Replication of Characteristic Visual Motifs of Indian Rural Art Forms Using a Generative Adversarial Network

Anhiti Mandal[1] and Olivia Rose[#]

[1]The Shri Ram School, India
[#]Advisor

ABSTRACT

The usage of Machine Learning in analyzing and creating art has accelerated substantially over the last couple of years, being used to mass produce novel pieces on a scale that artists would never have imagined feasible previously. This paper will focus on the study of an important Computer Vision tool- Generative Adversarial Networks (GANs) and whether or not it can learn and replicate the visually distinctive, defining characteristics of different forms of visual art, specifically rural Indian art forms. Indian rural art is a rich, underexplored area of visual media that consists of highly stylized, distinctive characteristics. We trained GANs to learn, abstract, and reproduce the visual motifs of five different forms of Indian rural art to understand whether GANs can learn these defining artistic attributes unsupervised and further our understanding of the differences between human intuition and machine perception of art. The textural and feature extraction ability will be analyzed through human perception as well as Fréchet Inception Distance to evaluate the quality of the generated images. This will benefit artists by allowing them to create artworks in unexplored genres, as well as benefit commercial industries, such as home décor and fashion, through the creation of different products in various art forms in an efficient manner.

Introduction

In recent years, deep learning has been proven to be one of the most powerful tools in artificial intelligence and is starting to emerge in daily life. In the context of computer vision, deep learning has been widely used in face recognition, social media filters, street view image recognition, image retrieval, and so on. Another interesting discussion today within deep learning is how it might impact and shape our future cultural and artistic production. However, generating artworks is challenging, especially for Indian rural art forms. We used a modified Deep Convolutional GAN (DCGAN) architecture to train using our dataset of 5 different Indian art forms, to automatically synthesize realistic paintings. (Wang et al., 2017)

Methodology

Data Set

Training data consisted of 1,560 total images from five distinct Indian Rural art forms with 312 images in each form. (Kumar et al., 2018)

Kalamkari

Kalamkari, originating from the southern Indian states of Andhra Pradesh and Telangana, derives its name from the words 'kalam,' meaning pen, and 'kari' meaning craftsmanship. Made on cotton cloth, historically this form was used extensively as decorative backdrops in temples, depicting scenes and stories from sacred Indian texts, such as the Mahabharata and Ramayana. Today new motifs like the 'Tree of Life' – a deep-rooted tree while growing towards the sky, connecting to the earth as well as to the underworld, and a symbol of nourishment for living beings, as well as birds and animals are frequent subjects. Inspiration from Indian forts, palaces, and temples is also common. The colours used are mostly indigo, mustard, rust, black and green, which give it a unique earthy tone. (S. N. Patwardhan, n.d.)

Figure 1. Kalamkari Art

Madhubani

The Madhubani art form, from the eastern state of Bihar, is known for its five different styles. The 'Bharni' art form, painted in bright, vibrant colours, depicts Hindu Gods and Goddesses. The 'Kachni' art form depicts animals, flowers, and other natural things in monochrome, to highlight certain peculiar features. 'Tantrik' style paintings represent traditional texts and are used for specific religious festivals. The 'Godna' style also signifies natural creatures, including the tree of life, and can be seen on paintings as well as tattoos. 'Kobar' paintings depict Hindu wedding ceremonies, primarily made on the walls of the married couple's home. Madhubani paintings use only natural dyes, made from flowers, vegetables, and leaves, usually wilted ones that have fallen from a tree or flowers left over at a temple. The outlines are made with charcoal, mixed with cow dung. (Aishwarya, 2022)

Figure 2. Madhubani Art

Pattachitra

Pattachitra ('patta' means canvas and 'chitra' means painting) art comes mostly from the states of Odisha and West Bengal and is made on dried palm leaves, hand-made canvas cloth or paper, tussar silk, aluminium utensils, showcasing Indian mythology or rural Indian life. A unique format is the 'chitra-pothies'- several palm leaves stacked on top of each other, held together between painted wood covers using a string, with the painting depicting a certain mythological theme. The work is known to be very intricate and uses only natural hand-made bright colours (made from powdered coloured stones, leaves, turmeric, and burning wick), painted with brushes made from the hair of domestic animals. (Mokashi, 2021)

