International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

Using Generative Adversarial Network (GAN) to Produce Artistic painting

Anushka Naik[1], Amogh Sanzgiri[2]

[1] *Student, Dept. Of Information Technology, Goa College of Engineering, Goa, India*
[2] *Professors, Dept. of Information Technology, Goa College of Engineering, Goa, India*

---***---

Abstract - *This paper uses Generative Adversarial Networks (GANs) to create an antique Indian painting style using JPEG photos, emulating the manner of the well-known Indian artist Raja Ravi Varma. In order to bridge the gap between artificial intelligence and creative expression, the goal is to investigate the potential of GANs to produce aesthetically pleasing and stylistically rich artworks. Using a carefully chosen dataset, the GAN architecture is trained as part of the approach, enabling the model to pick up on the complex compositions, textures, and patterns specific to his painting genres. By use of the antagonistic interaction between a discriminator and generator, the GAN aims to generate artworks that combine elements of computational creativity with conventional creative approaches. The project's output has the potential to be used in content development, digital art creation, and the democratization of artistic tools. This initiative, which uses GANs, adds to the field of generative art by offering a forum for the fusion of artificial intelligence and human creativity.*

***Key Words*: Generative Adversarial Netwok (GAN), artistic style, Raja Ravi Verma, AI painting, CycleGAN, StyleTransfer**

1. INTRODUCTION

Artistic expression has long been a reflection of cultural identity, creativity, and the evolving narrative of human civilization. It has helped us to connect our past with present and highlighting our evolution, importance and dominance over time. Ajanta and Ellora caves is one such significant evidence. But with time art and their artists are disappearing not due to advancement in technology but also due to its technique. Traditional painting technique is time consuming, presence of artist physically and unavailability of raw materials. Indeed, high resolution cameras, small storage devices and fast printing machines have made people shift their interest from traditional art form. Though with such existing problems the craze for traditional art is still among us and is growing over time . In such case there is much need to preserve our traditional art by using modern techniques such as Generative Adversarial Network (GAN).

The advent of Generative Adversarial Networks (GANs)[8] has revolutionized the field of artificial intelligence, particularly in the domain of creativity. GANs, with their ability to generate realistic and novel content, present a unique opportunity to explore traditional artistic styles with the computational power of modern technology. [5] GAN has two components and works like a game based model. A Generator (G) produces the sample and Discriminator (D) tries to distinguish between G's produced sample and original painting. If D is successful in identifying fake then a penalty is imposed on G and if D is not successful in identifying fake then a penalty is imposed on D.Due to this penalty the generator and discriminator learn and improve its performances.

Though GAN have evolved over time but their area of research is limited in particular domain. Most GAN have focused on western art but ignored importance of Asian especially Indian art style. One of the many resons behind this is less dataset availibility and its diversity due to various painters. The culture depicting the art changes from place to place even though the artform may be the same.

2. RELATED WORK

2.1 Generative Adversarial Network (GAN)

Generative adversarial networks (GANs), introduced by Goodfellow[8] , is an emerging technology for both unsupervised and semi-supervised learning. They are implicit density generative models, and they are characterized by two main components: a generator G, and a discriminator D. The basic idea of GANs is to set up a game between the generator and discriminator. The former tries to generate samples that are intended to come from the real data distribution, while the latter examines real and generated samples in order to distinguish between real or fake data. A common analogy is to think of the generator as an art forger, and the discriminator as an art expert. The forger tries to create forgeries which are increasingly similar to real paintings, in order to deceive the art expert. The expert, at the same time, learns more and more sophisticated ways to discriminate between real and false artworks. One of the most crucial points of GANs is that the generator has no direct access to the real data: the only manner for it to learn is through interaction with the discriminator. By contrast, the discriminator has access to both real and generated data. This behavior can be expressed via a min max game, where the generator tries to minimize the gain of the discriminator, while the discriminator tries to do the opposite.

The adversarial modeling framework is most straightforward to apply when the models are both multilayer perceptrons. [7] To learn the generator's distribution p_g over data x, we define a prior on input noise variables $p_z(z)$, then represent a mapping to data space as $G(z; \theta_g)$, where G is a differentiable function represented by a multilayer perceptron with parameters θ_g. [6,7] We also define a second multilayer perceptron $D(x; \theta_d)$ that give an outputs in a single scalar. D(x) represents the probability that x came from the data rather than p_g.

$$\log(1 - D(G(x))) \quad (1)$$

We train D to maximize the probability of assigning the correct label to both training examples and samples from G. We simultaneously train G to minimize.

$$\min_G \max_D V(D,G) = \mathrm{E}x \sim p\mathrm{data}(x)[\log D(x)] + \mathrm{E}z \sim pz(z)[\log(1 - D(G(z)))] \quad (2)$$

