Comparative Analysis of Generative Adversarial Network (GAN)

Simran Saini, Himanshu, Aaryan Bhardwaj, Siddharth Singh Sud
Computer Science, BML Munjal University, Gurugram

ABSTRACT

This project explores the capabilities of adversarial training and cycle consistency constraints in unpaired image-to-image translation tasks, focusing on transforming images between two distinct domains: horses and zebras. Inspired by the success of CycleGAN, we aim to demonstrate the potential for realistic image transformations across diverse visual domains. Furthermore, leveraging the ChestX-ray dataset, which includes labeled images categorized as "NORMAL" and "PNEUMONIA," also we have Vanilla GAN which is implemented on brain MRI images for brain tumor detection dataset .We extend our investigation to the realm of medical imaging. Here, we employ an Auxiliary Classifier Generative Adversarial Network (ACGAN) architecture to generate chest X-ray images, offering insights into potential applications in medical image synthesis. The generator model architecture comprises multiple convolutional layers, supplemented by transpose convolutional layers. These layers work together to up sample the input noise vector, ultimately producing high-fidelity and realistic images representative of the target domain.

Through this research, we showcase the efficacy of adversarial training techniques and cycle consistency constraints not only in artistic image synthesis but also in the generation of medical images, thereby highlighting the versatility and potential impact of such approaches across various domains.

I. INTRODUCTION

A Generative Adversarial Network (GAN) is a type of artificial intelligence algorithm composed of two neural networks, the generator and the discriminator, which are trained together in a competitive manner. The generator generates synthetic data, such as images, while the discriminator evaluates the authenticity of the generated data. Through iterative training, the generator learns to produce increasingly realistic data, while the discriminator becomes more adept at distinguishing between real and fake data. GANs have demonstrated remarkable success in generating high-quality synthetic data across various domains, including images, audio, and text. Generative Adversarial Networks (GANs) operate on a fundamentally adversarial principle, comprising two neural networks—the generator and the discriminator—that engage in a competitive learning process. The generator is tasked with synthesizing data, such as images, from random noise, aiming to produce samples that are indistinguishable from real data. On the other hand, the discriminator acts as a binary classifier, distinguishing between real and fake data. During training, the generator generates fake samples, and the discriminator evaluates their authenticity. The discriminator provides feedback to the generator, guiding it to produce more realistic samples. Simultaneously, the discriminator updates its parameters to become more proficient at distinguishing real from fake data. This adversarial interplay continues iteratively, with both networks improving over time through a process akin to a minimax game, where the generator aims to minimize the discriminator's ability to differentiate between real and fake data while the discriminator strives to accurately classify the samples. As training progresses, the generator learns to generate increasingly realistic samples, converging towards a distribution that closely resembles the true data distribution. The ultimate goal of a GAN is to produce synthetic data that is indistinguishable from real data, capturing the underlying patterns and structure of the dataset it was trained on. This process enables GANs to generate high-quality synthetic data across various domains, facilitating tasks such as image synthesis, data augmentation, and anomaly detection.

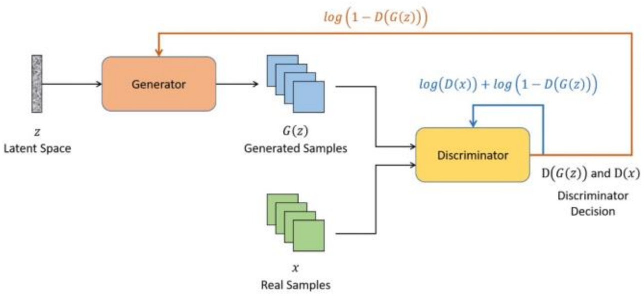

Fig 1: Working architecture of GAN

In the realm of artificial intelligence and computer vision, the ability to translate images between different visual domains has garnered significant attention and acclaim. This endeavour, known as image-to-image translation, holds promise for a myriad of applications, ranging from artistic expression to medical imaging. One prominent approach in this domain is the CycleGAN framework, which exemplifies the power of adversarial training and cycle consistency constraints in achieving realistic transformations between disparate visual domains.

