchmark dataset called CUHK face sketch database (CUFS). From the empirical study, it is observed that the proposed SIGAN architecture with underlying deep learning models could outperform existing GAN models in terms of Fréchet inception distance (FID) with 38.2346%. In future, we intend to improve over SIGAN architecture with an encoder for face appearance and an encoder for face labels towards better performance. Another direction for future work is to stack GAN models towards improving performance further.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Buddannagari Latha	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Athiyoor Kannan Velmurugan	✓	✓		✓			✓			✓	✓	✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

No conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] M. Sigala, A. Beer, L. Hodgson, and A. O'Connor, *Big data for measuring the impact of tourism economic development programmes: A process and quality criteria framework for using big data*. 2019. doi: 10.1007/978-981-13-6339-9_4.

[2] G. Nguyen *et al.*, "Machine learning and deep learning frameworks and libraries for large-scale data mining: A survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.

[3] M. S. Sannidhan, G. Ananth Prabhu, D. E. Robbins, and C. Shasky, "Evaluating the performance of face sketch generation using generative adversarial networks," *Pattern Recognition Letters*, vol. 128, pp. 452–458, Dec. 2019, doi: 10.1016/j.patrec.2019.10.010.

[4] W. Chao, L. Chang, X. Wang, J. Cheng, X. Deng, and F. Duan, "High-Fidelity Face Sketch-To-Photo Synthesis Using Generative Adversarial Network," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 4699–4703. doi: 10.1109/ICIP.2019.8803549.

[5] L. Yan, W. Zheng, C. Gou, and F.-Y. Wang, "IsGAN: identity-sensitive generative adversarial network for face photo-sketch synthesis," *Pattern Recognition*, vol. 119, p. 108077, Nov. 2021, doi: 10.1016/j.patcog.2021.108077.

[6] H. Bi, N. Li, H. Guan, D. Lu, and L. Yang, "A multi-scale conditional generative adversarial network for face sketch synthesis," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3876–3880. doi: 10.1109/ICIP.2019.8803629.

[7] W. Wan, Y. Yang, and H. J. Lee, "Generative adversarial learning for detail-preserving face sketch synthesis," *Neurocomputing*, vol. 438, pp. 107–121, May 2021, doi: 10.1016/j.neucom.2021.01.050.

[8] W. Wan and H. J. Lee, "Generative adversarial multi-task learning for face sketch synthesis and recognition," in *Proceedings - International Conference on Image Processing, ICIP*, 2019, vol. 2019-Septe, pp. 4065–4069. doi: 10.1109/ICIP.2019.8803617.

[9] J. Zheng, W. Song, Y. Wu, R. Xu, and F. Liu, "Feature encoder guided generative adversarial network for face photo-sketch synthesis," *IEEE Access*, vol. 7, pp. 154971–154985, 2019, doi: 10.1109/ACCESS.2019.2949070.

[10] K. Du, H. Zhou, L. Cao, Y. Guo, and T. Wang, "MHGAN: multi-hierarchies generative adversarial network for high-quality face sketch synthesis," *IEEE Access*, vol. 8, pp. 212995–213011, 2020, doi: 10.1109/ACCESS.2020.3041284.

[11] Z. Li, C. Deng, E. Yang, and D. Tao, "Staged sketch-to-image synthesis via semi-supervised generative adversarial networks," *IEEE Transactions on Multimedia*, vol. 23, pp. 2694–2705, 2021, doi: 10.1109/TMM.2020.3015015.

[12] D. Wadhwa, U. Maharana, D. Shah, V. Yadav, and P. Pandey, "Human sketch recognition using generative adversarial networks and one-shot learning," in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, Aug. 2019, pp. 1–6. doi: 10.1109/IC3.2019.8844885.

[13] L. Wang, W. Chen, W. Yang, F. Bi, and F. R. Yu, "A state-of-the-art review on image synthesis with generative adversarial

networks,” *IEEE Access*, vol. 8, pp. 63514–63537, 2020, doi: 10.1109/ACCESS.2020.2982224.

[14] J. Zhao, X. Xie, L. Wang, M. Cao, and M. Zhang, “Generating photographic faces from the sketch guided by attribute using GAN,” *IEEE Access*, vol. 7, pp. 23844–23851, 2019, doi: 10.1109/ACCESS.2019.2899466.

[15] J. Zheng, Y. Wu, W. Song, R. Xu, and F. Liu, “Multi-scale feature channel attention generative adversarial network for face sketch synthesis,” *IEEE Access*, vol. 8, pp. 146754–146769, 2020, doi: 10.1109/ACCESS.2020.3015312.

[16] Y. Fang, W. Deng, J. Du, and J. Hu, “Identity-aware CycleGAN for face photo-sketch synthesis and recognition,” *Pattern Recognition*, vol. 102, p. 107249, Jun. 2020, doi: 10.1016/j.patcog.2020.107249.

[17] M. Zhu, J. Li, N. Wang, and X. Gao, “A deep collaborative framework for face photo-sketch synthesis,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 3096–3108, Oct. 2019, doi: 10.1109/TNNLS.2018.2890018.

[18] Y. Lin, S. Ling, K. Fu, and P. Cheng, “An identity-preserved model for face sketch-photo synthesis,” *IEEE Signal Processing Letters*, vol. 27, pp. 1095–1099, 2020, doi: 10.1109/LSP.2020.3005039.

