PHAROS

THE GREEK AI FACTORY

Generative Models in MRI Data

Zoi Katsantoni, MSc Graduate
Vasileios Kochliaridis, PhD Candidate
Vlahavas Ioannis, Prof.

ARISTOTLE UNIVERSITY OF THESSALONIKI

Introduction

The rise of AI in healthcare is significantly improving diagnoses and driving the transition towards personalized medicine.

A major challenge in this progress is data scarcity, specifically the immense difficulty in acquiring high-quality volumetric imaging, such as MRIs.

- Financial barriers in data collection and **infrequent disease occurrence**.
- Strict legal and ethical constraints (protecting patient privacy).

The Solution - Synthetic Data: Artificially generated information that retains the statistical characteristics of real data, eliminating the risk of patient identification

Purpose

Main Objective

The generation of realistic and clinically useful synthetic 3D brain MRIs.

Methodology

Utilization and comparison of Generative Models, such as Generative Adversarial Networks (GANs) and Diffusion Models

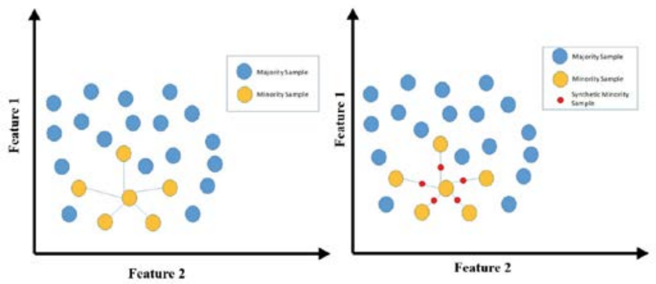
Technical Challenges

1) Handling the complex correlations between 3D data.
2) Preserving clinical validity and anatomical correctness in generated images.

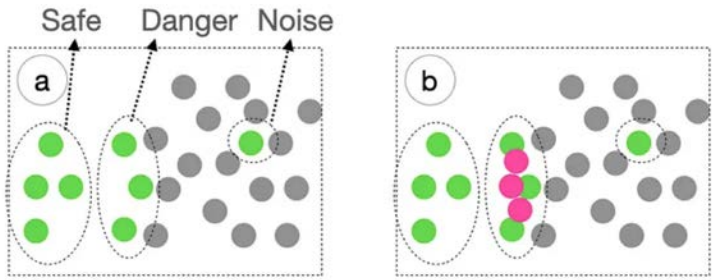
Synthetic Data Generation

Traditional Methods

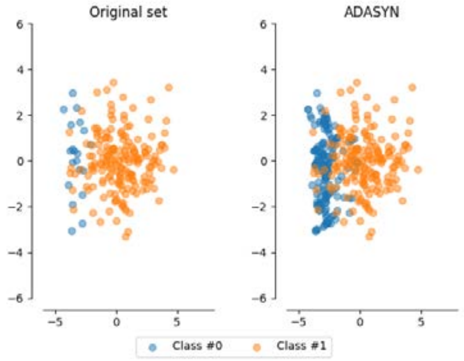
SMOTE



Borderline SMOTE



ADASYN



Problems

- Complex data structures
- The Linearity Assumption
- Loss of Spatial and Structural Context

Generative Models

All of the synthetic data generation approaches developed in this thesis are fundamentally based on these two state-of-the-art architectures, adapted specifically for MRI synthesis.

R3GAN

- Developed by **Brown & Cornell Universities** (Huang et al., 2024) in the paper "The GAN is dead; long live the GAN! A Modern Baseline GAN".
- Based on **StyleGAN2**
- Evaluated and initially trained on large-scale datasets like **FFHQ**, **ImageNet**, and **CIFAR.**

Guided Diffusion

- Developed by **OpenAI** (Dhariwal & Nichol, 2021) in the paper *"Diffusion Models Beat GANs on Image Synthesis"*.
- Built upon standard Diffusion Models, it introduces a "**Guidance**" mechanism to easily control and direct the image generation process.
- Evaluated and initially trained on **ImageNet**

Dataset

WashU Medicine

Open Access Series of Imaging Studies (OASIS)

- OASIS 1 (Cross sectional data)
  1 MRI session per person
- OASIS 2 (Longitudinal data)
  > 2 MRI sessions per person
- OASIS 3 (Longitudinal Multimodal Neuroimaging)
  MRI + PET scans
- OASIS 4 (Clinical Cohort)
  Neurospychometric, and neuroimaging assessments.

