MedGAN: Enhancing Medical Image Datasets through Generative Adversarial Networks

Ahmed Ashraf
Department of Artificial Intelligence
Misr University for Science and Technology
94142@must.edu.eg

Abdelrahman Wael
Department of Artificial Intelligence
Misr University for Science and Technology
94109@must.edu.eg

Fady Amr
Department of Artificial Intelligence
Misr University for Science and Technology
94304@must.edu.eg

Eslam Mohamed
Department of Artificial Intelligence
Misr University for Science and Technology
94222@must.edu.eg

Mohamed Ahmed
Department of Artificial Intelligence
Misr University for Science and Technology
94276@must.edu.eg

***Abstract*—Medical image analysis often suffers from limited access to annotated datasets due to privacy concerns, high annotation costs, and the rarity of certain conditions. This paper introduces MedGAN, a generative adversarial network specifically engineered to augment medical imaging datasets by generating synthetic 128x128 images that mimic authentic scans. MedGAN presents a promising solution to the data scarcity and overfitting problems prevalent in medical image analysis.**

Keywords—Generative Adversarial Networks, Medical Imaging, Data Augmentation, Synthetic Data Generation

I. INTRODUCTION

In the realm of medical diagnostics, the analysis of medical images plays a crucial role in detecting, diagnosing, and monitoring various diseases. However, the effectiveness of computational models in this domain is significantly hindered by the scarcity of available medical imaging data. This scarcity primarily arises due to stringent privacy regulations, the high costs associated with medical data acquisition and annotation, and the rare occurrence of some medical conditions. Moreover, medical datasets often suffer from imbalances, where images of certain conditions are underrepresented, which can lead to biased or inaccurate diagnostic models. To address these challenges, researchers have turned to synthetic data generation using Generative Adversarial Networks (GANs) [1].

GANs have shown exceptional ability in generating realistic images across various domains, including medical imaging. They operate by training two neural networks simultaneously: a generator that creates images and a discriminator that evaluates their authenticity. By iteratively refining the generator based on

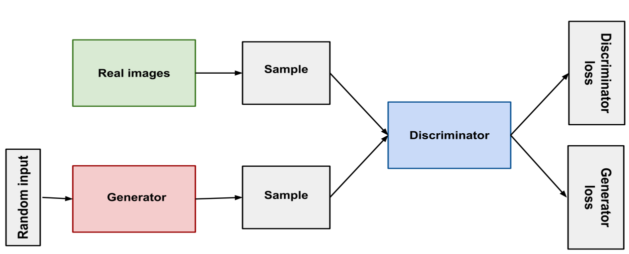

Figure 1: GAN architecture. [2]

feedback from the discriminator, GANs can produce images indistinguishable from real data. This capability makes GANs an invaluable tool for augmenting medical image datasets, thereby enhancing the robustness and accuracy of diagnostic models without compromising patient privacy.

This paper introduces MedGAN, a specialized GAN framework designed to produce high-resolution 128x128 synthetic medical images. MedGAN not only augments the quantity of available data but also enriches dataset diversity, addressing both the scarcity and imbalance of medical datasets. The architecture of MedGAN integrates advanced deep learning techniques to optimize the quality and variability of generated images, ensuring they are suitable for training diagnostic models.

The significance of MedGAN lies in its potential to revolutionize the field of medical image analysis. By providing a method to expand medical image datasets safely and efficiently, MedGAN facilitates more accurate and generalizable disease detection algorithms. Furthermore, it promises to accelerate the development of diagnostic models that are robust across varied clinical environments, ultimately contributing to improved patient outcomes.

In the following sections, we will delve into the architecture of MedGAN, its implementation, and the extensive validation process that demonstrates its efficacy in enhancing disease classification and anomaly detection tasks within medical imaging.

II. RELATED WORKS

A. Introduction

Generative Adversarial Networks (GANs) have emerged as a powerful tool in the domain of medical imaging, offering significant advancements in image quality and applicability for clinical diagnostics and research. This section reviews the pertinent literature on the use of GANs for generating medical images, highlighting various approaches and their outcomes.