[19] N. Hajarolasvadi, M. A. Ramirez, W. Beccaro, and H. Demirel, “Generative adversarial networks in human emotion synthesis: a review,” *IEEE Access*, vol. 8, pp. 218499–218529, 2020, doi: 10.1109/ACCESS.2020.3042328.

[20] M. Z. Khan *et al.*, “A realistic image generation of face from text description using the fully trained generative adversarial networks,” *IEEE Access*, vol. 9, pp. 1250–1260, 2021, doi: 10.1109/ACCESS.2020.3015656.

[21] P. Li, B. Sheng, and C. L. P. Chen, “Face sketch synthesis using regularized broad learning system,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5346–5360, Oct. 2022, doi: 10.1109/TNNLS.2021.3070463.

[22] M. Zhang, Y. Li, N. Wang, Y. Chi, and X. Gao, “Cascaded face sketch synthesis under various illuminations,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1507–1521, 2020, doi: 10.1109/TIP.2019.2942514.

[23] Q. Liu, H. Zhao, Y. Wang, F. Zhang, and M. Ramasamy, “Sketch to portrait generation with generative adversarial networks and edge constraint,” *Computers and Electrical Engineering*, vol. 95, 2021, doi: 10.1016/j.compeleceng.2021.107338.

[24] X. Li, F. Gao, and F. Huang, “High-quality face sketch synthesis via geometric normalization and regularization,” in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2021, pp. 1–6. doi: 10.1109/ICME51207.2021.9428348.

[25] J. Guo and Y. Liu, “Facial parts swapping with generative adversarial networks,” *Journal of Visual Communication and Image Representation*, vol. 78, p. 103152, Jul. 2021, doi: 10.1016/j.jvcir.2021.103152.

[26] H. Bi, Z. Liu, L. Yang, K. Wang, and N. Li, “Face sketch synthesis: a survey,” *Multimedia Tools and Applications*, vol. 80, no. 12, pp. 18007–18026, May 2021, doi: 10.1007/s11042-020-10301-0.

[27] S. Jiang, Z. Tao, and Y. Fu, “Geometrically editable face image translation with adversarial networks,” *IEEE Transactions on Image Processing*, vol. 30, pp. 2771–2783, 2021, doi: 10.1109/TIP.2021.3052084.

[28] K. K. Babu and S. R. Dubey, “CSGAN: cyclic-synthesized generative adversarial networks for image-to-image transformation,” *Expert Systems with Applications*, vol. 169, p. 114431, May 2021, doi: 10.1016/j.eswa.2020.114431.

[29] W. Xia, Y. Yang, and J.-H. Xue, “Cali-sketch: stroke calibration and completion for high-quality face image generation from human-like sketches,” *Neurocomputing*, vol. 460, pp. 256–265, Oct. 2021, doi: 10.1016/j.neucom.2021.07.029.

[30] N. A. Mashudi, N. Ahmad, and N. Mohd Noor, “LiWGAN: a light method to improve the performance of generative adversarial network,” *IEEE Access*, vol. 10, pp. 93155–93167, 2022, doi: 10.1109/ACCESS.2022.3203065.

[31] U. Osahor and N. M. Nasrabadi, “Text-guided sketch-to-photo image synthesis,” *IEEE Access*, vol. 10, pp. 98278–98289, 2022, doi: 10.1109/ACCESS.2022.3206771.

[32] K. Ko, T. Yeom, and M. Lee, “SuperstarGAN: generative adversarial networks for image-to-image translation in large-scale domains,” *Neural Networks*, vol. 162, pp. 330–339, May 2023, doi: 10.1016/j.neunet.2023.02.042.

[33] X. Wang and X. Tang, “Face photo-sketch synthesis and recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009, doi: 10.1109/TPAMI.2008.222.

[34] W. Zhang, X. Wang, and X. Tang, “Coupled information-theoretic encoding for face photo-sketch recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 513–520. doi: 10.1109/CVPR.2011.5995324.

BIOGRAPHIES OF AUTHORS

Buddannagari Latha is pursuing her Ph.D. in artificial intelligence at Koneru Lakshmaiah Education Foundation, Vijayawada, Andhra Pradesh. She holds an M.Tech in Computer Science Engineering from JNTUH (2015) and a B.Tech in Computer Science & Engineering from Sridevi Women's Engineering College, JNTUH (2008). She is presently working as an assistant professor in the Department of Computer Science and Engineering at Rajiv Gandhi University of Knowledge Technologies - Basar, Telangana, and has 10 years of teaching experience. Her research interests include deep learning and machine learning. She can be contacted at email: latha.reddy5808@gmail.com.

Athiyoor Kannan Velmurugan graduated in computer science and engineering from Arunai Engineering College, affiliated with the University of Madras, Chennai. He obtained his M.Tech. in computer science and engineering from Dr. MGR Educational and Research Institute (Deemed University), Chennai, and earned his Ph.D. in Computer Science and Engineering from St. Peter’s Institute of Higher Education and Research (Deemed University), Chennai. He is currently working as a professor in the Department of Computer Science and Engineering at K L Deemed to be University, Vaddeswaram, Vijayawada, Andhra Pradesh. He has 23 years of academic experience and has published numerous research articles in reputed journals, book chapters, and conference proceedings. His areas of interest include wireless sensor networks, internet of things, artificial intelligence, and machine learning. He can be contacted at email: akvelmurugan74@gmail.com.