Dataset

Open Access Series of Imaging Studies (OASIS 1)

416 subjects, aged 18 to 96 years.

~100 subjects diagnosed with very mild to moderate dementia

The remaining are non-demented

- Demographics (gender, education, socioeconomic status)
- Clinical assessments (CDR & MMSE scores)

3 or 4 individual T1-weighted MRI scans per subject.

Open Access Series of Imaging Studies (OASIS 1)

Dataset

Demographic data provided

- **Gender**
- **Age**
- Dominant Hand
- Socioeconomic status
- **Education**
  1. less than high school grad
  2. high school grad
  3. some college
  4. college grad
  5. beyond college

Derived anatomic volumes

- Estimated total intracranial volume (eTIV) (mm3)
- Normalized whole brain volume (nWBV)

Clinical Data

- **Mini-Mental State Examination (MMSE)**
- **Clinical Dementia Rating**
  - 0: non-demented
  - 0.5: very mild dementia
  - 1: mild dementia
  - 2: moderate dementia

  All participants with dementia (CDR >0) were diagnosed with probable AD

Dataset

Open Access Series of Imaging Studies (OASIS 1)

Averaging → Atlas Register → Masking → Brain Tissue Segmentation

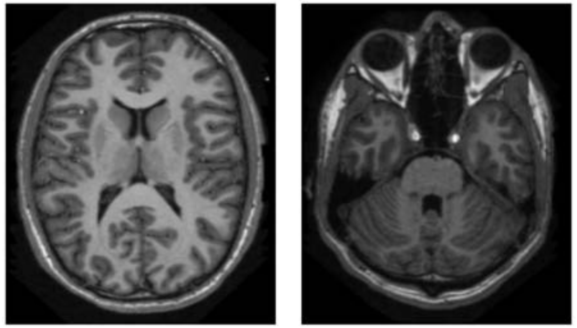
Averaging

- Smoother images,
- Motion artifacts corrected
- 1 MRI per person

Atlas Register

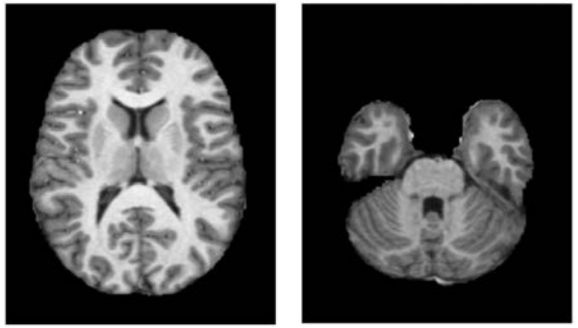
T88_111 contains the MRI:

- Registered in T88 Atlas
- Gain field-corrected



Masking

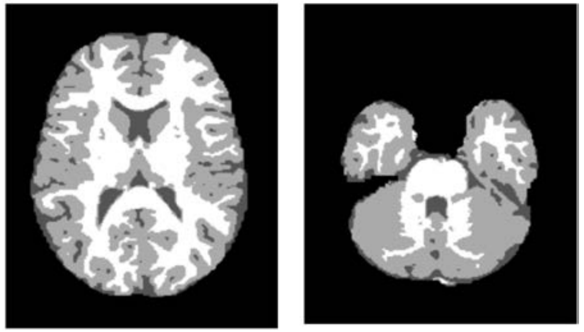
Masking contains removing non-brain tissues. Isolating the brain from other anatomical structures (e.g. eyes, skull etc)



Brain Tissue Segmentation

Categorization of brain structures into three distinct classes:

Grey matter | Cerebrospinal fluid (CSF) | White matter | Background



Methodologies

1. 3D Architecture
2. Channel Stacking
3. Positional Conditioning

3D Architecture

Model Adaptation

Transforming the R3GAN architecture into the 3D space by replacing two-dimensional convolutional layers (Conv2D) with three-dimensional ones (Conv3D).

