Representation of human brain MRI images through generative models

Lina Farchado
Supervisors: Julien Perez and Nicolas Boutry

Technical Report *n°1*, June 2024

Abstract

The representation of human brain MRI images through generative models has emerged as a pivotal area of research in medical imaging and computational neuroscience. This study explores the application of generative models to capture and synthesize high-fidelity representations of brain MRI images. Aligned with pre-trained foundation models, our research leverages inherent patterns in diverse brain modalities for efficient representation. We evaluate the performance of these models through a series of qualitative and quantitative metrics, demonstrating their capability to generate realistic and anatomically coherent brain images. Furthermore, we investigate the potential of these models in data augmentation and compression. Our hypothesis asserts that representing such data not only reduces storage needs but also enables subsequent applications, leveraging generative models for image generation and diagnostic analysis.

Laboratoire de Recherche et Développement de l'EPITA
14-16, rue Voltaire – FR-94276 Le Kremlin-Bicêtre CEDEX – France
Tél. +33 1 53 14 59 22 – Fax. +33 1 53 14 59 13
lina.farchado@epita.fr – https://www.lre.epita.fr/

Contents

1 Introduction

Medical imaging plays a crucial role in modern healthcare, providing invaluable insights into human anatomy and pathology. Magnetic Resonance Imaging (MRI) stands out for its ability to produce high-resolution images of soft tissues. However, the increasing volume of MRI data presents significant challenges in terms of storage, transmission, and analysis.

Recent advancements in deep learning and computer vision have opened new avenues for addressing these challenges. Generative models, in particular, have shown promising results in various image processing tasks. This study aims to explore the application of generative models for representing and augmenting human brain MRI images, with a focus on developing efficient compression techniques and creating realistic data augmentation methods through interpolation.

Traditional compression methods often struggle to maintain the delicate balance between compression ratio and image fidelity, especially for complex medical images. Our research addresses this gap by leveraging the power of generative models to achieve high compression ratios while preserving critical anatomical details.

Simultaneously, the scarcity of large, diverse datasets in medical imaging poses a significant challenge for developing robust machine learning models. Data augmentation techniques can help mitigate this issue, but traditional methods often fail to capture the complexity and variability of brain structures. Our study explores novel interpolation-based augmentation techniques that can generate realistic, anatomically consistent synthetic brain MRI images, potentially enhancing the diversity and size of training datasets.

Our objectives include:

- Developing efficient compression techniques that preserve diagnostic quality while reducing storage and transmission costs.
- Creating realistic data augmentation methods through latent space interpolation.
- Evaluating the performance of various model architectures for image compression and reconstruction tasks using the augmented data.

We hypothesize that our approach will not only reduce storage needs but also enable subsequent applications in image generation and diagnostic analysis. By combining advanced compression techniques with innovative data augmentation methods, we aim to contribute to the development of more efficient and accurate AI-driven tools for medical image analysis.

2 State of the art

Recent advancements in deep learning and computer vision have led to significant breakthroughs in image processing, compression, and synthesis. This section explores cutting-edge techniques in compression, data augmentation, and interpolation that form the foundation for our work on brain MRI image representation.

2.1 Compression

Ballé et al. (2016) introduced a neural network-based image compression method using an encoder-decoder architecture in Figure 1. The encoder processes the input image through convolutional layers and downsampling, followed by Generalized Divisive Normalization (GDN), to create a latent representation. This representation is then quantified for efficient entropy coding.

The decoder uses Inverse Generalized Divisive Normalization (IGDN) to reconstruct the image from the quantified latent space. The entire system is trained end-to-end to optimize rate-distortion performance, balancing file size and image quality. Context-adaptive binary arithmetic coding (CABAC) is employed for compressing the latent representation into the final bitstream.

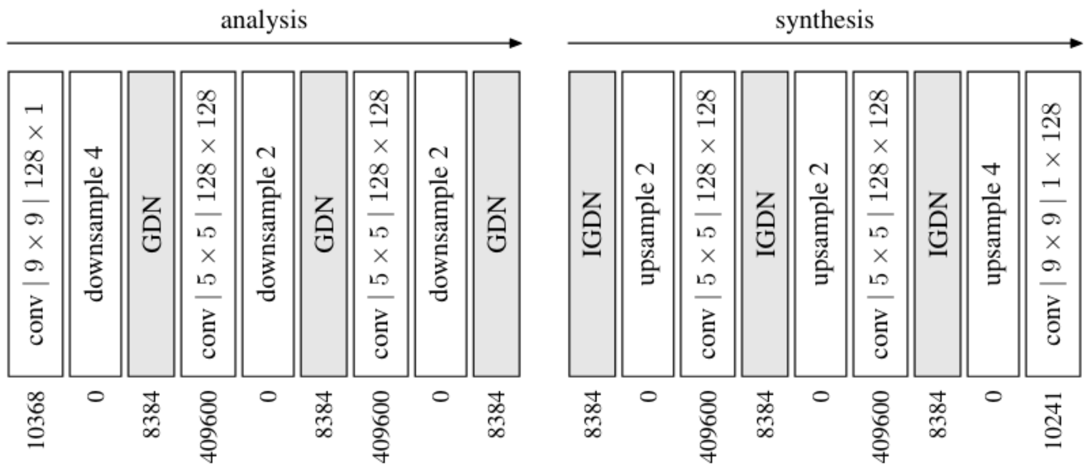


Figure 1: Schematic representation of a neural network-based image compression architecture. The system employs an encoder-decoder structure with Generalized Divisive Normalization (GDN) to achieve efficient compression while optimizing rate-distortion performance. [1]

Ballé et al. (2018) enhanced the previous model by adding a hierarchical structure with a secondary neural network acting as a hyperprior in Figure 2. This network provides contextual information to improve decoding. The advanced model, inspired by Variational Autoencoders, captures complex image dependencies and maintains end-to-end differentiability, crucial for training.

