Multimodal Medical Image Synthesis for Underrepresented Populations

Nissanth Neelakandan Abirami, Sam Devavaram Jebaraj

Abstract

The project aims to address the data scarcity in glioma MRI datasets, particularly from underrepresented regions such as Sub-Saharan Africa. This scarcity hinders the development of diagnostic models capable of generalizing across diverse populations. By leveraging both the 2D and 3D StyleGAN architectures, this project generates spatially coherent MRI slices and volumes respectively.

This work demonstrates that StyleGANs can effectively bridge the imaging data gap for glioma segmentation in underrepresented populations, providing synthetic data with strong anatomical quality.

Keywords: Brats 2024, Medical image synthesis, GAN, Generator, StyleGAN 2, 3D GAN, Adaptive Discriminator Augmentation.

I. Introduction

To address the issue of brain MRIs being biased towards a certain section of the world population, the project focuses on generating high-quality synthetic glioma slices from underrepresented sectors using a StyleGAN framework.

StyleGAN, introduced by Karras et al. (2019), is a state-of-the-art generative adversarial network (GAN) architecture that enables fine-grained control over generated images using latent style vectors. For this project, both 2D and 3D StyleGANs are implemented to synthesize multi-parametric MRI slices and volumetric data respectively.

In the 2D GAN, MRI slices are synthesized at a resolution of 256 x 256 x 1, conditioned on tumor types and slice planes. For the 3D GAN, volumes of size 128 x 128 x 128 are generated. The integration of advanced techniques like Adaptive Instance Normalization (AdaIN), noise injection, weight and dynamic style modulation enhances the diversity and fidelity of the generated images.

3D StyleGAN successfully generates individual MRI slices (Axial, Coronal, and Sagittal). This architecture however, was only able to generate the shape of the skull and the computational cost of training was substantial. 2D StyleGAN, on the other hand, was successful in generating a high-definition MRI slice with Glioma.

The outcomes of this project are evaluated using metrics such as: Adversarial loss from the generator and discriminator as well as Fréchet Inception Distance (FID). By generating realistic and diverse synthetic MRI data, this work contributes towards bridging the imaging data gap, thereby improving the generalizability for segmentation and classification.

II. Problem Statement

Medical imaging datasets suffer from significant geographic imbalances, limiting the generalizability of machine learning models for glioma detection and segmentation.

This project addresses the issue by generating high-quality synthetic glioma MRI images using 2D and 3D StyleGAN frameworks. The aim is to bridge the data gap and enhance model performance for segmentation and classification tasks.

III. Literature review

1) StyleGAN (Karras et al., 2019)

StyleGAN introduced a novel style-based generator architecture that enables fine-grained control over the image synthesis process by incorporating style modulation into the latent space.

Unlike traditional GANs that directly map latent vectors to output images, StyleGAN employs a Mapping Network to transform the latent vector into an intermediate style vector . This style vector modulates the feature maps at various layers of the generator using Adaptive Instance Normalization (AdaIN).

- Generator Components:

  Mapping Network: Converts the latent code into the style space .

  AdaIN Layers: Normalize and scale feature maps with style vectors.

  Noise Injection: Adds controlled randomness to improve texture diversity.

  Progressive Upsampling: Gradually increases image resolution.

  Relevance: StyleGAN's style modulation techniques and AdaIN layers serve as the foundation used in this project.

2) Conditional GAN (Mirza & Osindero, 2014)

Conditional GANs extended the original GAN framework by introducing conditional inputs, such as class labels, to guide the generation process. This conditioning is achieved by concatenating the input condition vector with the latent vector (for the generator) or with the input data (for the discriminator).

- Key Components:

  Conditioned Generator: Accepts both the latent vector and conditioning labels as input.

  Conditioned Discriminator: Evaluates real/fake images while ensuring alignment with conditioning labels.

  Relevance: In this project, the Conditional GAN framework was to be extended towards anatomical slice type (Axial, Coronal, or Sagittal).

3) 3D GANs (Wu et al., 2016)

3D GANs extended traditional GAN architectures to model and generate 3D volumetric data using voxel grids. The generator outputs a 3D voxel grid, while the discriminator evaluates these volumetric outputs.

