Representation of human brain MRI images through generative models

Lina Farchado
Supervisors: Julien Perez and Nicolas Boutry

Technical Report *n°2*, January 2025

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 interpolation-based approaches for data augmentation in brain MRI analysis, focusing on the generation of anatomically coherent synthetic images. We propose and evaluate three distinct interpolation configurations. Our methods demonstrate significant improvements in segmentation tasks. The proposed approaches maintain anatomical plausibility while introducing beneficial variations, as evidenced by low Mean Squared Error scores and improved Dice coefficients across tissue classes. These results suggest that our interpolation strategies offer viable solutions for augmenting brain MRI datasets while preserving critical anatomical features, potentially enhancing the robustness of downstream medical imaging applications.

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

The analysis and processing of brain MRI images represent a critical challenge in modern medical imaging, particularly in the context of limited dataset availability. Data augmentation has emerged as a crucial strategy to address this challenge, enabling the expansion of training datasets while maintaining anatomical validity. Traditional augmentation techniques, such as geometric transformations or intensity adjustments, while useful, may not capture the complex anatomical relationships present in brain MRI data. This limitation has motivated the development of more sophisticated approaches that can generate synthetic images while preserving biological plausibility. Our research focuses on developing novel interpolation-based methods for brain MRI data augmentation. We propose three distinct approaches to maintain anatomical coherence while introducing beneficial variations that can enhance model robustness. By operating in the latent space and incorporating spatial awareness, our approaches aim to generate synthetic images that preserve the complex structural relationships present in brain anatomy. The primary contributions of this study are:

- Development of three interpolation-based augmentation strategies specifically designed for brain MRI data
- Introduction of a distance-based weighting scheme for multi-image interpolation that preserves anatomical continuity
- Comprehensive evaluation of augmentation effectiveness through both qualitative and quantitative metrics
- Validation of the proposed methods through detailed segmentation experiments across multiple tissue classes

Our results demonstrate significant improvements in segmentation performance, particularly for gray and white matter tissues, suggesting that our approaches offer viable solutions for enhancing brain MRI analysis pipelines. Having established the importance of data augmentation in brain MRI analysis, we now review relevant prior work in the field.

2 State of the art

Recent advances in deep learning and computer vision have led to significant developments in medical image analysis. Two key areas particularly relevant to our

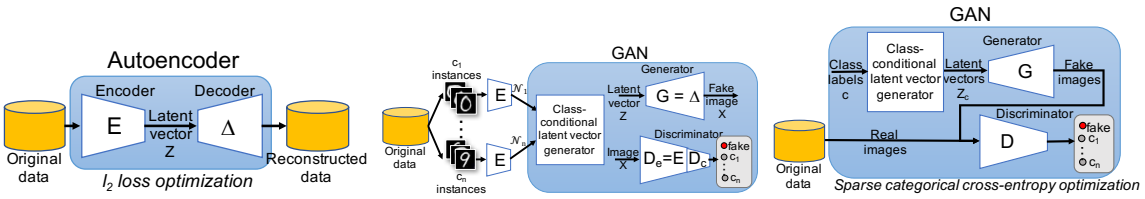
work are data augmentation techniques and interpolation methods. Data augmentation has proven crucial for improving model robustness and addressing data scarcity in medical imaging, while novel interpolation approaches have enabled the generation of anatomically plausible synthetic images. This section reviews significant contributions in both areas, providing context for our proposed methodology.

2.1 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 1.

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 1: 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. [1]

2.2 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.

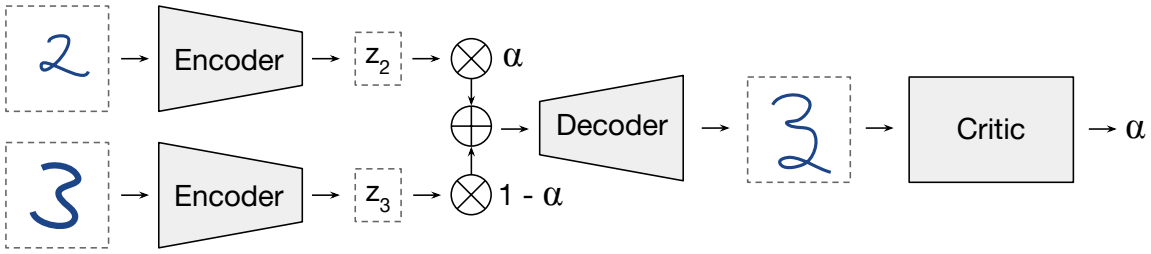


Figure 2: ACAI (Adversarially Constrained Autoencoder Interpolation) approach for improving interpolation in autoencoders. The method introduces a critic network that predicts the interpolation coefficient α, encouraging the generation of realistic interpolated points.[2]

As it is represented in the Figure 2, 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 $\hat{x}_\alpha = g_\phi(z_\alpha)$.