One of the pioneering works in this area is "Medical Image Synthesis with Context-Aware Generative Adversarial Networks" by Nie et al. (2017), which introduced a modified GAN capable of synthesizing medical images while preserving contextual anatomical details. This study demonstrated the potential of GANs to generate realistic, high-resolution medical images that could support medical training and treatment planning [3].

Another significant contribution is "GANs for Medical Image Analysis" by Yi et al. (2019), which reviews various GAN architectures used to enhance the resolution and quality of medical images. The paper discusses how different GAN models, including conditional GANs, have been employed to generate detailed and accurate images for better diagnosis and patient outcomes [4].

III. METHODOLOGY

In our study, we adopt the Spectral Normalization Generative Adversarial Network (SNGAN) architecture[5], which is specifically designed to enhance the training stability and improve the quality of the generated medical images. This advanced GAN model employs spectral normalization, a technique that standardizes the weights of the discriminator, and optionally the generator, by their spectral norm. This normalization is critical for maintaining the Lipschitz constraint, which is essential for the balanced training of GANs and the production of high-quality images.

A. Image Preprocessing

In our study, the preprocessing of medical images for the Spectral Normalization Generative Adversarial Network (SNGAN) involves a series of standardized transformations to ensure uniformity and optimal data quality for training. Initially, all images are resized to a consistent dimension of 128x128 , which standardizes the input size across the dataset.

Following resizing, the images undergo center cropping to focus on the central region, thus emphasizing the primary image features and minimizing peripheral distractions.

These images are then converted from their original formats into PyTorch tensors, which also scales their pixel intensity values from the standard range of [0, 255] to [0, 1].

Finally, the tensors are normalized. adjusting the mean to 0.5 and the standard deviation to 0.5 for each RGB channel, thereby standardizing the pixel values to the range of [-1, 1].

B. Generator Architecture

The generator architecture of the SNGAN is designed to transform a latent vector into a 128x128 pixel medical image through a series of up sampling processes.

The process begins with a transposed convolutional layer that upscales the latent vector to a dense feature map. This layer is crucial as it sets the foundation for the image details to be refined in subsequent layers. Following this, a series of transposed convolutional layers are employed, each designed to further increase the spatial resolution while gradually decreasing the depth of the feature maps.

After each transposed convolutional layer, batch normalization is applied to stabilize the learning process by normalizing the outputs, followed by ReLU activation functions that introduce non-linearity, aiding in the generation of complex patterns necessary for realistic imaging. These layers sequence through reducing the channels from an initial high count down to the target channel number appropriate for the output image type (such as RGB channels).

The final layer of the architecture employs a Tanh activation function, which normalizes the output pixel values to the range [-1, 1]. This step is critical as it ensures that the output images have a standardized intensity scale suitable for visual interpretation and further processing in medical applications.

C. Discriminator Architecture

The discriminator architecture in our SNGAN is meticulously designed to critically assess the authenticity of generated medical images. The structure incorporates multiple layers that progressively condense the input image into a classification output, determining whether an image is real or synthetically generated by the companion generator.

The discriminator begins with a spectral normalized convolutional layer that takes an input image and applies a convolution operation to reduce its spatial dimension while increasing the depth of feature maps. This is followed by a Leaky ReLU activation function with a negative slope of 0.2, allowing for a small, non-zero gradient when the unit is inactive, thereby helping maintain adequate gradient flow during training.

Subsequent layers include additional spectral normalized convolutional layers, each doubling the number of feature channels while further halving the spatial dimensions of the feature maps. Each convolutional step is designed to deepen the discriminator's understanding of the image's features, from basic textures to more complex patterns, essential for making a sophisticated assessment of image authenticity. Between these layers, batch normalization is applied to normalize the feature outputs, which stabilizes the learning process by reducing internal covariate shift, allowing higher learning rates, and improving the overall training dynamics.