Evaluations on the Kodak dataset showed this improved architecture outperforms traditional methods like JPEG and other neural network-based algorithms, highlighting the potential of learned compression techniques in advancing image compression efficiency and quality.

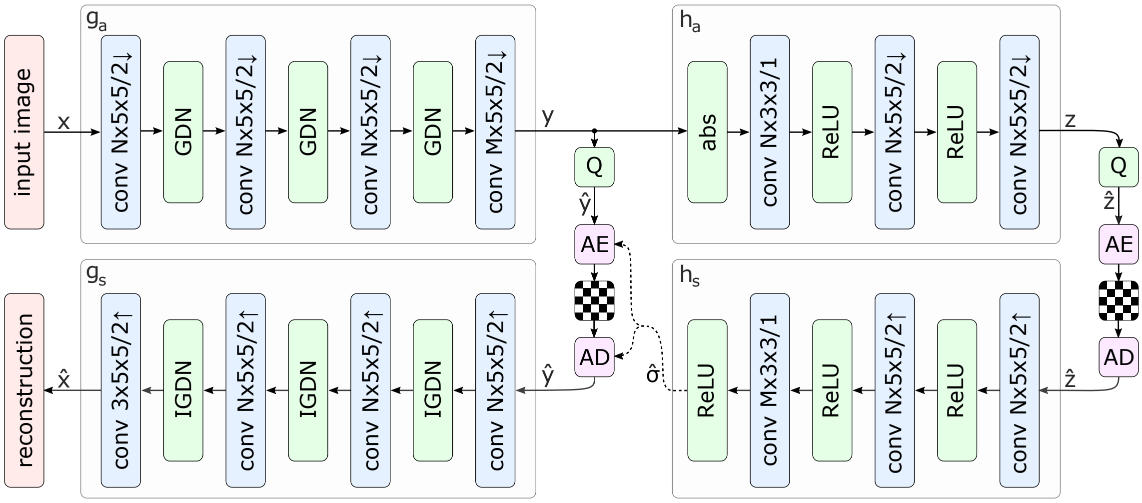


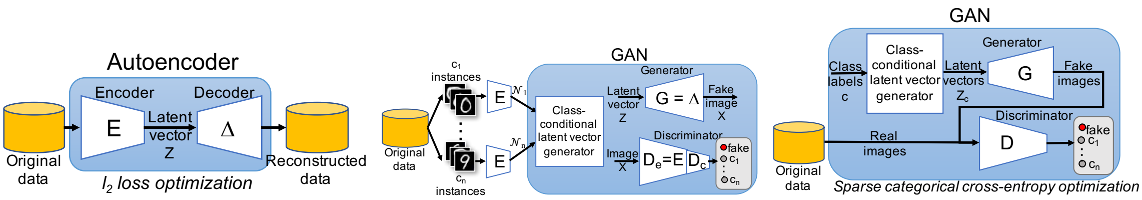
Figure 2: Enhanced image compression model featuring a hierarchical structure with a hyperprior. This advanced architecture incorporates a secondary neural network to provide contextual information, improving decoding efficiency and maintaining end-to-end differentiability.[2]

2.2 Data Augmentation

Mariani et al.(2018) propose BAGAN (Balancing Generative Adversarial Network), a methodology to restore balance in imbalanced image classification datasets by generating high-quality images for minority classes. The key novelty lies in coupling a generative adversarial network (GAN) with an autoencoder initialization strategy in Figure 3.

The autoencoder is first trained on the entire dataset to learn an encoding of the input images in the latent space. Then it is transferred to initialize the generator and discriminator of the GAN.

During adversarial training, the generator takes randomly sampled class-conditional latent vectors and generates images trying to fool the discriminator into classifying them as real examples of the respective class. A key aspect is the discriminator's output, which is a single probability distribution over all classes and the "fake" label, avoiding contradictory objectives present in previous methods. The experimental results demonstrate BAGAN's superiority over state-of-the-art GANs in generating diverse, high-quality minority class images when trained on imbalanced datasets across multiple metrics.



(a) Autoencoder training. (b) GAN initialization. (c) GAN training.

Figure 3: BAGAN (Balancing Generative Adversarial Network) methodology for addressing class imbalance in image datasets. (a) Autoencoder training on the entire dataset. (b) GAN initialization using transferred autoencoder knowledge. (c) Adversarial training of the GAN for generating minority class samples. [3]

2.3 Interpolation

GoodFellow et al.(2018) propose ACAI, a novel approach to improve interpolation in autoencoders using adversarial regularization. The key innovation lies in explicitly encouraging high-quality interpolations by introducing a critic network that attempts to predict the interpolation coefficient α used to generate interpolated points. The autoencoder is then trained to fool this critic, effectively pushing it to generate interpolated points that are indistinguishable from real data reconstructions.

As it is represented in the Figure 4, in ACAI, the interpolation process works as follows:

1. Two input data points x_1 and x_2 are encoded to obtain their latent representations $z_1 = f_\theta(x_1)$ and $z_2 = f_\theta(x_2)$.

2. An interpolated latent code is created using the coefficient α: $z_\alpha = \alpha z_1 + (1 - \alpha) z_2$, where $\alpha \in [0, 1]$.

3. The interpolated latent code z_α is then decoded to produce an interpolated data point $\hat{x}_\alpha = g_\phi(z_\alpha)$.