Figure 3. Pattachita Art

Warli

Warli paintings are a tribal art form, from one of the largest tribal regions of forest dwellers in Maharashtra, in western India. These are painted white, using a mixture of rice flour and water, with a bamboo stick that is chewed at the end to give it the texture of a paintbrush. They are painted only to mark special occasions such as weddings, festivals, or harvests. They resemble prehistoric cave paintings and usually depict scenes of human figures engaged in activities like hunting, dancing, sowing, and harvesting, and also elements of nature, using basic geometric shapes. The circle represents the sun and the moon, the triangle depicts mountains and pointed trees, and the square indicates a sacred enclosure or a piece of land. (*Warli Paintings from India*, 2020)

Figure 4. Warli Art

Tanjore

Tanjore or Thanjavur paintings, mostly related to Hindu Gods and Goddesses and episodes from religious texts, originate from the southern state of India – Tamil Nadu. Painted in rich vivid colours involving combinations of red, blue, and green, their key features are the delicate gold or silver foils overlaid on the paintings and an elaborate inlay of glass beads and even precious and semi-precious stones. The paintings are made on canvas pasted (using gum made of tamarind) on a wooden plank (usually from a jackfruit tree or teak), walls, wooden panels, glass, paper, mica, and even ivory. The painting always has the main subject at the centre, surrounded by subsidiary figures or elements like temple arches, birds, and animals. Once complete, it is always framed with a glass panel, for protecting it. (Rajvanshi, 2021)

Figure 5. Tanjore Art

Pre-Processing

All images were resized to a size of 64x64x3 using a transformer for the DCGAN and used three colour channels in training images. The size of the z latent vector is 100 and the size of the feature maps is 64x64.

Figure 6. Training Images

Model

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are composed of two adversarial networks, a generator and a discriminator. It can be formulated as a two-player minimax game. The generator G generates images from random noise, i.e., it learns a mapping from a random noise vector z to an image y, G: z → y. The discriminator D simultaneously learns a mapping from an image x to some values between 0 to 1, which indicates the probability that the input comes from the true data distribution. The goal of the discriminator D is to distinguish samples from the generative model and samples from the training data, while the goal of G is to generate images that are indistinguishable from those from training images (i.e., fool the discriminator D). (Wang et al., 2017)

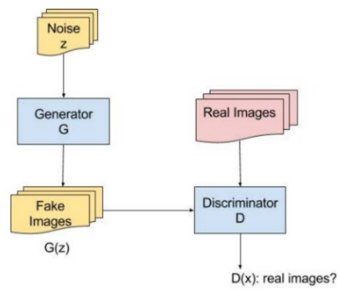

Figure 7. Structure of a GAN (*adapted from:* (Wang et al., 2017)

Deep Convolutional Generative Adversarial Networks

Deep Convolutional Generative Adversarial Networks (DCGANs) (Radford et al., 2015) is an improved version of the original GAN which adds convolutional layers in both generator and discriminator's architecture. The DCGAN is different from the original GAN and has a deeper architecture. The discriminator uses strided convolutions and the generator uses transpose convolutions. Batch normalization used in both the generator and discriminator and fully connected hidden layers have been removed. This makes DCGANs more stable than original GANs and they output sharper images. However both models show instability in terms of mode collapse and gradient vanish. (Wang et al., 2017)

Hyperparameters

We find that the model achieves the best performance with a batch size of 16 for training, a learning rate for the Generator and Discriminator at 0.0003 and 0.0001 respectively, and is trained for 125 epochs. We used transpose convolutions for the generator and strided convolutions for the discriminator with the images undergoing 5 convolutions. The activation function used for the generator is Rectified Linear Unit (ReLu) followed by the Tanh function and Leaky Rectified Linear Unit for the discriminator (Leaky ReLu) followed by the Sigmoid function. Batch normalization is used in both discriminator and generator. The Adam optimizer is used as a loss function.