Computational Limitations

It highlighted the exponential increase in memory and computational requirements (need for High Performance Computing) due to the 3D space.

Curse of dimensionality

Excessive memory requirements

Drastic reduction in batch size

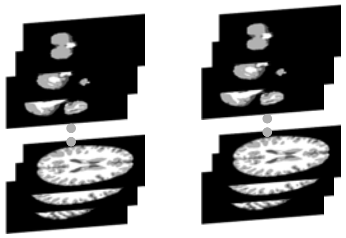
→ Training instability and mode collapse

→ Failed to converge to realistic results, producing blurry and indistinct structures

Complete training failure!

Channel Stacking

MRI size for model input:
(176,128,128)



Data shape

The Opportunity in Medical Data:

MRI images are grayscale

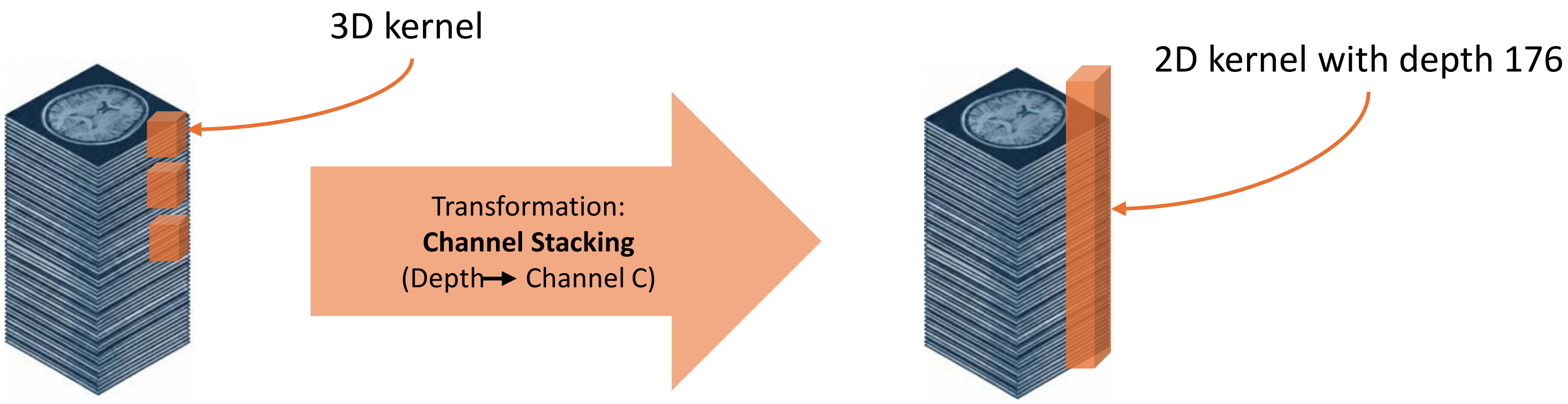
The channel dimension (used for RGB colors) remains unused

Mapping the depth axis to the channel dimension, reshaping the input from (1, D, H, W) to (D, H, W)