Advantages of GAN[8]

- Image synthesis is the ability to produce images, starting from another type of information. This information can be random noise, a text describing the image, or a feature of the image.
- Image-to-image translation is translating the possible representation of one scene into another, such as mapping grayscale images to RGB, or generating an image from only the edges.
- The generator, in GANs, learns a mapping between an arbitrary latent space and data space, in a completely unsupervised manner. The generator associates the feature code values to the actual semantic attributes of the output.
- In the previous applications, the goal of the model is to train the generator with the help of the discriminator, which acts like a teacher. Usually, after the learning phase, the discriminator is discarded, and only the generator is used. In semi-supervised learning, this paradigm is shifted, since the objective is to train the discriminator, with the help of the generator.

2.1 Arbitary Style Transfer [3]

Arbitrary style transfer refers to the process of applying the artistic style of one image (the style reference) to the content of another image (the content reference) in a way that preserves the key features of the content image while adopting the artistic characteristics of the style image. This technique is a subset of neural style transfer, which utilizes deep neural networks to achieve the transfer of artistic styles

2.2 Neural Style Transfer [4]

Neural style transfer is an optimization technique performed by taking two image, one is normal image and a artistic image and blend them together so the output image looks like the content image, but "painted" in the style of the style reference image. This is implemented by optimizing the output image to match the content statistics of the content image and the style statistics of the style reference image. These statistics are extracted from the images using a convolutional network.

2.3 Image to Image Translation [7,6]

Image-to-image translation is a generative artificial intelligence (AI) technique that translates a source image into a target image while preserving certain visual properties of the original image. This technology uses machine learning and deep learning techniques such as generative adversarial networks (GANs); conditional adversarial networks(cGANs); and convolutional neural networks (CNNs) to learn complex mapping functions between input and output images.Image-to-image translation allows images to be converted from one form to another while retaining essential features.The goal is to learn a mapping between the two domains and then generate realistic images in whatever style a designer chooses. This approach enables tasks such as style transfer, colorization and super-resolution, a technique that improves the resolution of an image. The image-to-image technology encompasses a diverse set of applications in art, image engagement, data augmentation and computer vision, also known as machine vision. For instance, image-to-image translation allows photographers to change a daytime photo to a nighttime one, convert a satellite image into a map and enhance medical images to enable more accurate diagnoses.

2.4 StarGAN [6]

For image-to-image translation, StarGAN represents a major breakthrough in generative adversarial networks (GANs), especially when it comes to multi-domain and attribute-manipulation scenarios. StarGAN is a flexible option for a range of image translation problems since it provides an integrated framework that can manage several domains in a single model. One of StarGAN's primary innovations is its one-to-many translation capability, which enables the simultaneous translation of a single input image into several target domains. Applications that need to manipulate style or control features, like altering the color of hair, a person's expression, or other visual elements in photos, must have this feature.

Through a shared generator and discriminator architecture, which allows the generator to learn to map input images to different output domains and the discriminator to distinguish between actual and created images across all domains, StarGAN is able to operate across numerous domains. This configuration reduces computational complexity and resources by enabling effective training and inference for handling a variety of domains without the requirement for separate models for each domain. To further enhance translation quality, StarGAN also includes a domain classification loss to enforce the generator's capacity to generate realistic images in each target domain. Researchers and practitioners working on image synthesis, style transfer, and domain adaptation tasks across many visual domains have grown to favor StarGAN because of its versatility, attribute control, and one-to-many translation capabilities.

2.5 CycleGAN [7]

In problems where getting paired data is difficult or impracticable, CycleGAN has emerged as a key framework for unpaired image-to-image translation. An expansion of conventional GANs designed especially for situations involving unpaired data is called CycleGAN. Using cycle-consistency constraints to enforce meaningful translations, CycleGAN can learn mappings between two domains without the need for directly paired instances, which is one of its main advantages. This method makes the process of acquiring data easier while simultaneously improving the model's ability to generalize to new data and other picture distributions.

Two generators and two discriminators, each tasked with translating images across two domains and distinguishing real from created images, make up CycleGAN's architecture. A key component of training is the cycle-consistency loss, which makes sure that when an image is reconstructed from one domain to another and back again, it closely resembles the original, thereby transferring style or qualities without sacrificing substance. Because of this mechanism, CycleGAN generates images with higher realism and coherence, which makes it suited for a variety of image translation tasks, including artistic rendering, object transfiguration, and style transfer. With its ease of setup, efficiency when processing unpaired data, and capacity to generate translations of excellent quality, CycleGAN has solidified its place as a top framework for picture synthesis and domain adaptation.