- Generator Components:

  Starts with a latent code and progressively applies 3D convolutions to generate a voxel grid.

  Incorporates techniques like upsampling and noise injection for diversity and quality.

- Discriminator Components:

  Applies 3D convolutional layers to extract features and evaluate real/fake outputs.

  Relevance: The 3D GAN framework serves as the foundation for the 3D StyleGAN implemented in this project.

4) StyleGAN2-ADA (Karras et al., 2020)

StyleGAN2 improved upon the original StyleGAN architecture by introducing several enhancements, including Adaptive Discriminator Augmentation (ADA). ADA dynamically applies data augmentations (e.g., flipping, cropping) to the discriminator during training. These augmentations stabilize the training process and prevent overfitting, particularly when working with limited datasets.

- Key Features:

  ADA: Applies augmentations to the discriminator only, without impacting the generator.

  Improved Regularization: Stabilizes training and reduces mode collapse.

Higher Fidelity: Achieves improved image quality with fewer artifacts.

Relevance: ADA techniques are integrated into the training process for both the 2D and 3D GANs in this project, enabling stable convergence.

5) Multi-modal MRI Synthesis (Zhan et al., 2021)

Zhan et al. proposed a GAN-based framework for multi-modal MRI synthesis, incorporating mechanisms like feature merging to improve the quality of generated data. The model generates synthetic images across multiple modalities (e.g., T1, T2, FLAIR) while preserving anatomical structures.

- Key Contributions:

Multi-modal Conditioning: Incorporates multiple modalities into the GAN framework.

Feature Fusion: Combines features from different modalities to enhance synthesis quality.

6) GAN Dissection (Bau et al., 2019)

GAN Dissection explores the internal workings of GANs by visualizing how latent variables influence the generation of specific image features. By analyzing intermediate layers, this approach identifies semantic components and their contributions to the final output.

- Key Insights:

GANs can be dissected to understand the relationship between latent vectors and generated features.

Provides interpretable insights into feature synthesis mechanisms.

IV. Timeline

Date	Task
Nov 11	Submitted project proposal
Nov 20	Implemented a custom StyleGAN on MNIST dataset - Successful
Nov 26	Trained StyleGAN on 2D axial slices of T2W MRI scans (32x32x1 resolution) - Successful
Dec 2	Upgraded architecture for 3D MRI scans (32x32x32 resolution) - Results not convincing
Dec 4	Reverted to 2D slices
Dec 9-11	Experimented with variations: replaced Adaptive Instance Normalization with Weight Modulation and tested Feature Quantization
Dec 10	Prepared project documentation
Dec 13	Submitted final project

V. Dataset Description

The dataset used for this project is the BraTS-Africa Collection, available on The Cancer Imaging Archive (TCIA).

This dataset consists of multi-parametric MRI scans, including T1, T1C, T2W, and T2-FLAIR modalities and its annotations for tumor sub-regions, such as enhancing tumor (ET), non-enhancing core (NETC), and surrounding edema. The number of subjects used for training were just 91, each having their own corresponding modalities along with a segmentation mask.

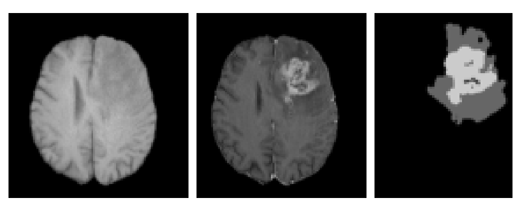

Figure 1: T1C, T2W, Segmentation mask of the same slice

T1 images provide high-resolution anatomical details with clear tissue contrast between gray matter, white matter, and cerebrospinal fluid.

T2 images emphasize water content in the brain, making edema and hyperintense soft tissues more prominent. NETC and surrounding edema are well-captured in T2-weighted scans.

VI. Dataset Preprocessing

Preprocessing plays a crucial role in ensuring that the 2D GAN and 3D GAN models effectively learn the underlying data distribution while overcoming challenges like data scarcity, noise, and inconsistency. The preprocessing pipeline is carefully designed to standardize and prepare the BraTS-Africa dataset.