The core idea behind CycleGAN lies in its ability to learn mappings between two domains without the need for paired data, thereby obviating the laborious task of collecting matched image pairs for training. Instead, it leverages unpaired images from each domain and employs adversarial training to learn the mapping functions implicitly. This methodology has been demonstrated to yield impressive results, enabling the creation of compelling transformations between diverse domains such as horses and zebras. Beyond artistic endeavours, the potential applications of image-to-image translation extend into the realm of medical imaging, where the ability to generate realistic synthetic images holds immense value. In this context, the ChestX-ray8 dataset provides a rich source of labelled chest X-ray images, offering an opportunity to explore the synthesis of medical imagery using advanced deep learning techniques. Leveraging this dataset, our project delves into the synthesis of chest X-ray images, aiming to generate both "NORMAL" and "PNEUMONIA" labelled images for diagnostic purposes.

To accomplish this task, we employ an Auxiliary Classifier Generative Adversarial Network (ACGAN) architecture, which extends the traditional GAN framework by incorporating an auxiliary classifier into the discriminator. This addition facilitates not only the generation of realistic images but also the control over specific attributes or classes within the generated samples. By incorporating class information into the discriminator, the ACGAN framework enables more precise control over the image generation process, ensuring that the synthesized images adhere to desired characteristics. Central to our approach is the generator model, which is tasked with translating noise vectors into realistic images. This model architecture comprises several convolutional layers followed by transpose convolutional layers, facilitating the upsampling of the input noise vector to generate high-resolution images. Inspired by the success of CycleGAN, our model architecture is built upon similar principles, with two generator networks (one for each domain) and two discriminator networks.

The generators are responsible for translating images from one domain to the other, while the discriminators differentiate between real and generated images, providing feedback to the generators to improve their performance iteratively. To enhance the stability and effectiveness of training, we employ UNet-based architectures for both generators and discriminators, leveraging the inherent skip connections to facilitate information flow between different layers. Additionally, instance normalization is utilized to stabilize the training process, ensuring consistent and reliable convergence of the model.

In summary, our project endeavours to harness the capabilities of advanced deep learning architectures, namely ACGAN and CycleGAN, for the purpose of unpaired image-to-image translation. By extending these frameworks to the domain of medical imaging, we aim to demonstrate their efficacy in synthesizing realistic chest X-ray images, thereby potentially contributing to advancements in diagnostic imaging and healthcare.

II. METHODS

Data Collection and Preprocessing:

Chest X-ray images that were collected from the ChestX-ray8 dataset, containing images labeled as "NORMAL" and "PNEUMONIA".

The images were preprocessed by resizing them to a uniform size of 64x64 pixels and converting them from BGR to RGB color format. Additionally, pixel values were normalized to the range [0, 1]. For the CycleGAN we used the "horse2zebra" dataset from TensorFlow Datasets, consisting of unpaired images of horses and zebras. For Vanilla GAN we have used MRI brain images from Kaggle.

These datasets have different images which are trained further as you will get to know in the methodology part. The picture in fig 2 shows various GAN's used in three datasets.

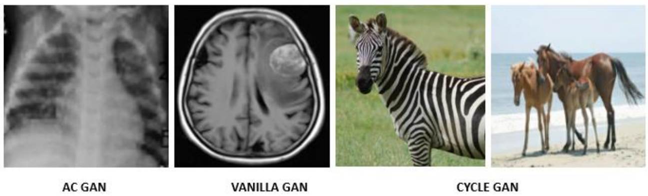

Fig 2: Samples of images of datasets

Model Architecture:

An Auxiliary Classifier Generative Adversarial Network (ACGAN) architecture was utilized to generate chest X-ray images. The generator model comprised multiple convolutional layers followed by transpose convolutional layers, enabling the upscaling of the input noise vector to produce realistic images. Conversely, the discriminator model employed convolutional layers to discern between real and generated images, incorporating an auxiliary classifier to predict class labels (NORMAL or PNEUMONIA).