3D Architecture vs Channel Stacking

3D Convolutions

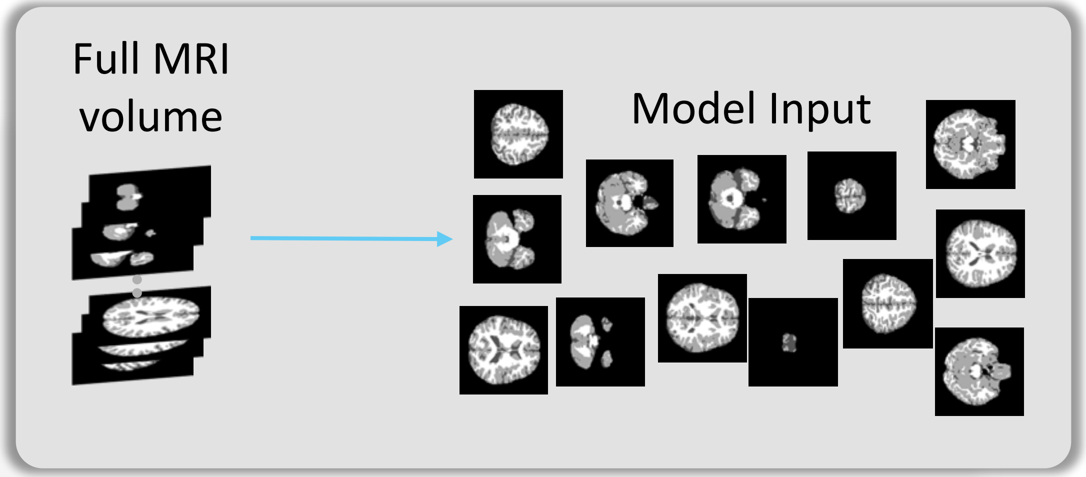
Channel Stacking



3D kernel moves in X, Y, and Z directions, processing sub-volumes
Model input size: (1,176,128,128)

2D kernel, spanning all channels, moves only in X and Y, processing the entire depth at once
Model input size: (176,128,128)

Positional Conditioning



Data shape

The Opportunity in Medical Data:

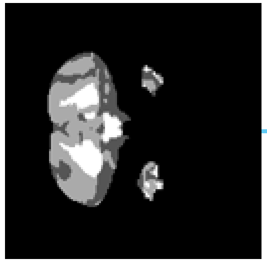
The second approach involved giving the MRI as separate pieces instead of a full volume

The input to the model was 2D image with dimensions of (1, 128, 128)

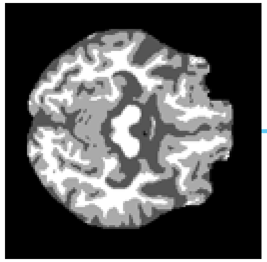
Problem: In this 2D method, the spatial information is lost!

Positional Conditioning

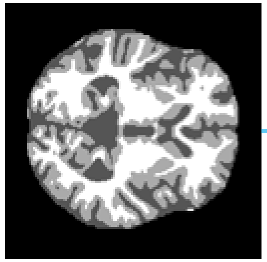
Slice ID: An integer that indicates the exact height of the slice along the depth axis.