1. Data Normalization:

   MRI images have varying intensity ranges due to differences in acquisition devices and settings. To standardize the data:

   Pixel intensities are normalized to the range [-1, 1] using Min-Max scaling.

   This range aligns with the output range of the Tanh activation function used in the generator.

2. Resizing and Resampling:

   2D GAN: Individual slices extracted from Axial, Coronal, and Sagittal planes are interpolated to 256 x 256 pixels. Since the expected input for the StyleGAN must be a square, this ensures uniform input dimensions for the generator and discriminator.

   3D GAN: The entire volumetric MRI data is resampled to 128 x 128 x 128 dimensions to reduce computational overhead while preserving sufficient spatial resolution.

3. Slice Extraction and Labeling:

   For the 2D GAN, slices are extracted from three orthogonal planes:

   Axial Plane: Horizontal slices ([1, 0, 0])

   Coronal Plane: Frontal slices ([0, 1, 0])

   Sagittal Plane: Side slices ([0, 0, 1])

   Each slice is labeled using one-hot encoding, which acts as the conditioning input to the generator and discriminator, enabling the generation of specific slice types.

4. Noise Reduction:

   MRI scans may contain noise and artifacts that degrade image quality. A combination of Gaussian smoothing and intensity standardization is applied to filter out noise and improve data consistency.

5. Augmentation Techniques:

   Given the limited size of the BraTS-Africa dataset, augmentation techniques are applied to artificially increase dataset variability:

   Random Flipping: Horizontal and vertical flipping.

   Random Cropping: Cropping regions of the slices to simulate diversity.

   Intensity Shifts: Randomly altering pixel intensities to simulate acquisition variability.

6. Feature Quantization:

   For the discriminator, feature quantization is applied to its intermediate feature maps. By discretizing the features, the discriminator's learning process becomes more stable and regularized.

VII. Model Architecture

The architecture of the 2D GAN and 3D GAN is based on StyleGAN, with specific modifications to address the challenges of MRI data synthesis. Each component of the generator and discriminator is designed to ensure spatial consistency, feature diversity, and alignment with the conditioning metadata.

3D GAN Architecture

Generator

The 3D generator synthesizes volumetric MRI data conditioned on clinical metadata:

1. Mapping Network:

   Similar to the 2D GAN, the mapping network converts the latent vector and metadata into a style vector.

2. Progressive Upsampling:

   The generator starts with a learned constant tensor of size 4 x 4 x 4 and progressively samples it using 3D transpose convolutions.

3. AdaIN Layers:

   AdaIN layers modulate the intermediate feature maps, ensuring the generator outputs volumetric data aligned with the conditioning input.

4. Output:

   The final output is a volumetric MRI scan of size 128 x 128 x 128, with pixel values normalized to the range [-1, 1].

Discriminator

1. The 3D discriminator evaluates the generated volumes that is similar to that of a 3D ProGAN.

2. A series of 3D convolutional layers are designed to process volumetric data through progressive downsampling in 3D, reducing the spatial resolution of the input volume (128 x 128 x 128) => (64 x 64 x 64) and so on.

3. Residual blocks would still be used to improve learning efficiency and stability.

2D GAN Architecture

Generator

The 2D generator synthesizes individual MRI slices conditioned on the slice type (Axial, Coronal, or Sagittal). The generator uses:

1. Mapping Network:

   Converts the latent vector and the conditional input (slice type) into a style vector.

2. Weight Modulation:

   Instead of directly modifying the normalization statistics like AdaIN, the weights are modulated by a learned style vector, adjusting how the feature maps are processed in the generator.

3. Noise Injection:

   Controlled Gaussian noise is added at multiple layers of the generator to introduce subtle texture variations, improving the diversity and realism of the generated images.

4. Progressive Upsampling:

   The generator starts with a learned constant tensor (e.g., 4 x 4) and progressively samples it using convolutional transpose layers to achieve the final resolution of 256 x 256 pixels.