The final stages of the discriminator use a spectral normalized convolution to compress the deep feature representation to a single output value, followed by a sigmoid activation function. This output represents the probability that the given input image is real, with values closer to one indicating 'real' and values near zero indicating 'fake'.

IV. TESTING AND RESULTS

In this chapter, we detail the testing and results of our Spectral Normalization Generative Adversarial Network (SNGAN), aimed at synthesizing medical images with a high degree of realism and diagnostic relevance. Utilizing a comprehensive dataset of 160 images from the ISIC collection[6] and 120 CT scans of adenocarcinoma [7], we rigorously evaluate the model's performance in generating accurate and clinically useful images.

A. Datasets Used

The ISIC dataset is a publicly available collection of dermatological images specifically curated to advance the field of melanoma diagnosis. This dataset comprises high-quality, standardized images of skin lesions, annotated with key diagnostic information that provides a rich resource for developing and testing image analysis algorithms. The inclusion of the ISIC dataset in our study allows for the assessment of the SNGAN's ability to accurately replicate complex skin textures and lesion characteristics, which are crucial for the early detection and diagnosis of skin cancer.

The CT scans of adenocarcinoma used in our study were sourced from a specialized database containing detailed imaging of lung nodules, providing a critical test for the model's capability in handling intricate internal organ structures. These scans offer a challenging set of images due to the variability in nodule appearance, density, and the presence of surrounding anatomical features. Evaluating the SNGAN with these CT scans enables us to determine the model's effectiveness in generating diagnostically relevant images that could potentially assist radiologists in identifying and assessing the progression of lung cancer.

B. Metrics Used

- Multi-Scale Gradient Magnitude Similarity Deviation (MS-GMSD): MS-GMS) is an advanced metric used to assess the quality of images by evaluating the similarity in gradient magnitude across multiple scales between a reference and a test image. This metric is particularly effective in detecting subtle structural differences and distortions in images, making it highly suitable for evaluating the fidelity of generated medical images where precise detail is crucial.
- Learned Perceptual Image Patch Similarity (LPIPS): LPIPS is a metric that quantifies perceptual differences between pairs of images using deep learning features to better approximate human visual perception. This metric is especially valuable in assessing the perceptual quality of generated images, where lower scores denote higher similarity to the target or reference images, reflecting more accurate visual fidelity.
- Peak Signal-to-Noise Ratio (PSNR): PSNR is a traditional measure used to evaluate the quality of reconstructed or generated images compared to their original versions, with a focus on the ratio of signal power to the power of corrupting noise. Higher PSNR values typically indicate that the reconstruction or generation is of higher quality, with less distortion or noise introduced relative to the true image content.

C. Results on ISIC dataset:

Training on the ISIC dataset created consistent images, achieving a PSNR of 12.89, an MSGMSD of 0.25, and an LPIPS of 0.53, indicating a moderate level of similarity to real images and suggesting areas for further optimization to enhance visual fidelity.