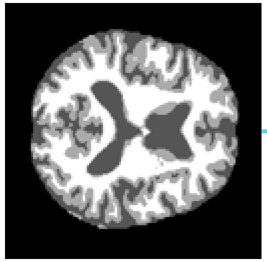
Slice ID: 30



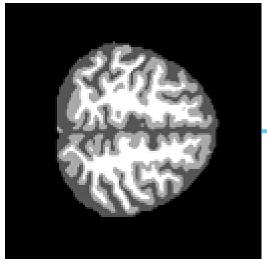
Slice ID: 62



Slice ID: 74

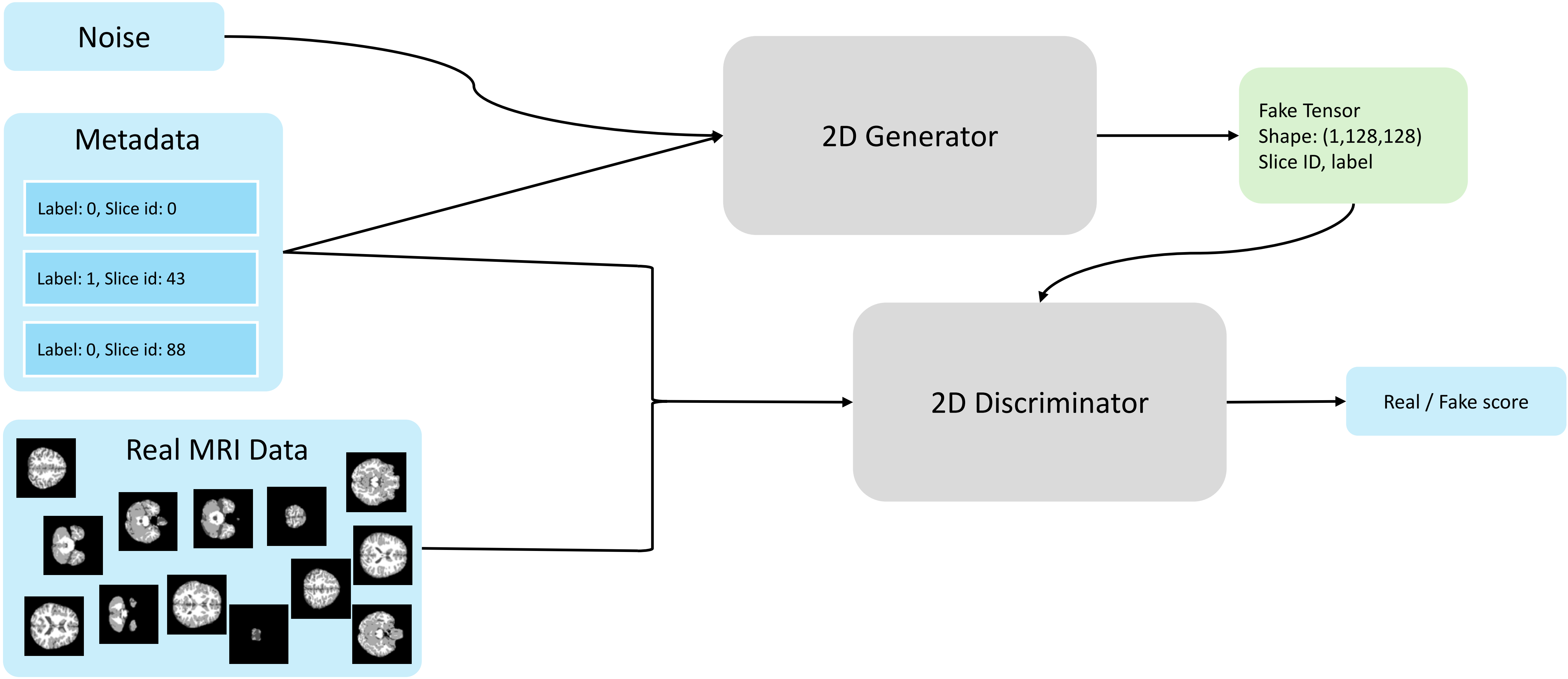


Slice ID: 88



Slice ID: 123

Positional Conditioning



Methodologies

Metadata
Gender, age, MMSE score, education

3D Architecture

- Conditional R3GAN

Channel Stacking

- Conditional R3GAN
- Conditional R3GAN + metadata
- Guided Diffusion

Positional Conditioning

- Conditional R3GAN
- Conditional R3GAN + metadata
- Guided Diffusion

Evaluation Metrics

Fréchet Inception Distance (FID)

It compares the statistical distribution of the generated images to the distribution of the real images.

Interpretation: Lower is better.

Peak Signal-to-Noise Ratio (PSNR)

It indicates how clean the image is from noise.

Interpretation: Higher is better.

Structural Similarity Index Measure (SSIM)

SSIM attempts to mimic how the human visual system perceives image quality, focusing on structural information rather than absolute pixel-level errors.

Interpretation: Closer to 1 is better.

Results – 3D Architecture

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420

Bad performance.

- very high FID
- near-zero SSIM
- In practice, the network failed to capture even the basic structural representation of the MRI volume (pure noise results)

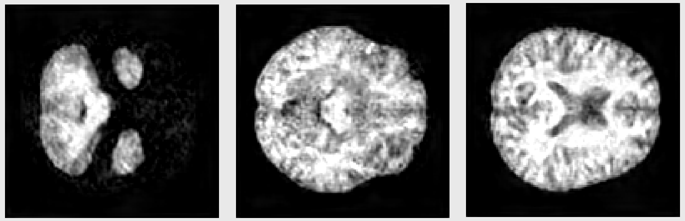
Computational Resource Constraints: The poor performance is not an architectural flaw, but a result of the **exponential complexity of 3D data**.

GPU memory limitations forced necessary compromises (e.g. smaller batch size), preventing the model from reaching optimal convergence.