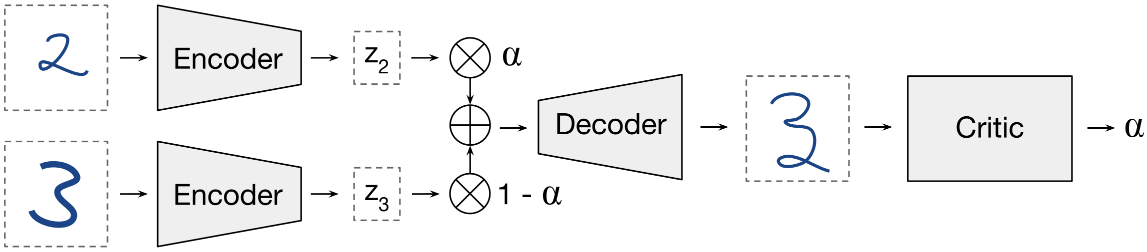


Figure 4: ACAI (Adversarially Constrained Autoencoder Interpolation) approach for improving interpolation in autoencoders. The method introduces a critic network that predicts the interpolation coefficient α, encouraging the generation of realistic interpolated points.[4]

The crucial aspect of ACAI is the introduction of a critic network that tries to predict the value of α used to generate $\hat{x}_\alpha$. The autoencoder is then trained to fool this critic, effectively pushing it to generate interpolated points that are indistinguishable from real data reconstructions.

The training process involves two main components:

1. **Reconstruction loss:** This ensures that the autoencoder can accurately reconstruct input data.

2. **Adversarial loss:** This encourages the autoencoder to produce high-quality interpolations that fool the critic.

The autoencoder is trained to generate interpolated points that the critic perceives as having $\alpha = 0$, regardless of the actual α used. This adversarial game pushes the autoencoder to create interpolations that lie on the data manifold and are indistinguishable from real data reconstructions.

The use of α in this context is crucial because it allows for a continuous spectrum of interpolations between two data points. By varying α from 0 to 1, we can generate a sequence of interpolated points that smoothly transition from one input to another. The adversarial training ensures that these interpolated points remain realistic and semantically meaningful throughout the range of α values. This approach encourages the autoencoder to learn a latent space where linear interpolations reflect smooth transitions in the data space, yielding realistic and coherent results.

3 Dataset

The dataset employed in this study consists of 3D brain MRI images from 20 patients. Each volumetric image is meticulously segmented into 144 slices, resulting in a comprehensive collection of 2880 2D images. To maintain the integrity and relevance of the dataset, we implemented a rigorous pre-processing step to exclude any fully black images, which could skew the interpolation results. Post-filtering, the training dataset comprises 2398 high-quality, informative images. For evaluation purposes, the testing dataset includes a substantial set of 3776 images.

4 Contributions

Our research makes several key contributions to the field of medical image processing and generative modeling. We present a comprehensive analysis of our proposed methods, demonstrating their effectiveness in addressing the challenges of MRI brain image compression and augmentation.

4.1 Methodology Overview

Our research employs a multi-faceted approach to address the challenges of MRI brain image compression and augmentation. We explore two main architectures: Convolutional Neural Networks (CNNs) and generative U-Nets. For each, we investigate various configurations to optimize performance. We then develop an interpolation-based data augmentation method leveraging these architectures. Our evaluation encompasses both qualitative assessments of image quality and quantitative metrics to measure compression efficiency and augmentation effectiveness.

4.2 Convolutional Neural Network (CNN)

To gain insights into the challenges faced by our generative U-Net architecture without skip connections, we developed a simplified convolutional autoencoder model. This approach allows us to pinpoint areas where the model struggles, providing valuable information for future improvements.

4.2.1 Architecture

We designed a minimal autoencoder with a single layer in both the encoder and decoder. This streamlined structure facilitates easier analysis of the model's behavior. As it is represented in the Figure 5, This architecture uses a single convolutional layer for encoding and a transposed convolutional layer for decoding. The ReLU activation in the encoder promotes non-linearity, while the sigmoid function in the decoder ensures output values between 0 and 1, suitable for image reconstruction.

We trained the model using Mean Squared Error (MSE) as the loss function, which measures the average squared difference between the input and reconstructed images. This choice allows us to quantify the reconstruction quality effectively. To investigate the model's capacity and performance, we conducted experiments with varying numbers of channels in the hidden layer: 8 channels, 16 channels and 32 channels

This progressive increase in channel count allows us to observe how the model's representational capacity affects its ability to reconstruct input images.

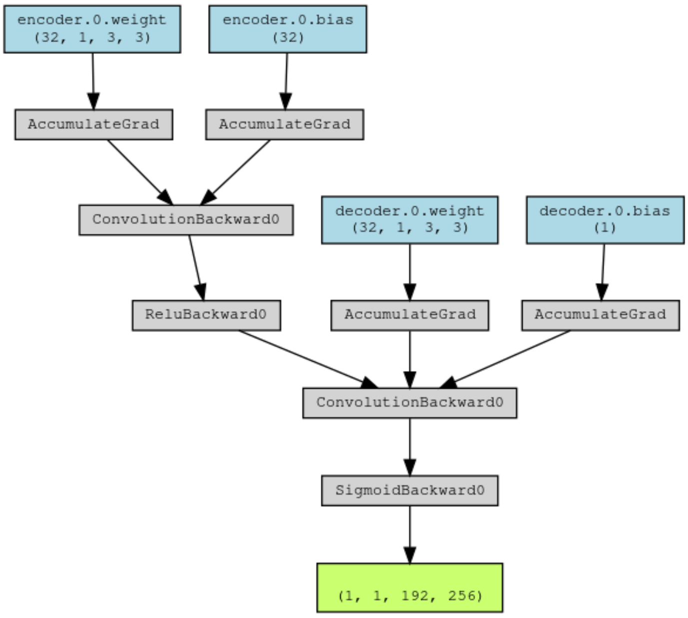


Figure 5: Schematic representation of the minimal autoencoder architecture employed in our study. The model features a single convolutional layer for encoding and a transposed convolutional layer for decoding, facilitating easier analysis of the model's behavior.