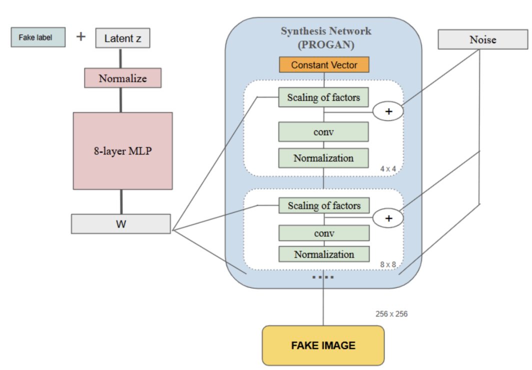

Figure 2: Generator Architecture

Discriminator

The discriminator evaluates the authenticity of generated slices and ensures alignment with the conditioning label:

1. Feature Extraction:

   Uses a series of convolutional layers to extract features from the input slice. This follows the same architecture as that of the discriminator of a normal Progressive GAN (ProGAN)

2. Adaptive Discriminator Augmentation:

   ADA dynamically augments the discriminator's input data (e.g., flipping, cropping) during training to stabilize learning. This is especially useful for limited datasets, allowing the discriminator to generalize better with fewer images.

3. Top-k Training for Generator:

   During training, gradients from low-quality fake images are zeroed out to focus the generator's learning process on high-quality outputs. This accelerates convergence and improves image quality.

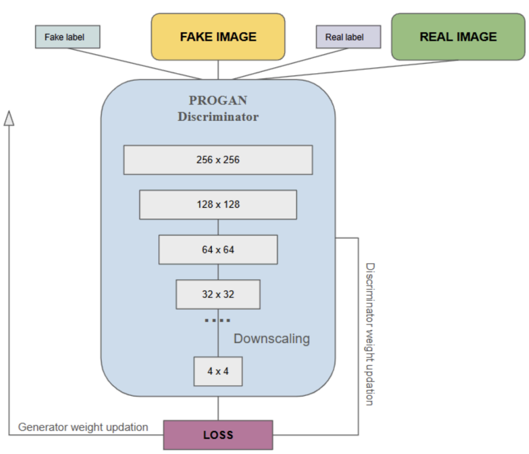

Figure 3: Discriminator Architecture

Losses:

The generator loss evaluates how well the generator fools the discriminator into classifying its outputs as real. The generator tries to minimize the discriminator's success by producing realistic data.

$$\mathcal{L}_G = -\mathbb{E}_{z\sim P_z}[\log D(G(z))]$$

The discriminator's loss measures its ability to correctly distinguish between real and fake data. It aims to maximize the probability of assigning real data a high score and fake data a low score.

$$\mathcal{L}_D = -\mathbb{E}_{x\sim P_{\text{real}}}[\log D(x)] - \mathbb{E}_{z\sim P_z}[\log(1 - D(G(z)))]$$

The Fréchet Inception Distance (FID) evaluates the quality and diversity of generated images by comparing their statistics to real images. Lower FID indicates better alignment and helps monitor long-term progress.

$$\text{FID} = \|\mu_{\text{real}} - \mu_{\text{gen}}\|^2 + \text{Tr}(\Sigma_{\text{real}} + \Sigma_{\text{gen}} - 2(\Sigma_{\text{real}}\Sigma_{\text{gen}})^{\frac{1}{2}})$$

VIII. Results

The results achieved through the implementation of both 2D and 3D StyleGAN demonstrate significant improvements. Below are the detailed results for both models:

Before using the Brats Dataset, the project turned its focus towards generating numbers using the custom conditioned StyleGAN.

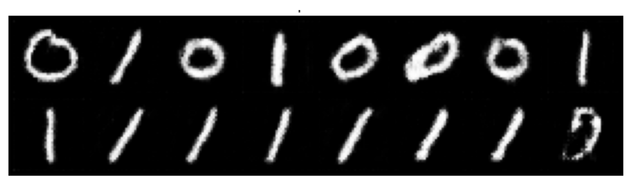

Figure 4: Generated digits from Custom StyleGAN

Since each digit was a 32x32 image, we were able to verify the working even with a short training period of 10 epochs as shown in Fig 4.