Results – Channel Stacking

		FID	PSNR	SSIM
3D Model	3D R3GAN			20
Channel Stacking	Conditional R3GAN			97
	R3GAN + Metadata			22
	Guided Diffusion			72

Despite capturing the correct geometry, the output is not yet ready for clinical application. Severe **high-frequency noise** obscured fine anatomical details degrading the tissue texture.



Converting 3D data into a 2D format:

- drastically reduced the computational burden and significantly boosted performance.
- learned successfully the macroscopic structure and geometry of the brain, achieving a "pseudo-3D" perception

Results – Channel Stacking

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072

The Neutral Impact of Metadata:
increased training time without offering substantial qualitative upgrades functionally or visually.

Results – Channel Stacking

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072

- catastrophic failure of the generative process
- visual inspection reveals entirely unstructured noise

Results – Positional Conditioning

- successfully eliminated the high-frequency noise of previous methods
- produced the sharpest and most structurally accurate individual images among the GAN models.

The model excelled as a 2D image generator, but it failed as a 3D volume synthesizer.

Without neighboring slice context, the model frequently repeated the same slice across different depths!

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072
Positional Conditioning	Conditional R3GAN	211.703	24.68	0.7331
	R3GAN + Metadata	219.034	16.92	0.6285
	Guided Diffusion	16.641	24.85	0.7904

Results – Positional Conditioning

Adding metadata worsened the results.

Visually, the model produced slices identical to the previous version, containing the same problems (repeating slices across an MRI)

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072
Positional Conditioning	Conditional R3GAN	211.703	24.68	0.7331
	R3GAN + Metadata	219.034	16.92	0.6285
	Guided Diffusion	16.641	24.85	0.7904

Huge Difference!
Why?

Results – Positional Conditioning

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072
Positional Conditioning	Conditional R3GAN	211.703	24.68	0.7331
	R3GAN + Metadata	219.034	16.92	0.6285
	Guided Diffusion	**16.641**	**24.85**	**0.7904**

High quality MRI slices!

Guided Diffusion successfully learned the physiological anatomical progression!