During training, the ACGAN model employed the RMSprop optimizer with distinct learning rates for the discriminator (0.0001) and the generator (0.00005). Training involved alternating steps between the discriminator and generator, where the former was trained to differentiate between real and fake images, while the latter aimed to generate convincing images to deceive the discriminator. The training regimen spanned 3000 epochs with a batch size of 32, with early stopping implemented to mitigate overfitting.

The project employs the "horse2zebra" dataset sourced from TensorFlow Datasets, comprising unpaired images of horses and zebras. Its model architecture draws inspiration from CycleGAN, featuring two generator networks (one per domain) and two discriminator networks. Generators facilitate image translation across domains, while discriminators discern between real and generated images. Both generators and discriminators adopt UNet-based architectures, with instance normalization to foster training stability.

Various loss functions drive the model's optimization. Adversarial loss incentivizes generators to generate convincing images that deceive discriminators. Cycle consistency loss ensures the coherence of translation cycles (e.g., horse to zebra and back to horse) with the original image. Identity loss encourages generators to maintain input image content when the target domain remains the same.

Training unfolds with batches of unpaired images from horse and zebra domains. It involves alternating optimization steps for generators and discriminators. Generators strive to minimize adversarial and cycle consistency losses, while discriminators refine their ability to discriminate between real and generated images. Periodic saving of checkpoints throughout training facilitates model resumption and evaluation, ensuring robustness and progress monitoring. This comprehensive approach enables the model to learn effective mappings between domains and generate realistic translations, contributing to advancements in unpaired image-to-image translation tasks.

For Vanilla GAN, the architecture comprises two key components: the generator and the discriminator, which together form a Generative Adversarial Network (GAN). The generator takes random noise as input and transforms it through a series of layers, including dense and transposed convolutional layers, to generate images. The generator's objective is to produce images that are indistinguishable from real ones. On the other hand, the discriminator acts as a binary classifier, distinguishing between real images from the dataset and fake images generated by the generator. It employs convolutional layers followed by flattening and dense layers to perform classification. Through adversarial training, where the generator tries to fool the discriminator while the discriminator tries to distinguish real from fake, both networks improve iteratively.

During training, the generator and discriminator are optimized in tandem but in opposite directions. The generator aims to minimize the discriminator's ability to distinguish fake images, while the discriminator aims to correctly classify real and fake images. This adversarial process fosters competition between the two networks, driving them to improve over time. Additionally, the training loop involves sampling random noise, generating fake images, and updating both networks' parameters based on their performance in discriminating between real and fake images. This architecture enables the generation of synthetic data that closely resembles the real data distribution, with applications spanning image synthesis, data augmentation, and anomaly detection, among others.

III.RESULTS AND DISCUSSION

A. Evaluation of the trained model for X-rays part encompassed various metrics, including accuracy, precision, recall, and F1-score, alongside the Fréchet Inception Distance (FID) score to assess image quality and diversity. Subsequent validation on an independent dataset gauged the model's generalization capacity, while testing on a separate dataset containing generated images validated its ability to produce realistic chest X-ray images.

B. To further assess image fidelity, a pre-trained VGG16 model underwent fine-tuning using the generated images and associated labels. This fine-tuned model underwent training for 60 epochs, with early stopping based on validation loss. Performance analysis involved the examination of confusion matrices, classification reports, and evaluation metrics, offering insights into the model's accuracy in image classification and generation.

C. Post-training evaluation of the model for Cycle GAN involves assessing its performance on a subset of test images sourced from the horse domain. Leveraging the trained generator, horse images are transformed into zebra-like representations, with the resultant images subjected to visual scrutiny. By comparing the translated images with their corresponding input counterparts, the visual inspection allows for a qualitative evaluation of the model's ability to produce realistic translations.