Limitations

Due to the unavailability of a ready data set for Indian rural paintings, the dataset is self-created and used for training the model after being transformed to 64x64 size. The generated images show a replication in the white spaces as exists in the training images which have been retained during the transformation due to differing sizes. The self-created and minimal nature of the images has resulted in images with a large scope of improvement in terms of replication of stylistic features as well as object recognition and extraction. Increasing the size of the data set will drastically improve results for both the Frechet Inception Distance metric as well as the generated artwork. Moreover, limitations in computational power did not support model training for longer than 150 epochs, although longer training time and further experimentation would likely yield better results.

Results

Generated Artwork

Indian artwork is known to showcase a high spatial frequency and follows the Rasa Theory: anaesthetic theory on which many Indian pieces have been based explains how the viewer, not the artist, is vital in the creation of the art and is used to induce within them a feeling or psychological state, individual to each viewer rather than representing the artist's idea.

Figure 8. Real Images

Figure 9. Fake Images

Our GAN has shown the ability to replicate the textures and colourations of these paintings, keeping with the same theory of Rasa, and also shows an impression of the spatial details in a more stochastic pattern. The GAN was unable to reproduce the various religious motifs in the Kalamkari and Pattachitra art such as faces, bodies, and symbols of the recurring gods and goddesses; however, it has been able to identify this defining feature and create an approximate figuration of the same. The arches of the Indian god paintings are a defining characteristic of Tanjore paintings portraying the gold arches of Indian temple architecture, which can be seen in the generated artwork as well. The GAN failed to capture the circular 'chakra' motifs in the Warli Artwork and fake images show no such pattern. Due to the progressive growth through convolutions in the images, the final output size of the images is relatively small, which is why the traditional GAN architecture is unable to replicate the fine details of Indian art. A limitation of the data set is that the differing sizing has resulted in a white border during transposition which the GAN has identified and replicated; however, it didn't quite learn this feature completely as the border is seen to surround all four sides while it only exists on one or two sides in the generated work.

Model Performance as a Function of Training

Generally, the goal of training in deep learning is to increase the overall accuracy of the model's performance. For instance, the performance of a convolutional neural network (CNN) for image classification should increase in accuracy as a result of training, thereby showing that the network is learning which visual features are useful in performing accurate classification. A method for determining model performance during training is by calculating the model's error, frequently referred to as *loss*, as a function of training time. Loss functions compare the target and predicted output values to measure how well the neural network models the training data. When training a CNN, we aim to minimize this loss between the predicted and target outputs regression settings, therefore showing that the model is increasing in classification accuracy. However, since GAN training aims to synthesize images that successfully fool the discriminator into classifying them as real, we must calculate and balance two competing loss functions: one for training the discriminator, and one for training the generator. For successful GAN training, we aim to minimize the GAN's overall loss when training the discriminator, showing that the discriminator is learning the features that best differentiate real from generated images; however, when training the generator, we aim to maximize the overall loss to demonstrate that the generator is learning to produce more and more realistic features that successfully "trick" the discriminator into classifying artificial images as real. (Dwivedi, 2022) We implemented a simple loss function that calculates the squared difference between the expected and predicted values for each training iteration. As expected, the discriminator loss decreased as a function of training, while the generator loss increased over the course of training. These results demonstrate that the GAN was successfully improving on generating novel images that progressively resembled real samples of Indian rural art.