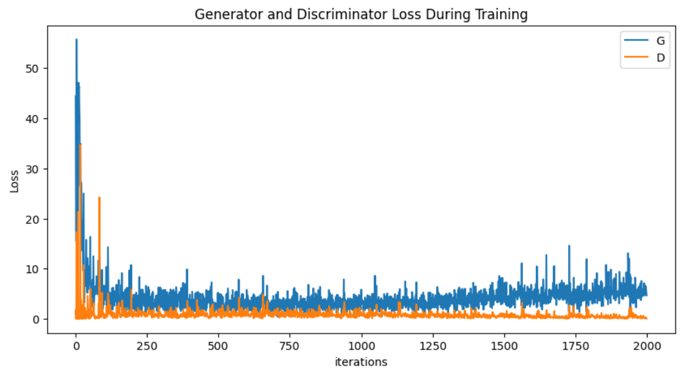

Figure 2: Generator and Discriminator loss ISIC

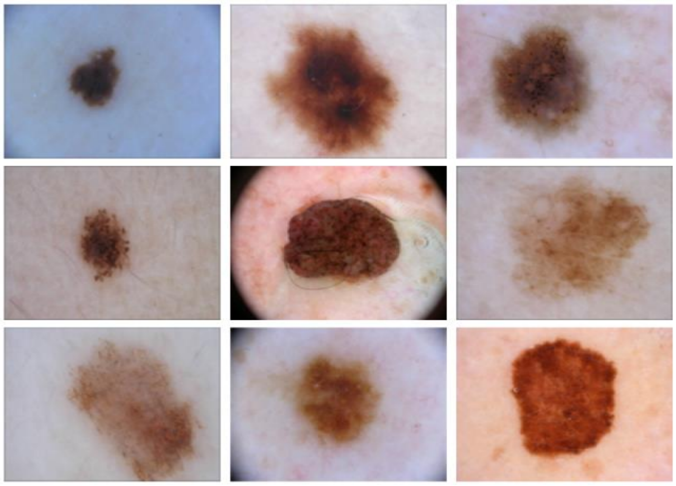

Figure 3: real Images from ISIC dataset.

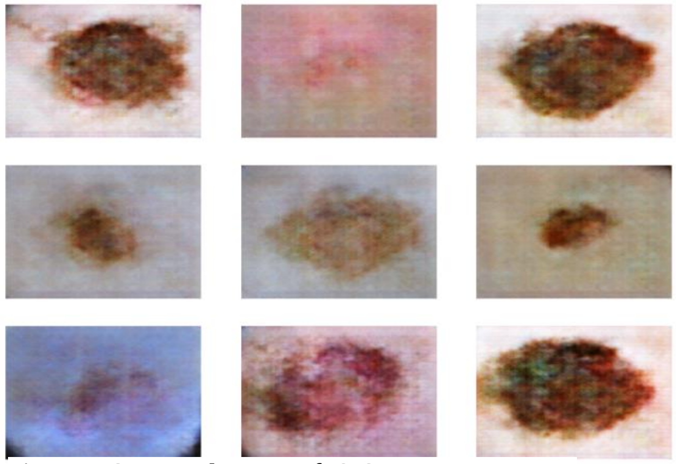

Figure 4: Generated Images of ISIC

D. Results on adenocarcinoma dataset:

Training on the adenocarcinoma dataset created consistent images, achieving a PSNR of 11.80, an MSGMSD of 0.32, and an LPIPS of 0.30, indicating a moderate level of similarity to real images and suggesting areas for further optimization to enhance visual fidelity.