Through this result visualization process, the CycleGAN model's capability to seamlessly bridge the gap between the horse and zebra domains becomes evident. The translated images exhibit compelling resemblances to authentic zebra images, showcasing the model's ability to capture and replicate key visual characteristics inherent to zebras. This qualitative evaluation underscores the model's proficiency in image translation, affirming its capacity to generate convincing and visually appealing transformations.

In conclusion, the Cycle GAN framework emerges as a potent tool for unpaired image-to-image translation tasks, demonstrating its adeptness at learning domain mappings and generating faithful translations without the need for paired training data. The visually plausible translations produced on previously unseen test images underscore the model's effectiveness and highlight its potential for a myriad of practical applications in various domains.

For Vanilla GAN, the generator and discriminator models used in this architecture are relatively simple, consisting of a few convolutional layers with Leaky ReLU activations and a final layer with a tanh activation in the generator for image generation and a sigmoid activation in the discriminator for binary classification. These models may produce satisfactory results on simple datasets, but for more complex tasks or higher-resolution images, more sophisticated architectures and training strategies might be necessary to achieve better results. To evaluate the performance of the GAN and assess the quality of the generated images, metrics such as the Inception Score (IS), Frechet Inception Distance (FID) is done.

D. Research

These are some of related works done by various authors in field of Generative Adversarial Network.

TABLE I

Literature survey		
Author	**Approach**	**Results**
Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros: "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" (CycleGAN)	CycleGAN, using Summer-to-Winter dataset where which can translate images from one domain to another without paired examples. They achieved impressive results in various image translation tasks, such as style transfer, object transfiguration, and season transfer.	90%
Augustus Odena, Christopher Olah, and Jonathon Shlens	Used CIFAR-10 and CIFAR-100 these datasets for DCGAN to generate images conditioned on specific class labels, making it suitable for tasks like image-to-image translation, image super-resolution.	88%

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al	Introduced StarGAN, a framework for multi-domain image-to-image translation. StarGAN can translate images across multiple domains while preserving the original content. It achieves state-of-the-art results in tasks like facial attribute manipulation, where it can change facial attributes (e.g., age, gender, and expression) in images.	91%