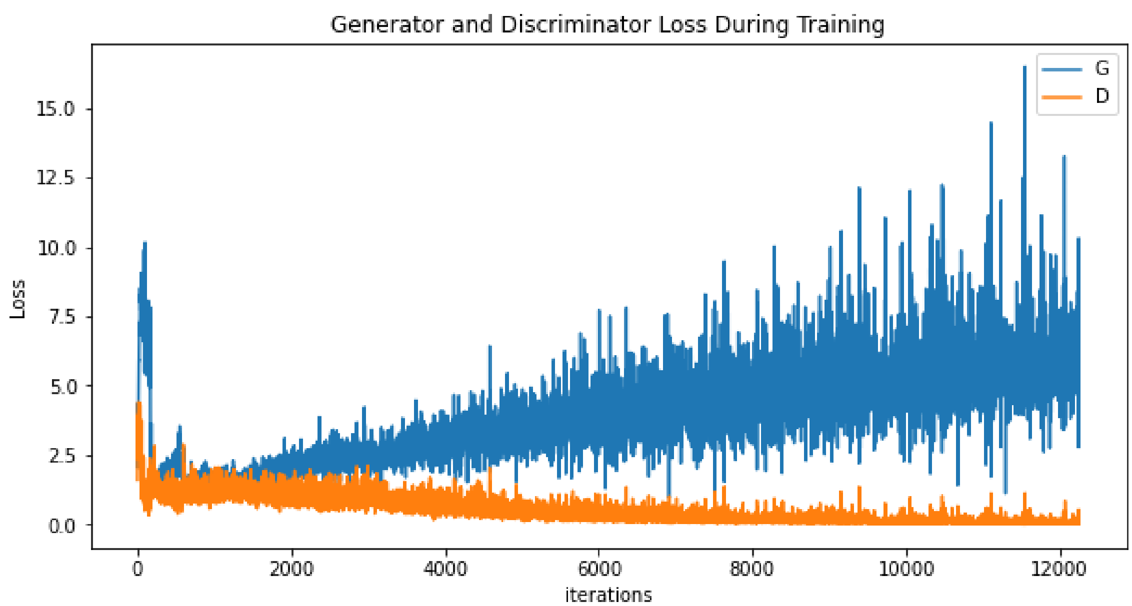

Figure 10. Loss Plot for Generator and Discriminator

Quantitative Analysis: Frechet Inception Distance

In addition to the loss function, GAN performance can be evaluated by calculating the similarity between generated images and real training images. A way to calculate this image similarity is by using the Frechet Inception Distance score (Heusel et al., 2018), also referred to as FID. FID calculates the similarity of the image statistics for real and generated images using Inception v3, a pretrained network for image classification. (Brownlee, 2019a). FID works by feeding real and generated images into the pretrained Inception model, which produces feature vectors for both sets of images. The distance between the distributions of feature vectors for real vs. synthesized images thus becomes the final FID metric; lower FID scores indicate smaller distances between real and generated images, which is interpreted as higher similarity between synthesized and real images. A "perfect" FID score would approach 0 (no distance between the two sets of image statistics), with poor-quality synthetic images receiving FID scores as high as 500 (Heusel et al., 2018).

We report a final FID of 166.265 for our best performing model. Published FID scores from GANs trained on other image sets are frequently reported in the ranges of 2.8 (Karras et al., 2020) to 65.9 (Ostrovski et al., 2018), though almost exclusively resulting from training sets of hundreds of thousands to one million images. The original FID paper (Heusel et al., 2018) shows plots of FID performance over training and under different noise conditions, where the plots typically plateau once they reach around a score of 100-150. Our score is in line with these reported values, particularly with regard to the number of training epochs. The FID does show a drastic improvement in score when the training epochs are increased, there was a jump from the previous score of 206.983 for 100 epochs to our current score of 166.265 only by training for 25 epochs more.

Conclusion

The GAN shows a reliance on texture and pattern features of the paintings to generate the output as compared to the high spatial frequency of the paintings. The model does not show object replication and is instead able to mimic the various colouration styles. The Rasa Theory, the aesthetic theory on which many Indian pieces have been based, explains how the viewer, not the artist, is vital in the creation of the art and is used to induce within them a feeling or psychological state, individual to each viewer rather than representing the artist's idea as is seen in Western art forms. Our GAN managed to bring in these differentiating features of the works and portray the concepts vital to Indian art by focussing on motifs, patterns, and colour theory of the artistic styles. As can be seen from the Loss Plot *(Figure 10)*, the discriminator loss decreased as a function of training, while the generator loss increased over the course of training. These results demonstrate that the GAN was successfully improving on generating novel images that progressively resembled real samples of Indian rural art. Due to the limited dataset and available computing power, we were only able to report a score of 166.265 in terms of Frechet Inception Distance although the model showed a drastic improvement with longer training periods, a previously tested batch that ran for 100 epochs and reported a score of 206.983 and there is a huge jump seen by just increasing the epochs by 25. Our results are in line with the original FID paper which shows a typical plateau near ~100-150.