3D GAN Results

The earlier versions of the 3D GAN generated volumetric MRI slices with visible artifacts and reduced anatomical clarity, particularly in the Sagittal and Coronal views. While the general tumor structure was recognizable, the volumes lacked fine-grained details, and the boundaries of tumor regions appeared blurred as shown below in Fig 5.

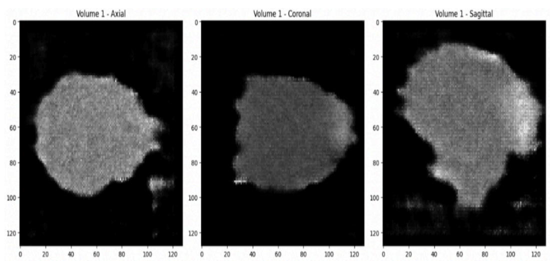

Figure 5: StyleGAN 3D (Initial Results)

After introducing noise injection, AdaIN, and additional training refinements, the final 3D GAN generated high-quality volumetric outputs with significant improvements in clarity. Axial, Coronal, and Sagittal views showed strong alignment with clear boundaries of the skull from real inputs. However, the spatial coherence with respect to the brain were better with mere recognition of tumor segments and a misaligned set of cortexes as seen below:

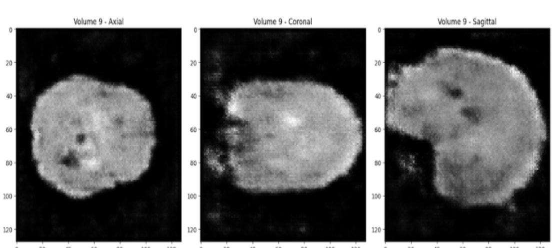

Figure 6: StyleGAN 3D (Best Results)

2D GAN Results

3D StyleGAN, our novel approach failed to capture the spatial details. This can be due to the limited resources in terms of availability of a powerful GPU. Hence, we reverted back to 2D StyleGAN focusing on generating individual Axial MRI slices.

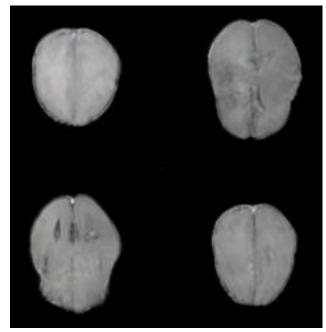

Figure 7: Axial slices generated through StyleGAN 2

Using the base model, the results were satisfactory as the model produced a 256 x 256 axial slice with better details, shown above, as compared to the previous attempt. The project also explored various alterations to the existing StyleGAN2 architecture as follows:

Adaptive Discriminator Augmentation (ADA):

ADA stabilized the training process and prevented overfitting, but also gave harper tumor boundaries compared to earlier results. However, assigning ada_target *p,* a lower value, the output images weren’t as satisfactory as the others.

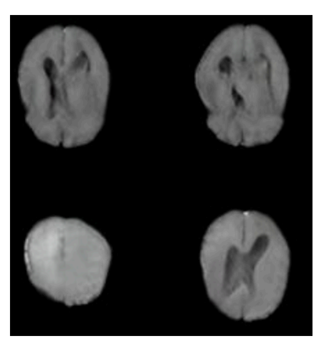

Figure 8: StyleGAN 2D with ADA

Attention Mechanism:

This showcased an enhanced focus on tumor regions, improving the delineation of critical anatomical features.

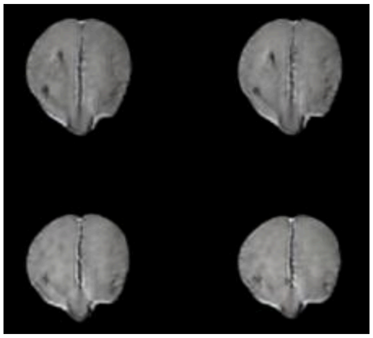

Figure 9: StyleGAN 2D with Attention

Feature Quantization:

Feature quantization introduced stability during training by regularizing the discriminator's intermediate features. This helped avoid mode collapse. But quantization loss perhaps led to huge misalignment of internal cortexes, causing a massive hole in some of the generated slices.