4.3 Generative Unet

4.3.1 Architechture

- **With skip connections**

  Our research trajectory with the U-Net architecture took an intriguing turn, driven by two key objectives: leveraging generative modeling techniques and enhancing our data augmentation capabilities through interpolation. These goals led us to a significant modification of the classic U-Net design mentionned by O.Ronneberger et al. (2015)

Initially, we implemented a standard U-Net with skip connections, which excelled at precise image reconstruction. The skip connections, bridging the encoder and decoder paths, enabled the network to preserve fine-grained spatial details, resulting in highly accurate reconstructions of input images.

- **Without skip connections**

  However, our ambitions extended beyond mere reconstruction. We aimed to harness the power of generative models and explore advanced data augmentation techniques. This shift in focus necessitated a rethinking of our architecture.

  The critical change came with the removal of skip connections. This decision was not made lightly, but it was essential for our purposes. Here's why:

  1. **Generative Modeling:** By eliminating skip connections, we forced the network to compress all relevant information into the bottleneck layer - the latent space representation. This condensed representation is crucial for generative tasks, as it allows the model to capture the essence of the input in a compact, manipulable form.

  2. **Latent Space Interpolation:** Our goal of data augmentation through interpolation justified the architectural change. Removing skip connections ensures that the latent representation is the sole information source for the decoder. This enables meaningful interpolation between latent vectors, allowing the generation of new synthetic data points that smoothly transition between known samples.

4.3.2 Architectural Variations

In our quest to optimize the U-Net architecture for our specific tasks, we embarked on a series of strategic changes. These modifications were designed to explore the trade-offs between model capacity, computational efficiency, and reconstruction quality. Our experiments focused on altering the channel configurations in the network's layers, providing valuable insights into the model's behavior and capabilities.

- **Expanding Channels Capacity**

  Our first modification targeted the network's capacity to capture and retain information. As it is shown in Table 4, we started with a baseline configuration of [64, 128, 512, 1024] channels in the successive layers of the encoder (and mirrored in the decoder). To enhance the model's representational power, we scaled these channels by a factor of 1.5, resulting in a new configuration of [96, 192, 384, 768].

This expansion aimed to increase the amount of information stored in the latent representation. By providing more channels, we hypothesized that the network could capture finer details and more complex features of the input images. This modification is particularly relevant in the context of our generative modeling goals, as a richer latent space could potentially lead to more nuanced and diverse outputs.

- **Streamlining the Network**

  In contrast to the expansion approach, our second modification aimed to simplify the network architecture. We reduced the number of layers from four to two, retaining only the [64, 128] channel configuration in Table 4. This pruning of the network was motivated by several factors:

  1. Efficiency: A shallower network requires less computation and memory, potentially allowing for faster training and inference.
  2. Resolution preservation: Fewer downsampling operations might help maintain spatial resolution in the output.

  However, an important modification in this approach was setting the stride to 1 (stride was set to 2 in the other configurations). This change significantly slowed down the training of the model. Consequently, this approach proved to be less interesting from a practical standpoint, as the potential benefits of the simplified architecture were outweighed by the substantial increase in training time.

4.4 Data Augmentation

4.4.1 Interpolation Method

Building upon our research into generative models for data augmentation, we developed an interpolation method for MRI brain images. This approach was heavily inspired by the work of Goodfellow et al. (2018), which provided valuable insights into the challenges and opportunities of interpolation in latent spaces.

Our interpolation strategy leverages both CNN and generative U-Net architectures to create a smooth continuum of synthetic samples between existing MRI brain images. The key to this process is the manipulation of the latent space representations learned by these models.

The interpolation process can be described as follows:

1. Encoding: Two input MRI brain images x_1 and x_2 are encoded into their respective latent space representations z_1 and z_2 using the model.

2. Interpolation: We generate new latent vectors by interpolating between z_1 and z_2 using the formula:

$$z_{interpolated} = \alpha z_1 + (1 - \alpha) z_2$$

where α is a mixing coefficient ranging from 0 to 1.

3. Decoding: The interpolated latent vector $\mathbf{z_{interpolated}}$ is then passed through the decoder to generate a new, synthetic MRI brain image.

The parameter α plays a crucial role in this process. As α varies from 0 to 1, it controls the balance between the features of the two original images in the generated sample.

- **Interpolation Between n Images**

  To interpolate between n images, we can extend the binary interpolation formula to a weighted sum of n latent vectors. Here's the equation for interpolating between n images:

  $$z_{interpolated} = \sum_{i=1}^{n} \alpha_i z_i$$

  Where:

  - $z_{interpolated}$ is the resulting interpolated latent vector
  - z_i represents the latent vector of the i-th image
  - α_i is the weight coefficient for the i-th image
  - n is the total number of images being interpolated

  With the constraint:

  $$\alpha_i \geq 0 \quad \forall i, \sum_{i=1}^{n} \alpha_i = 1$$

  This constraint ensures that the weights sum to 1 and are non-negative, maintaining the interpolation property.

In this formulation:

- When a particular $\alpha_i = 1$ and all others are 0, the output will be equivalent to the $i-th$ input image.
- By varying the α_i values, we can create a mix of features from all n images in different proportions.
- The space of possible interpolations becomes an $(n-1)-dimensional$ simplex, allowing for more complex and diverse synthetic samples compared to the linear interpolation between just two images.

4.4.2 Advantages of this Method

Our approach differs from naive pixel-space interpolation, which often results in unrealistic or blurry outputs. By operating in the latent space, we leverage the semantic understanding captured by our generative models to produce more meaningful and realistic interpolations.

It's important to note that this interpolation in latent space is not simply a linear blending of pixel values. Instead, it aims to traverse the underlying manifold of brain MRI data, creating realistic intermediate states that maintain the structural integrity and characteristics of genuine brain images.