Results – Positional Conditioning

Pure noise

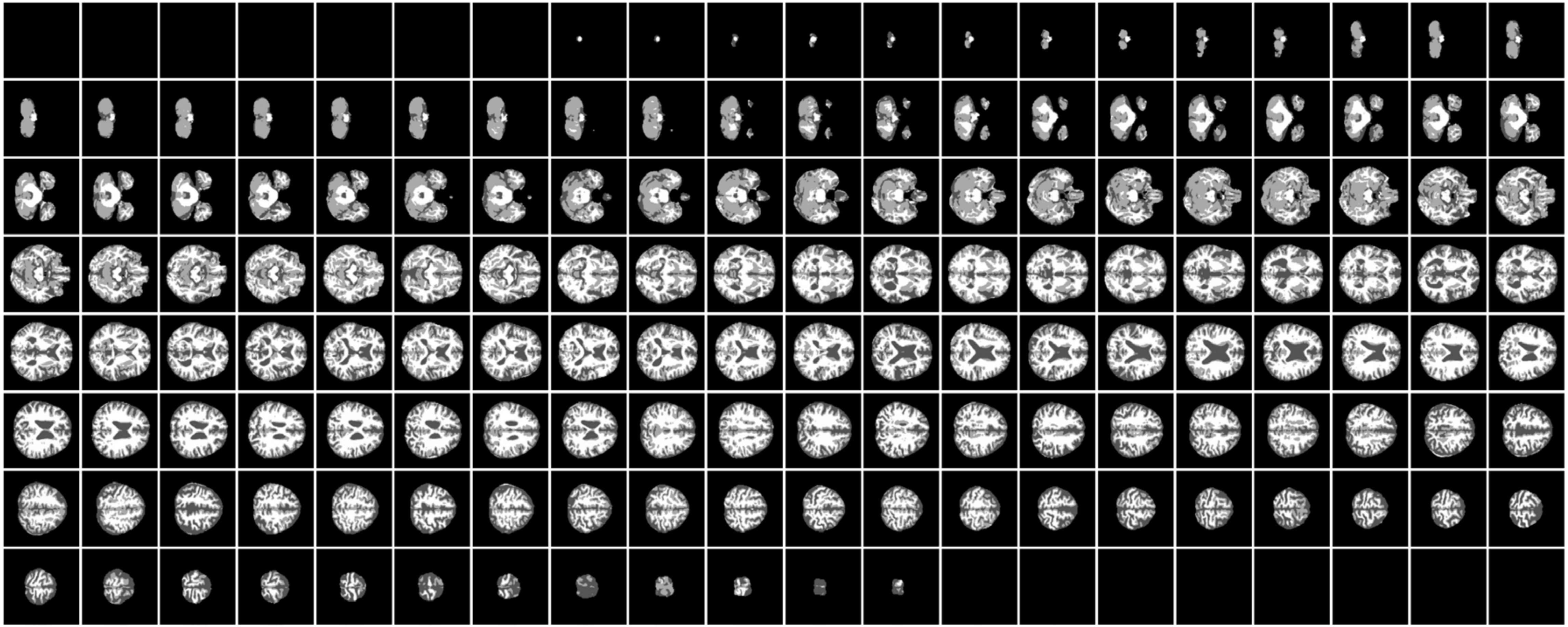
High Frequency noise

Repeating the same slice
(Lack of diversity)

All of these problems are highly penalized by FID!

		FID	PSNR	SSIM
3D Model	3D R3GAN	348.596	6.80	0.0420
Channel Stacking	Conditional R3GAN	280.160	21.63	0.5397
	R3GAN + Metadata	275.511	20.40	0.4822
	Guided Diffusion	445.835	5.84	0.0072
Positional Conditioning	Conditional R3GAN	211.703	24.68	0.7331
	R3GAN + Metadata	219.034	16.92	0.6285
	Guided Diffusion	**16.641**	**24.85**	**0.7904**