Future Exploration

So far, we have seen how well the DCGAN trained on a dataset of Indian Rural artwork for a limited training frame and analyzed the accuracy of the model in terms of the qualitative features of the original images and the generated images as well as the Frechet Inception Distance. However, due to the limitations in resources and possible architectural inadequacies for this particular kind of data, a lot more research can be done to further improve the generated results and expand the potential applications of the same.

1. Training on a larger dataset and for a longer training period would generate better results as the limited computing power and self-created nature of the dataset is a huge hindrance in the scope of improvement of the results.

2. A potential issue that can be resolved in these images is the occurrence of checkerboard artefacts, a strange repetitive pattern like a checkerboard in the images. Replacing the type of layers used during deconvolution as well as through upsampling.
3. Outputs of generative models could be vastly different for different models and different generations of models as well. Conducting a comparative analysis of how different models extract features and produce an output would be a promising direction to investigate.

Investigating these areas would be crucial to yield better results. Furthermore, expanding this generative technique to other culturally significant art forms would be an interesting way to explore and generate beautiful and rich artwork, and a fascinating study of the same by analyzing the machine interpretation vs the human perception of these works by focussing on the extracted features of the paintings. Additionally, interpolation of various art forms from distinct cultures and countries can be used to generate hybrid art by taking points in the latent space that lies between different works, which might prove to be a way to integrate these diverse cultures into a new media worth exploring.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Aishwarya. (2022). *Types of Madhubani Painting*. https://authindia.com/types-of-madhubani-painting/

Brownlee, J. (2019a). *How to Implement the Frechet Inception Distance (FID) for Evaluating GANs*. https://machinelearningmastery.com/how-to-implement-the-frechet-inception-distance-fid-from-scratch/

Castellano, G., & Vessio, G. (2021). Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview. *Neural Computing and Applications*, *33*(19), 12263–12282. https://doi.org/10.1007/s00521-021-05893-z

Dwivedi, H. (2022). *Understanding GAN Loss Functions*. https://neptune.ai/blog/gan-loss-functions

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. https://doi.org/10.48550/ARXIV.1406.2661

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium* (arXiv:1706.08500). arXiv. http://arxiv.org/abs/1706.08500

Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks* (arXiv:1812.04948). arXiv. http://arxiv.org/abs/1812.04948

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). *Analyzing and Improving the Image Quality of StyleGAN* (arXiv:1912.04958). arXiv. http://arxiv.org/abs/1912.04958

Kumar, S., Tyagi, A., Sahu, T., Shukla, P., & Mittal, A. (2018). Indian Art Form Recognition Using Convolutional Neural Networks. *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 800–804. https://doi.org/10.1109/SPIN.2018.8474290

Mokashi, A. (2021). *Pattachitra: The Heritage Art of Odisha*. https://www.livehistoryindia.com/story/living-culture/pattachitra

Ostrovski, G., Dabney, W., & Munos, R. (2018). *Autoregressive Quantile Networks for Generative Modeling* (arXiv:1806.05575). arXiv. http://arxiv.org/abs/1806.05575

Patwardhan, S. N. (n.d.). *KALAMKARI: A TRADITIONAL INDIAN ART FORM*. https://laasyaart.com/kalamkari-a-traditional-indian-art-of-perfection/

Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. https://doi.org/10.48550/ARXIV.1511.06434

Rajvanshi, N. (2021). *Tanjore Paintings: A Living Legacy*. https://www.livehistoryindia.com/story/art-history/tanjore-paintings

Wang, G., Chen, Y., & Chen, Y. (2017). *Chinese Painting Generation Using Generative Adversarial Networks*. http://cs231n.stanford.edu/reports/2017/pdfs/311.pdf

Warli paintings from India. (2020). https://indigoarts.com/galleries/warli-paintings-india