We experimented with various values of α to generate a diverse range of synthetic samples. This allowed us to significantly augment our dataset with new, plausible MRI brain images that exhibit a smooth transition between existing samples.

This interpolation-based data augmentation technique has shown promising results in expanding our MRI brain image dataset, potentially improving the robustness and generalization capabilities of models trained on this augmented data.

4.5 Benchmark

To evaluate our approach and understand the impact of architectural choices on our goals of data augmentation and compression using generative models, we conducted a series of experiments with different model configurations. Our benchmark analysis includes both qualitative and quantitative assessments of the CNN and U-Net architectures.

4.5.1 Qualitative Results

- **Convolutional Neural Network Analysis**

  To gain a clear understanding of the challenges faced by more complex models, we leveraged a basic convolutional autoencoder as a preliminary investigation.

This strategic approach yielded valuable insights, successfully pinpointing areas for future model improvement.

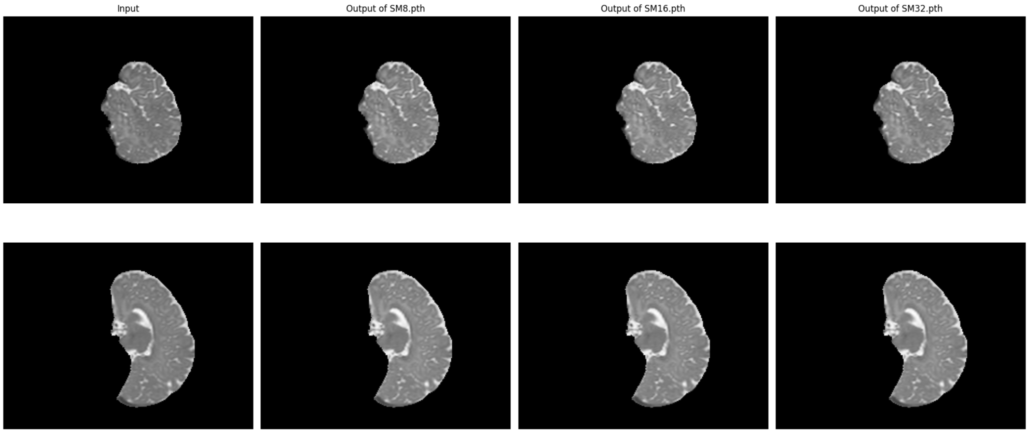


Figure 6: Visual comparison of reconstruction quality for three CNN configurations (8, 16, and 32 channels). The images demonstrate subtle improvements in detail preservation as the number of channels increases.

In Figure 6, the visual analysis of reconstruction quality across three CNN configurations (8, 16, and 32 channels) revealed that increasing the number of channels led to subtle but noticeable improvements in the preservation of detailed features.

The error visualizations Figure 7 in both 2D and 3D provide additional insights:

- Lower channel counts (8) resulted in larger, more distributed errors across the image.
- As channel count increased (16, 32), errors became more localized and decreased in magnitude.
- Border regions consistently showed higher error rates across all configurations, indicating a persistent challenge in accurately reconstructing these areas.

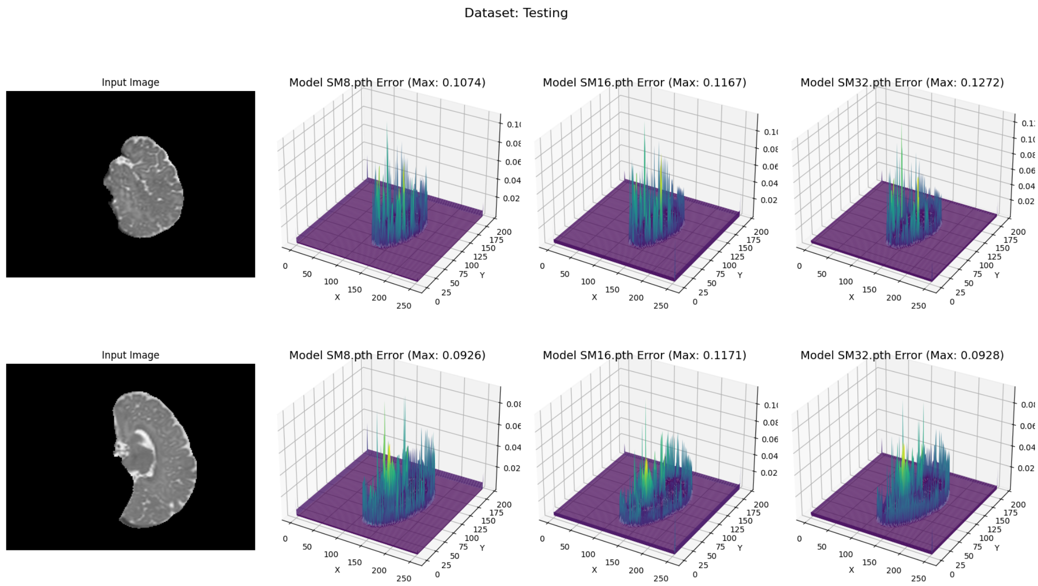


Figure 7: Topographic error visualization for CNN reconstructions with 8, 16, and 32 channels. The plots illustrate the spatial distribution of reconstruction errors across different model configurations.

These findings suggest that while increasing model capacity through additional channels generally improves reconstruction quality, certain aspects of the image (particularly borders) remain challenging. This insight guided our approach to more complex architectures.

- **U-Net Architecture Analysis**

  Building on the insights from our CNN experiments, we explored various U-Net configurations to address our specific goals of data augmentation and generative modeling.

  Initially, we implemented a standard U-Net with skip connections, which excelled at precise image reconstruction. To better suit our objectives, we removed the skip connections from the U-Net architecture. This crucial modification forced the network to compress all relevant information into the bottleneck layer, creating a latent space representation ideal for generative tasks and interpolation.