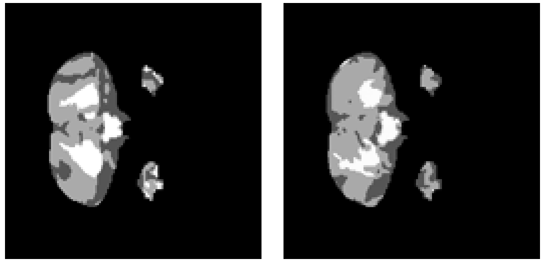
Results – Best Model



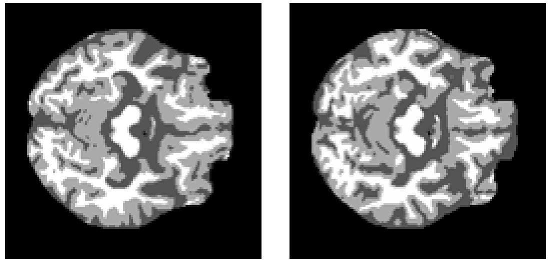
Results – Best Model

Real MRI Data | Generated MRI Data

Slice ID: 30

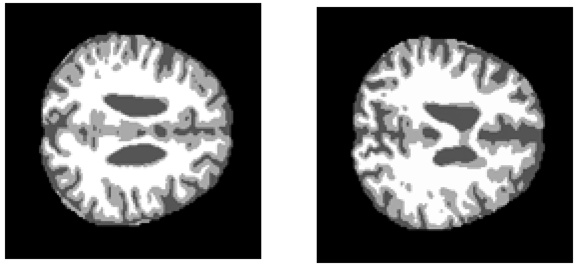


Slice ID: 62

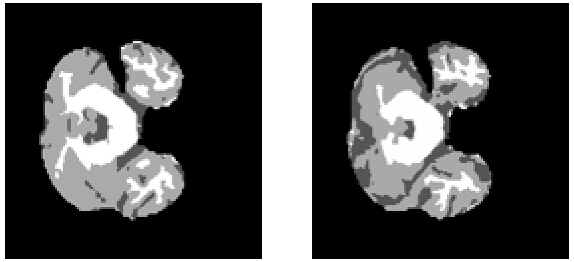


Real MRI Data | Generated MRI Data

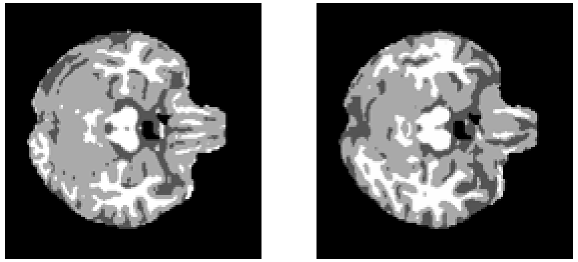
Slice ID: 105



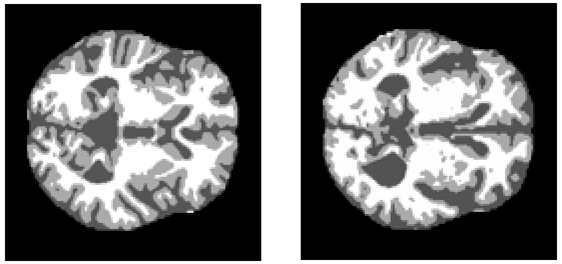
Slice ID: 43



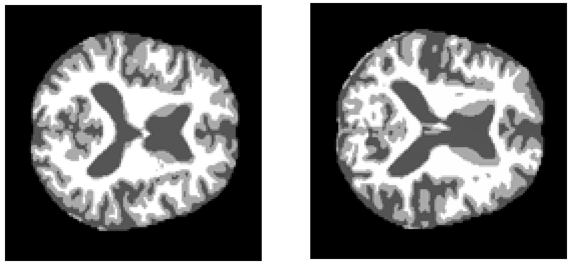
Slice ID: 55



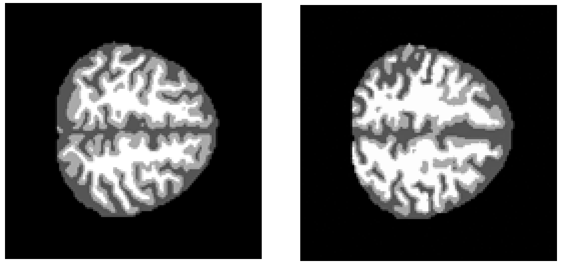
Slice ID: 74



Slice ID: 88



Slice ID: 123



Future Work

- Visual Turing Test
- Machine Learning Utility
- Privacy Preserving Metrics
- Extension to Higher Resolution
- Experiment using OASIS 2, 3, 4

Conclusions(1/2)

Native 3D Generation is Computationally Restrictive:

Attempting to train architectures like 3D R3GAN directly in 3D space leads to exponential memory demands and training failure, resulting in unstructured noise rather than usable images.

Dimensionality Reduction Aids Computation but Struggles with Quality:

Transforming 3D data into 2D formats via Channel Stacking drastically reduces the computational burden and captures basic geometry, but it introduces severe high-frequency noise that makes the images clinically unusable.

Channel Stacking is effective for GANs, but fails with Guided Diffusion.

Transforming 3D data into 2D formats via Channel Stacking produced pure-noise outputs when applied to Guided Diffusion models.

Conclusions (2/2)

Using Positional Conditioning GANs Excel in 2D but Fail in 3D Coherence:

While Positional Conditioning helped Conditional R3GAN produce sharp and structurally accurate individual 2D slices, it failed to synthesize a coherent 3D volume, frequently repeating the same slice across different depths.

Metadata Integration Yields Neutral to Negative Results:

The addition of clinical metadata increased training time without offering substantial qualitative upgrades and, in some cases, worsened the visual results.

Guided Diffusion is the best Model:

The Guided Diffusion model utilizing Positional Conditioning emerged as the best performer, successfully overcoming noise and diversity issues to accurately learn and generate the physiological anatomical progression of 3D brain MRIs.

PHAROS

THE GREEK AI FACTORY

Thank you!

Any questions?

Zoi Katsantoni

MRI Generation – Guided Diffusion

PHAROS

THE GREEK AI FACTORY

Generative Models in MRI Data

Zoi Katsantoni, MSc Graduate
Vasileios Kochliaridis, PhD Candidate
Vlahavas Ioannis, Prof.

ARISTOTLE UNIVERSITY OF THESSALONIKI