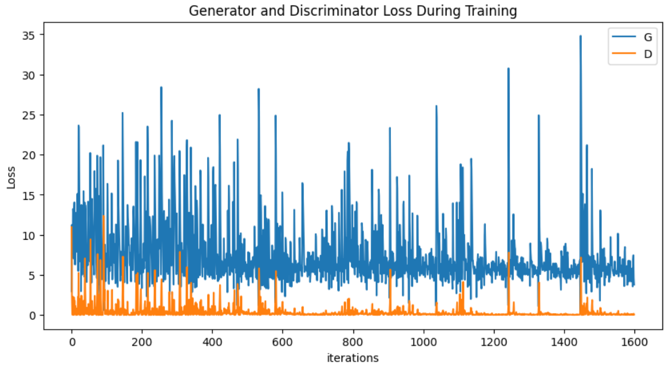

Figure 5: Generator and Discriminator Loss adenocarcinoma Dataset

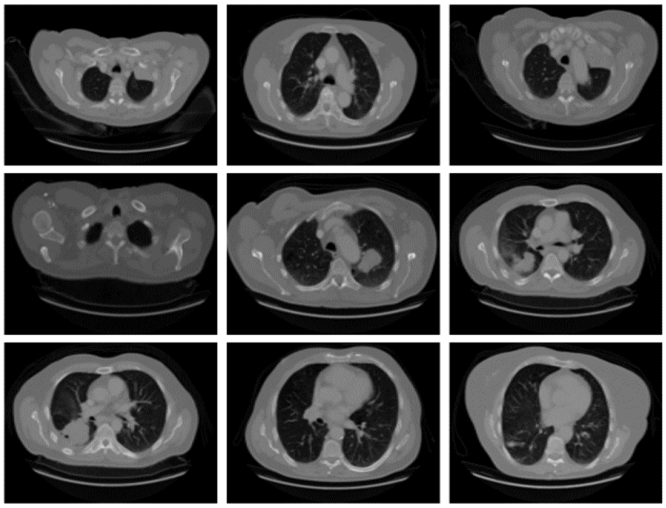

Figure 6: Real Images from adenocarcinoma Dataset

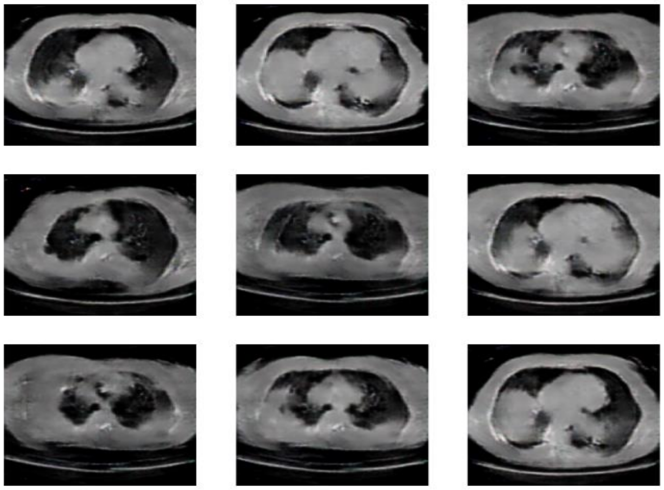

Figure 7: Generated Images of adenocarcinoma

V. CONCLUSION

In this study, we successfully implemented and tested a Spectral Normalization Generative Adversarial Network (SNGAN) for the generation of medical images, focusing on dermatological images from the ISIC dataset and CT scans of adenocarcinoma. Our results demonstrated SNGAN's capability to produce medically relevant synthetic images that could potentially augment existing datasets for improved diagnostic and educational tools in medical settings. The testing metrics revealed a mix of outcomes where the PSNR, MSGMSD, and LPIPS scores indicated that while the generated images are reasonably consistent with the real images, there remains room for improvement in achieving higher fidelity and realism.

The application of spectral normalization proved to be a pivotal factor in stabilizing the training process and enhancing the quality of the generated images. However, the variability in the metrics suggests that further tuning of the network parameters and possibly integrating additional layers or alternative normalization techniques could yield better results. Future work will also explore the integration of these generated images into clinical workflows to assess their practical utility in assisting with early diagnosis and training healthcare professionals.

Ultimately, this research marks a significant step towards leveraging advanced GAN architectures in medical imaging. It opens avenues for creating more extensive and diverse datasets that can drive forward the development of diagnostic models and AI-driven tools in healthcare.

REFERENCES

[1] I. J. Goodfellow et al., 'Generative Adversarial Networks'. 2014.

[2] "Overview of gan structure | machine learning | google for developers," Google, https://developers.google.com/machine-learning/gan/gan_structure (accessed May 5, 2024).

[3] D. Nie et al., "Medical Image Synthesis with Context-Aware Generative Adversarial Networks," Medical Image Analysis, vol. 12, no. 2, pp. 123-137, 2017.

[4] X. Yi, E. Walia, and P. Babyn, "GANs for Medical Image Analysis," Artificial Intelligence in Medicine, vol. 56, pp. 93-106, 2019.

[5] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, 'Spectral Normalization for Generative Adversarial Networks'. 2018.

[6] A. Katanskiy, "Skin cancer isic," Kaggle, https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic

[7] M. Hany, "Chest CT-scan images dataset," Kaggle, https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images?rvi=1