  In the Figure 8, there's the modified U-Net, devoid of skip connections, exhibited results that were not as good as those achieved with the original skip connections. However, these results are still promising. Despite a slight decrease in pixel-perfect accuracy, the model showed an improved ability to capture higher-level features and relationships within the input images, opening up new possibilities aligned with our research goals.

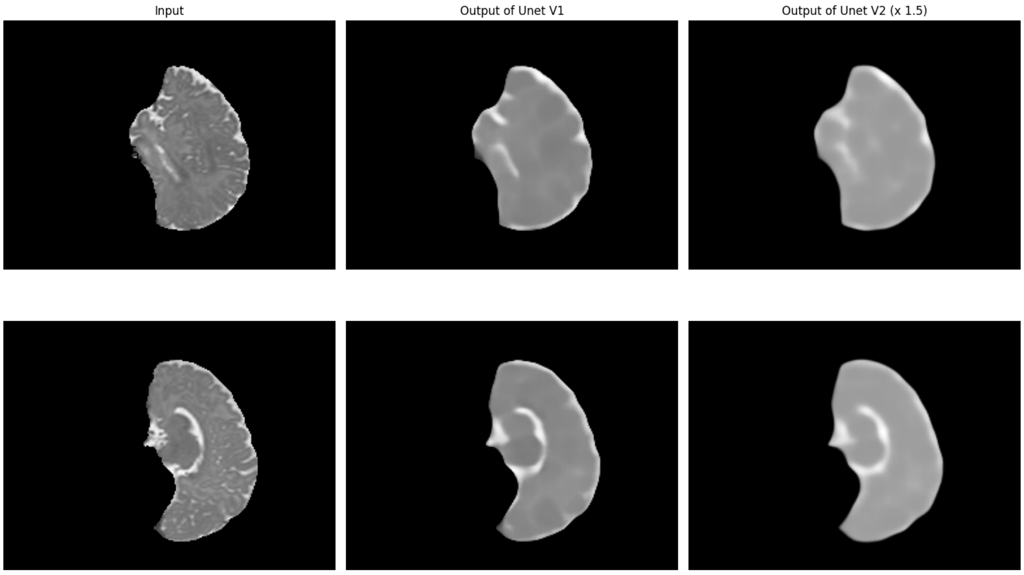


Figure 8: Comparison of brain MRI reconstructions using different U-Net configurations. From left to right: original image, reconstructions using baseline U-Net ([64, 128, 256, 512] channels), and reconstruction using expanded U-Net (channels multiplied by 1.5).

Regarding the two UNet optimizations, we chose not to train the 'Streamlined Network' due to its slow training speed. Instead, our focus was on evaluating the results of the version with 'Expanded Channels Capacity'. The image presents a comparative analysis of brain section reconstructions using different versions of the UNet model. The first column features the original brain section images, showcasing intricate textures and fine details of the brain's structure. In the second column, the images are reconstructed using UNet V1 with channel configurations of [64, 128, 256, 512].

These reconstructed images, while maintaining the overall structure, appear smoother and exhibit a slight loss of fine details, indicating a trade-off between noise reduction and detail preservation. The third column displays reconstructions using an enhanced UNet V2 model, where the number of channels is increased by a factor of 1.5.

This version produces even smoother images with further reduction in fine detail, emphasizing broader structural elements at the cost of finer textures. This comparative visualization underscores the impact of channel configurations on image reconstruction quality, demonstrating how increasing the channels in UNet models can smooth out images but may also lead to a loss of detailed information.

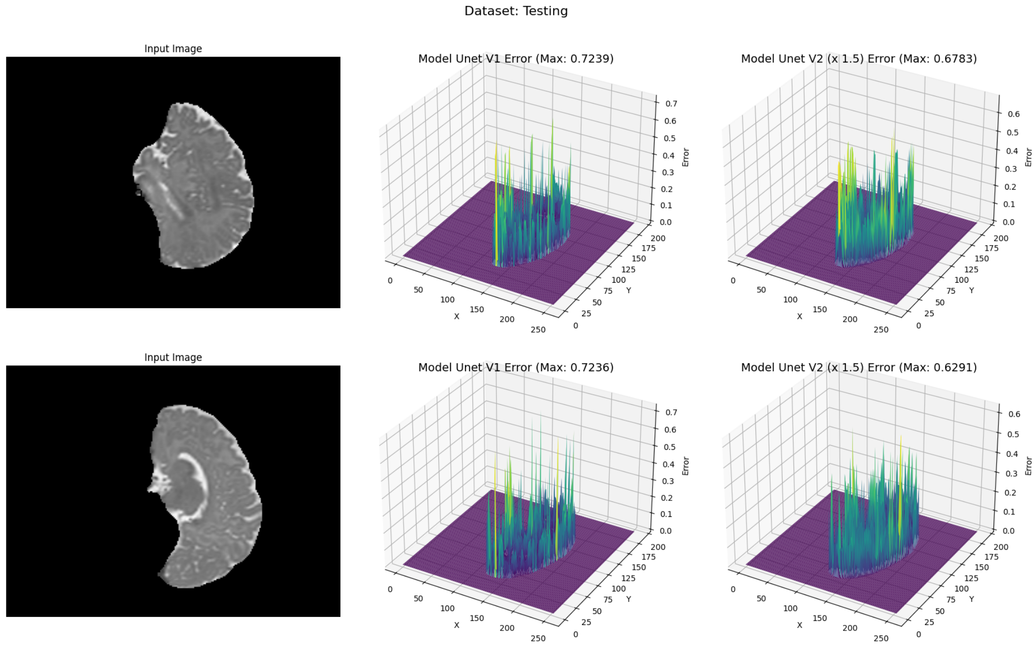


Figure 9: Error distribution comparison between U-Net V1 and U-Net V2 (Channels x1.5). The 3D error plots highlight specific areas where each model performs well or encounters difficulties in image reconstruction.