C. Figures

GAN's	FID SCORE
Vanilla Gan	40.845
ACGAN	15.2875
CycleGAN	34.561

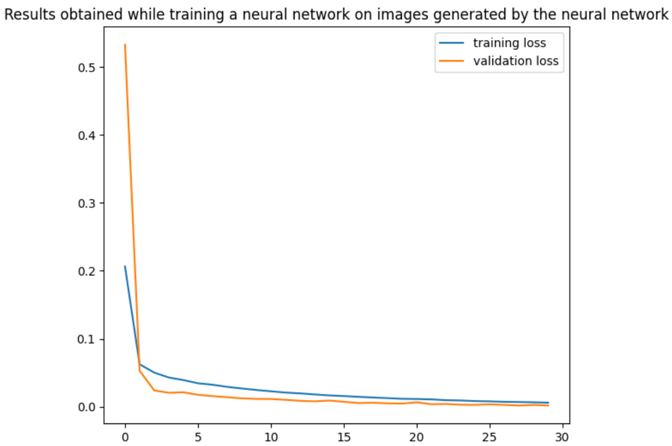

Fig 3: Train and Validation loss of ACGAN

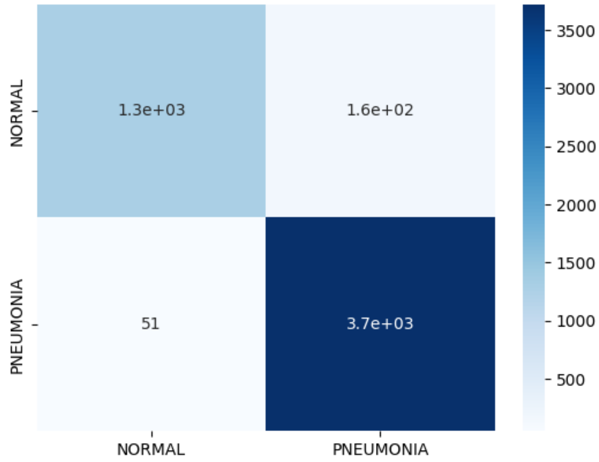

Fig 4: Confusion Matrix of ACGAN

D. Cycle GAN Results

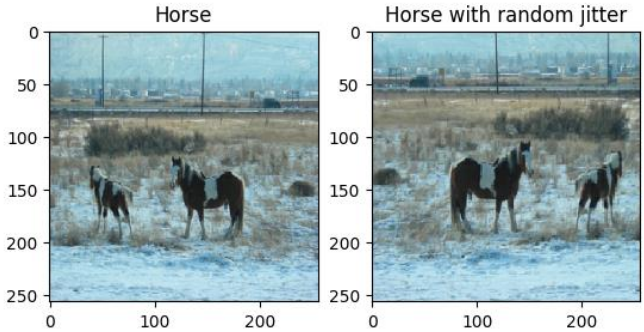

Fig 5: Image with random jitter

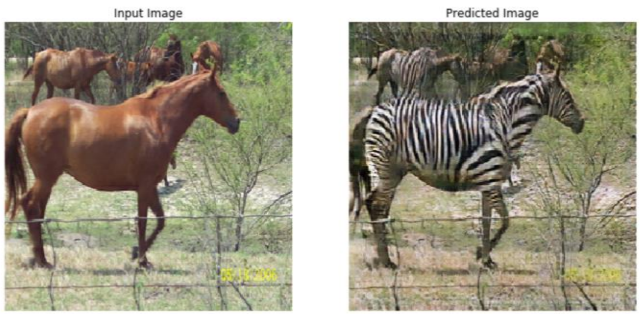

Fig 6: Transformed with CycleGAN

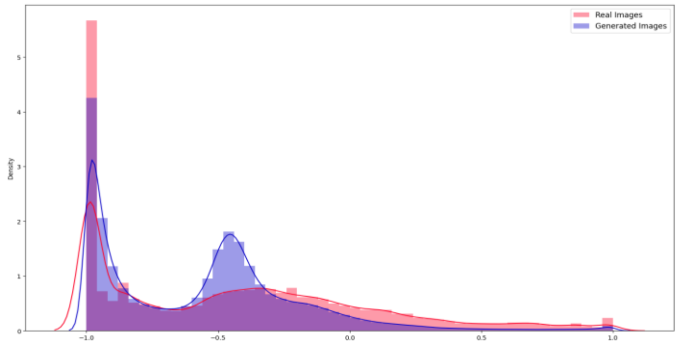

Fig 7: Comparing the generated images with the real samples by plotting their distributions. If the distributions overlap, that indicates the generated samples are very close to the real ones

IV. CONCLUSION

In this paper, Cycle GAN framework showcases remarkable capabilities in unpaired image-to-image translation, exemplified by its successful translation of images between the horse and zebra domains without the requirement for paired training data. Through qualitative evaluation, the model demonstrates its ability to generate visually plausible translations, capturing essential visual characteristics and producing compelling transformations. This underscores Cycle GAN's efficacy in learning domain mappings and generating realistic images, thus presenting a valuable tool for diverse applications, from artistic expression to domain adaptation in computer vision tasks.

Similarly, the ACGAN architecture proves to be a formidable approach for generating chest X-ray images, leveraging adversarial training to synthesize realistic medical imagery. With its auxiliary classifier enhancing control over generated images' attributes, ACGAN effectively produces chest X-ray images with discernible class labels (NORMAL or PNEUMONIA). Through rigorous training and evaluation, ACGAN showcases its potential in medical image synthesis, offering a pathway towards generating high-fidelity medical images for diagnostic and research purposes. Both CycleGAN and ACGAN represent significant advancements in the field of generative adversarial networks, promising avenues for further exploration and application in various domains.

V. REFERENCES

[1] 1. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

[2] . Odena, A., Olah, C., & Shlens, J. (2017). Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 2642-2651). JMLR. org.

[3] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[4] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8789-8797).