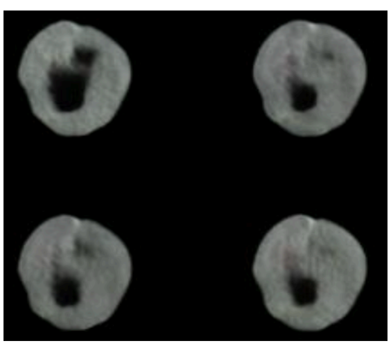

Figure 10: StyleGAN 2D with Feature Quantization

Top-K Training:

Slices generated using Top-K strategies had fewer artifacts, sharper tumor boundaries, and improved visual fidelity compared to baseline methods as seen in Fig 11.

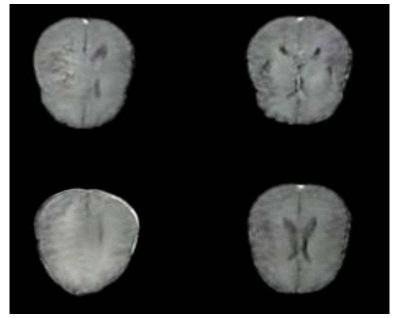

Figure 11: StyleGAN 2D with Top- K

The graph in Fig 12, shows the FID scores over epochs validates the impact of these architectural components: ADA, attention mechanisms, and Top-K training significantly reduced FID scores, indicating higher-quality images as training progressed. Baseline models without augmentations ("normal") consistently showed higher FID scores, highlighting the effectiveness of the enhancements.

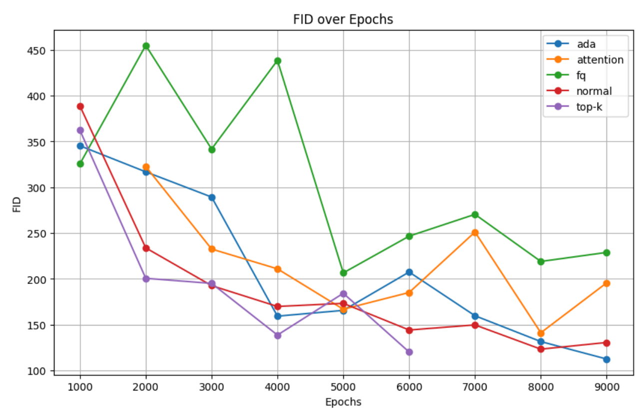

Figure 12: FID scores of different methodologies

The best one would be the Top-K generator model, as it achieved lower FID within 6000 epochs as compared to others which reached the same in 10000 epochs. However, it was terminated by 6000 due to the fact of consuming a lot of GPU power, even in a single core Nvidia A100.

Conclusion

The project successfully demonstrates the capability of 2D and 3D StyleGANs to generate high-quality synthetic MRI slices and volumetric data, addressing the challenges of data scarcity in underrepresented populations such as Sub-Saharan Africa. However, we would like to expand and visit back the 3D StyleGAN architecture with conditioning, hence this would be our extension. This would enable the same model to generate 3D volumes of the same Glioma lesions for two or multiple sectors of populations. This project really answered the question “Can a GAN used primarily for diversifying the fashion and art industry generate accurate medical MRI slices?”. Future work will focus on scaling the models to larger datasets and inclusion of multi-institutional data to further enhance applicability.

References

1. Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks*. CVPR.
2. Huang, X., & Belongie, S. (2017). *Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization*. ICCV.
3. Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets*. arXiv preprint.
4. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). *Training Generative Adversarial Networks with Limited Data*. NeurIPS.
5. Wu, J., Zhang, C., Xue, T., Freeman, W. T., & Tenenbaum, J. B. (2016). *Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling*. NeurIPS.
6. Bau, D., Zhu, J. Y., Strobelt, H., Lapedriza, A., Zhou, B., & Torralba, A. (2019). *GAN Dissection: Visualizing and Understanding Generative Adversarial Networks*. ICLR.
7. Zhan, Y., Liu, M., Xu, T., & Luo, J. (2021). *Multi-modal MRI Synthesis with GAN for Glioma Segmentation*. MICCAI.