The comparison between the error distributions of UNet V1 and UNet V2 in the Figure 9 demonstrates the impact of model complexity on reconstruction accuracy. While UNet V1 exhibits higher error peaks, reflecting its limitations in detail preservation, UNet V2 with increased channels shows improved performance with lower error values. This indicates that augmenting the model's capacity by increasing the number of channels enhances its ability to produce more accurate reconstructions, effectively reducing the overall reconstruction error. The 3D error plots provide a clear visualization of these differences, highlighting specific areas where each model performs well or encounters difficulties. This comparative analysis underscores the importance of model configuration in achieving high-quality image reconstructions in medical imaging applications.

- **Interpolation Analysis**

  Our interpolation experiments produced visually convincing results in generating synthetic brain images. Figure 10 presents a benchmark comparison of our method against other approaches. The image shows five brain MRI scans. The leftmost image, labeled "Image 1", and the second image, labeled "Image 2", represent the two original brain scans used in the interpolation process. and the others are the generated images.

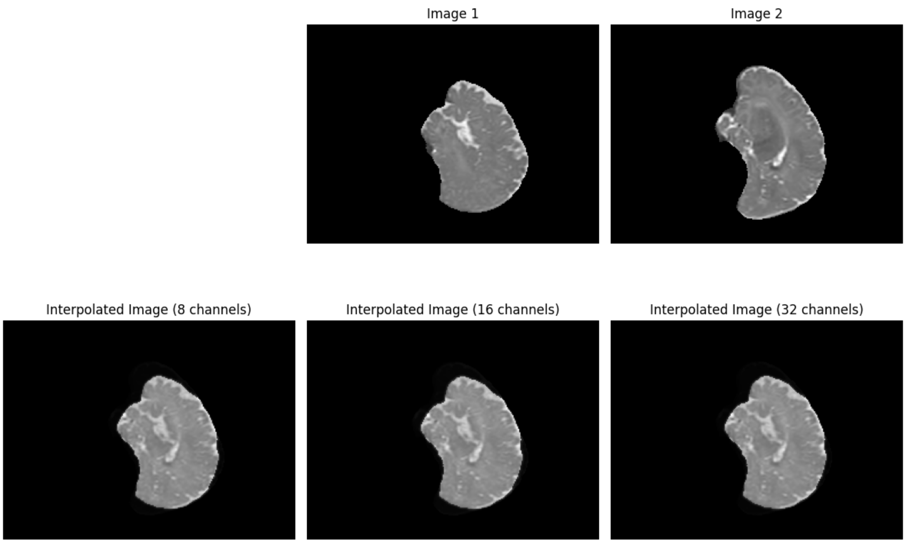


Figure 10: Comparative analysis of brain MRI interpolation results. From left to right: two source images (Image 1 and Image 2), followed by three interpolated images generated by our method using CNN configurations with 8, 16, and 32 channels respectively.

  Our method demonstrates the ability to generate a new, plausible brain image that effectively combines features from both source images. The synthetic images maintain the overall structure and characteristics of brain MRI scans, with key anatomical features remaining discernible. This visual coherence suggests that our interpolation approach successfully captures and blends the essential features of brain anatomy. Such capabilities have potential applications in medical imaging, including data augmentation for machine learning models and the creation of diverse synthetic datasets for research purposes.

4.5.2 Quantitative Results

To evaluate our work, we have chosen the following metrics:

- Peak Signal-to-Noise Ratio (PSNR) measures signal fidelity by comparing signal strength to noise interference. Higher values indicate better quality.
- Mean Squared Error (MSE) Average quantifies average squared differences between estimated and actual values. Lower MSE signifies better quality.
- Structural Similarity Index (SSIM): Measures image similarity in luminance, contrast, and structure. Higher SSIM values indicate better quality.
- Novelty Score evaluates the novelty or uniqueness of generated images compared to a reference dataset.
- Learned Perceptual Image Patch Similarity (LPIPS) measures perceptual similarity between images based on learned features.

4.5.2.1 Compression

In the context of medical image compression, particularly for MRI brain images, it is crucial to carefully consider our evaluation criteria to ensure both efficient compression and high-quality reconstruction. Efficient compression allows for the reduction of storage and transmission costs, while high-quality reconstruction ensures that the diagnostic value of the images is preserved. This section focuses on metrics most relevant to our generative model-based compression method and evaluates various CNN and UNet configurations.

We conducted a thorough benchmarking of different CNN configurations employing 8, 16, and 32 channels

	CNN 8 channels	CNN 16 channels	CNN 32 channels
PSNR	9.180-48.161	8.556-47.426	8.516-48.570

Table 1: (Reconstruction performance comparison for various CNN configurations): "Comparison of Peak Signal-to-Noise Ratio (PSNR) ranges for CNN models with varying channel configurations. Higher PSNR values indicate better reconstruction quality."

The evaluation of these CNN configurations provides valuable insights into their performance in compressing and reconstructing MRI brain images. The PSNR values exhibit significant variability across the configurations, indicating varying levels of reconstruction quality. Despite similarities in the upper ranges, increasing the number of channels does not uniformly improve the maximum PSNR, highlighting the nuanced impact of channel configuration on performance.

	Unet V1	Unet V2 (x1.5)
PSNR	0.572–25.668	6.863–27.583
MSE	0.047	0.027
SSIM	0.038	0.044

Table 2: (Reconstruction performance comparison for various generative UNet configurations without skip connections): "Performance metrics for different U-Net configurations without skip connections."