2.6 Pix2Pix [7]

A pioneer in the field of image-to-image translation, Pix2Pix is well known for producing outputs of excellent quality when given paired training data. Pix2Pix is a conditional generative adversarial network (GAN) version that concentrates on challenges for which training input-output pairs are available. Pix2Pix's pixel-level mapping skills are one of its main advantages; they enable accurate translation between domains, such as the conversion of grayscale images to color, the creation of realistic photographs from sketches, or the conversion of satellite images to maps. Because it allows for precise control over picture changes, Pix2Pix is an adaptable tool for a variety of computer vision and image synthesis applications.

Pix2Pix's architecture consists of a conditional GAN configuration with a discriminator that separates generated pairs from real pairs and a generator that learns to map input photos to output images in pairs. Pix2Pix can generate outputs that are both visually convincing and contextually meaningful by conditioning the generator on the input photos. This allows Pix2Pix to capture complex details and structures in the translated images. Furthermore, Pix2Pix uses a mix of adversarial loss and pixel-wise loss to guarantee the generated images' local fidelity and global coherence. Pix2Pix is well-suited for tasks requiring precise picture transformations, such as image colorization, image inpainting, and semantic segmentation to image synthesis, because these training objectives provide realistic and sharp outputs.

Table 1:- Comparison table of StarGAN, CycleGAN, Pix2Pix

Comarison of StyleGAN, CycleGAN, Pix2Pix		
StarGAN	**CycleGAN**	**Pix2Pix**
Uses paired dataset	Uses unpaired dataset	Uses paired dataset
Can generate into multiple target domain	Maintain content integrity	Used for pixel level mapping task
Control over attributes (hair, facial expression)	Style changes content remain same	Condition input images or labels

For this paper CycleGAN is used over StarGAN and Pix2Pix because of:

- Flexibility: In real-world situations where acquiring paired data can be difficult, CycleGAN's capacity to operate with unpaired data offers greater flexibility.
- Generalization: The cycle-consistency loss encourages generalization, which enhances resilience and allows for greater adaptability to new data.
- Domain Adaptation: CycleGAN's method of learning cross-domain mappings while maintaining content consistency makes it a viable contender for tasks centered on domain adaptation or style transfer without explicit matched examples.

- Easy to Use: CycleGAN is more approachable and useful for many image-to-image translation jobs due to its ease of setup and training, as opposed to Pix2Pix's dependence on paired data and StarGAN's complexity when handling numerous domains.

3. PROPOSED APPROACH

3.1 Approaches To Project

This project proposes a multifaceted approach that combines image processing techniques, rule-based transformations, and Generative Adversarial Networks (GANs) to achieve the transformation of ordinary JPEG images into artworks inspired by the timeless style of Raja Ravi Varma.

Data Collection: The dataset of Raja Ravi Varma paintings, ensuring representation of human portrait. This dataset forms the foundation for the subsequent stages of the project.Along with that CelebA dataset is collected

Preprocessing: Preprocessing techniques will be applied to the images, enhancing relevant features and preparing the dataset for training.

Splitting Data: The painting dataset is used for training the model and CelebA dataset is used for testing the trained model.

Model Selection: Selecting an appropriate model such as GAN or CNN. This project will be using GAN.

Model Architecture: The propose model will be consisting the following: Defining GAN Architecture

Generator (G): The generator accepts random noise or latent space and generates a target sample using it. After the discriminator's evaluation it improves and generates a better sample. Once the discriminator fails to identify a fake the recently generated sample is used as output.

Discriminator (D): The discriminator takes the input from generator G and compares it with already existing painting images present in the database. It produces an output in binary form i. e either 0 or 1. 1 indicates it has succeeded in identifying that image is fake and 0 indicates it has failed to identify the real image. This is an important step as it decides who should be penalized.

Loss Function: It is a penalty which is applied based on the discriminator's decision.it is used after the discriminator model.

Adversarial loss: [2] Training the generators to produce images in the target domain that are identical to genuine images based on the matching discriminator is the aim of the adversarial loss.

Cycle-Consistency Loss:[1] This loss imposes the requirement that a picture that is translated from one domain to another and back again must resemble the original image. The L1 or L2 distance between the original and reconstructed images is used to compute it.

Optimizer: It helps in adjusting the learning rate and decreases the training time of the GAN model.Along with that it can maintain flexible data and also handle noisy data.

Stochastic Gradient Descent It updates the model based on negative gradients and also tunes the learning rate.

Hyperparameter: It helps in determining the size of input vector, number of layers and number of hidden units in each layer.

Normalization and activation function: Normalization helps in scaling all the features at a similar scale.

Activation function is applied on the output layer where itwill capture patterns, model artistic style , transform image, etc.

1. Block Normalization: [2] Help in improving training of models and reducing covariate shift. Covariate shift occurs when the distribution of input data changes during training .It Is used in the Convolutional 2D layer .I have used this in both discriminator and Generator.

2. ReLU Activation function: [1] Rectified LinearUnit (ReLU) is used to speed up the process. It is used in Conv2D layer.I have used thin in generator

3. Leaky ReLU Activation function: [1] It is used to introduce non- linearity.

Training: The heart of the project lies in the utilization of a Generative Adversarial Network (GAN) to learn and reproduce the intricate patterns, textures, and stylistic details inherent in Raja Ravi Varma's paintings which will learn adversarially by G and D

Evaluation: Obtained generator samples will be compared by discriminator and determine fake or real.based on it loss will be assigned to either G or D.

Hyperparameter tuning Hyperparameters such as activation function padding strides, filter, kernel will be adjusted looking at the output obtained and based on requirements.

Result analysis: Loss curve of generator and discriminator is computed which will help to analyze the result obtained.

3.2 Proposed Algorithm [1]

1. Compile a dataset of pictures from two different categories, such snapshots and paintings.

2. The photos should be preprocessed (resized, normalized, and enhanced), then split into sets for training and validation.

3. Make two networks of generators G_1 and G_2, where G_1 translates pictures from domain A to domain B and vice versa for G_2

4. For both generators, use a comparable architecture, which is frequently based on a U-Net structure for detail capture.

5. Construct two networks of discriminators. D_1 and D_2 to differentiate between created and actual images in domains A and B, respectively.

6. To improve training stability, use PatchGAN discriminators to assess local image patches rather than the full image.

7. To trick the corresponding discriminators, define the adversarial loss (GAN loss) for each of the two generator networks.

8. Include cycle-consistency loss to guarantee that the translated image is almost identical to the original. The L1 or L2 distance between the original and reconstructed pictures is used to calculate this loss.

9. To ensure that an image translated from domain A to B looks comparable to other images in domain B, you can optionally include identity loss (and vice versa).

10. Initialize the generator and discriminator networks with random weights.

11. Alternately train the generator and discriminator networks using mini-batches:

12. Update the discriminator networks:

    Generate fake images for domains A and B using their respective generators.

    Train D_1 to distinguish between real images from domain A and fake images generated by G_2.

    Train D_2 similarly for domain B.

13. Update the generator networks:

    Generate translated images from domain A to B and back (and vice versa).

    Compute the adversarial loss to fool the discriminators and the cycle-consistency loss to preserve image structure.

    Update G_1 and G_2 using these losses.

14. Repeat this training loop for multiple epochs, monitoring generator and discriminator losses to ensure convergence.

15. To test the quality of translated images, use a different validation set and consider things like visual fidelity, image-to-image consistency, and domain transfer correctness.

16. For quantitative assessment, use evaluation criteria like the Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), or perceptual similarity measurements (such employing CNNs that have already been trained).

17. Deploy the trained generator networks for inference after the training process is over and desired outcomes are obtained.

18. Using the acquired mappings G_1 and G_2 , translate images between domains A and B.

19. If required, use any post-processing methods (such as color correction and denoising) to improve the translated photos' visual quality.

3.3 Dataset

The experiment is using Raja Ravi Verma's paintings, collected from various sources and CelebA dataset from kaggle, which consists of 120 and 2,0,000 samples respectively.

The dataset is divided into training (80%) and testing (20%) sets using a stratified random split. We applied standard preprocessing techniques, including [resizing, normalization and data augmentation].

3. CONCLUSIONS

This paper proposes a model that will transform digital image of human portrait into Raja Ravi Verma style paintings who was well known in the mid 18th century for his unique style of blending european oil painting with traditional Indian art. The GAN uses three generators and two discriminators, where first two generators and a discriminator will be used to produce colours and later these colours will be used in paintings to generate target output.

REFERENCES

[1] J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.

[2] Sergey Ioffe and Christian Szegedy. 2015. “Batch normalization: accelerating deep network training by reducing internal covariate shift” . In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 (ICML'15). JMLR.org, 448–456.

[3] R. Abdal, Y. Qin and P. Wonka, "Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 4431-4440, doi: 10.1109/ICCV.2019.00453.

[4] Johnson, J., Alahi, A., Fei-Fei, L. (2016). ”Perceptual Losses for Real-Time Style Transfer and Super-Resolution” . In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9906. Springer, Cham. https://doi.org/10.1007/978-3-319-46475-6_43

[5] Gatys, L. A.; Ecker, A. S.; Bethge, M. “Image style transfer using convolutional neural networks.” In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2414–2423, 2016.

[6] Y. Choi, et al., "StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018 pp. 8789-8797. doi: 10.1109/CVPR.2018.00916

[7] P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[8] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'14). MIT Press, Cambridge, MA, USA, 2672–2680.