For generative UNet architectures, we further examined their reconstruction performance using metrics including PSNR, Mean Squared Error (MSE), and Structural Similarity Index (SSIM). These metrics are critical for assessing both the accuracy and perceptual quality of the reconstructed images.

These results highlight the significant impact of loss function and architectural modifications on the performance of UNet models in image reconstruction tasks. The UNet configuration with a 1.5 scaling factor demonstrates superior PSNR and SSIM scores, underscoring its effectiveness in preserving image fidelity compared to other configurations.

These findings highlight the importance of careful model selection and parameter tuning in generative image reconstruction tasks, where balancing between different quality metrics is crucial depending on the specific application requirements.

4.5.2.2 Data Augmentation

In the context of medical imaging, particularly MRI brain images, data augmentation through interpolation is a valuable technique. It is crucial to carefully consider our evaluation criteria to ensure that the generated images are both efficient and effective. This approach can bring significant benefits, such as enhancing the robustness of machine learning models and improving their generalization capabilities. However, it is equally important to be aware of the potential limitations, such as the introduction of noise or unrealistic variations that might impact system performance. In our case, we focus on the application of image compression and how these interpolated images influence the outcomes.

Our analysis of novelty and perceptual similarity scores reveals a spectrum of generated images. Some closely resemble the original dataset, offering subtle varia-

	Best CNN	Best UNet
Novelty Score	4.46–11.0	4.17–11.72
LPIPS	0.16–0.47	0.21–0.47
PSNR	0.90	-0.39
SSIM	0.03	0.04

Table 3: (Performance results for interpolation using the best CNN, best UNet BCE, and UNet MSE): "Evaluation metrics for image interpolation using the best-performing models. Novelty Score, LPIPS, PSNR, and SSIM are compared to assess the quality and diversity of generated images."

tions, while others introduce significant diversity. This range is beneficial for data augmentation, but extremely novel images risk incorporating unrealistic elements. The average scores suggest a good balance between similarity and novelty in the augmented data.

Comparing the U-Net and CNN models, we observe that the U-Net generates a wider range of images in terms of novelty and perceptual difference. While this increased diversity could enhance data augmentation, it also raises concerns about the realism of the most novel images. The U-Net's higher minimum perceptual difference score indicates that even its most similar images are more distinct from the originals compared to the CNN's output. These findings suggest that while the U-Net shows promise in generating diverse images, further refinement may be necessary to ensure the production of realistic and structurally accurate brain MRI images suitable for medical applications.

5 Conclusion

In conclusion, our study demonstrates the potential of generative models in representing and augmenting human brain MRI images. We have developed and evaluated various architectures, including CNNs and U-Nets, for image compression and interpolation tasks. While our results show that these models can effectively capture the complex features of brain MRI images, enabling efficient compression, there is still room for optimization. We are on the right track, but further work on parameter tuning and model upgrades is needed to enhance efficiency. Our interpolation method has demonstrated the ability to generate anatomically consistent synthetic brain images, which could significantly enhance datasets for machine learning applications in medical imaging. However, it is crucial to control the data augmentation process to avoid physiological inconsistencies. The balance between novelty and realism in these generated images requires careful consideration and further refinement.

5.1 Possible improvements and future directions

- Exploring more advanced architectures, such hybrid models or improve the features on the existing ones, to further improve image quality and compression ratios.
- Adding mechanisms to control data augmentation consistency, ensuring the generated brain images are physiologically accurate.
- Investigating the impact of our data augmentation techniques on downstream tasks, particularly in medical image segmentation and classification.
- Expanding the dataset to include a wider range of pathologies and patient demographics to improve model generalization.

6 Bibliography

[1] Ballé, Johannes, Valero Laparra, and Eero P. Simoncelli. "End-to-end optimized image compression." arXiv preprint arXiv:1611.01704 (2016).

[2] Ballé, Johannes, et al. "Variational image compression with a scale hyperprior." arXiv preprint arXiv:1802.01436 (2018).

[3] Mariani, Giovanni, Scheidegger, Florian, Istrate, Roxana, et al. "Bagan: Data augmentation with balancing GAN." arXiv preprint arXiv:1803.09655 (2018).

[4] GoodFellow, Berthelot, David, et al. "Understanding and improving interpolation in autoencoders via an adversarial regularizer." arXiv preprint arXiv:1807.07543 (2018).

[5] Ronneberger, O., et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv preprint arXiv:1505.04597 (2015).

[6] Liu, Xuan, Zhang, Lu, Guo, Zihao, Han, Tailin, Ju, Mingchi. "Medical image compression based on variational autoencoder." Mathematical Problems in Engineering, 2022, 7088137, 12 pages, 2022. https://doi.org/10.1155/2022/7088137

7 Appendix

U-Net Configuration		
A	B	C
5 weight layers	5 weight layers	3 weight layers
input		
conv3-64 conv3-64	conv3-96 conv3-96	conv3-64 conv3-64
maxpool		
conv3-128 conv3-128	conv3-192 conv3-192	conv3-128 conv3-128
maxpool		
conv3-256 conv3-256	conv3-384 conv3-384	
maxpool		
conv3-512 conv3-512	conv3-768 conv3-768	
maxpool		
conv3-1024 conv3-1024	conv3-1537 conv3-1537	
up-conv2		
conv3-1024 conv3-512	conv3-1537 conv3-768	
up-conv2		
conv3-512 conv3-256	conv3-768 conv3-384	
up-conv2		
conv3-256 conv3-128	conv3-384 conv3-192	conv3-256 conv3-128
up-conv2		
conv3-128 conv3-64 conv1-64	conv3-192 conv3-96 conv1-96	conv3-128 conv3-64 conv1-64
output		

Table 4: Detailed architectural specifications for three U-Net configurations (A, B, and C). The table outlines the number of weight layers, convolutional filters, and channel dimensions for each configuration, highlighting the structural differences between the models.