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 Methodology

These advances in data augmentation and interpolation techniques provide the theoretical foundation for our work. Building upon these concepts, we now present our methodology for generating anatomically coherent synthetic brain MRI images through novel interpolation-based approaches.

3.1 Dataset and Preprocessing

The dataset employed in this study consists of T2-weighted 3D brain MRI images collected from 20 patients. For our primary interpolation experiments, we used 10 patients for the training phase and 10 different patients for testing. Each volumetric MRI image was meticulously segmented into 144 axial slices, producing a comprehensive collection of 2D images. To maintain the integrity of the analysis and prevent potential interpolation artifacts, we implemented a pre-processing step to exclude completely black images (images containing only zero-value pixels) that contained no meaningful information. Additionally, to evaluate the segmentation performance of our approach, we conducted experiments on a subset of the data consisting of 10 patients from our dataset. This subset was divided into 8 patients for training/validation and 2 patients for testing. After the pre-processing step to remove black images (728 from the training and validation set and 196 from the testing set), we obtained 424 high-quality, informative 2D slices for training/validation and 92 slices for testing.

3.2 Interpolation method

Our methodology for generating synthetic brain MRI images builds upon established principles in generative modeling while introducing novel approaches for maintaining anatomical consistency. We present a comprehensive framework that progressively extends from basic two-image interpolation to more sophisticated multi-image techniques, each designed to address specific challenges in medical image synthesis.

3.2.1 Interpolation between two images

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 CNN (Convolutional Neural Network) 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.

3.2.2 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:

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

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.

3.2.3 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.

3.3 Interpolation configurations

In this study, we explored several interpolation configurations to generate augmented brain MRI images. These configurations aimed to address the challenges of spatial distance, biological correctness, and latent space manipulation. The following subsections describe each configuration in detail.

3.3.1 Random Interpolated images

In this configuration, we interpolated between two randomly selected images without considering their spatial proximity. The generated image was obtained by a linear combination of the two input images, with a randomly generated weight α assigned to the first image and (1 - α) assigned to the second image. However, this approach might generate biologically incorrect brain images when the selected images are spatially distant (e.g., when interpolating between a slice from the middle of the brain volume and one from the extremities).

To address this issue, we have interpolated spatially close slices and slices belonging to the same patient. To do so, we have tested the data augmentation in several configurations, which are described in the following subsections. These configurations aim to ensure that the interpolated images are biologically plausible and maintain the structural integrity of the brain anatomy.

3.3.2 Spatially-aware interpolation configurations

- **Even and Odd index image interpolation**

  To ensure spatial proximity, we interpolated between images with even and odd indexes separately. For even index interpolation, we generated new images by interpolating between slices at positions 2 and 4, and compared the result with the slice at position 3. Similarly, for odd index interpolation, we interpolated between slices at positions 1 and 3, and compared the result with the slice at position 2. This approach maintains the spatial continuity of the brain structure.

- **Latent space noise injection**

  We introduced random noise in the latent space during the interpolation process to explore an additional data augmentation approach. The noise values were sampled from a uniform distribution between -0.01 and 0.01. This configuration was implemented as a complementary experiment to assess the impact of noise injection on the model's performance, rather than as an optimization of the previous interpolation methods.

- **Multi-image interpolation with distance-based weighting**

  We propose an approach for multi-image interpolation in the latent space that incorporates distance-based weighting to generate more biologically plausible brain images. Our method interpolates between n source images while giving higher importance to centrally located images through a carefully designed weighting scheme.

  The interpolated image is obtained through a weighted combination of latent representations. The weighting coefficients are generated through a three-step process that ensures higher weights for central images while maintaining controlled randomness:

  Initial weights are calculated based on the distance from the central point:

$$W_i = 1 - \frac{|i - \lfloor n/2 \rfloor|}{n+1} \tag{2}$$

  Where:

  - W_i represents the initial weight for the i-th image
  - i is the index of the current image in the sequence (0-based)
  - n is the total number of images being interpolated
  - $\lfloor n/2 \rfloor$ represents the floor division, giving us the central position
  - The denominator $(n + 1)$ ensures weights are properly normalized

  This weighting scheme ensures that images closer to the center of the sequence receive higher initial weights, reflecting their greater relevance to the target interpolation position.

  Random modulation is applied:

$$R_i = W_i \cdot rand(0, 1) \tag{3}$$

Normalization is performed to ensure the sum of coefficients equals 1:

$$\alpha_i = round(\frac{R_i}{\sum_{j=0}^{n-1} R_j}, 2) \tag{4}$$

In our specific implementation with $n = 4$ source images (at indices $i-2$, $i-1$, $i+1$, and $i+2$), this weighting scheme naturally assigns higher importance to the central images ($i-1$ and $i+1$) compared to the outer images ($i-2$ and $i+2$). The interpolated image is then compared with the ground truth image at position i. The normalization step ensures that while the weights maintain their relative proportions (favoring central images), their sum equals 1, effectively converting them into percentage contributions. For example, if $R_i = \{0.3, 0.5, 0.8, 0.4\}$, the normalization would yield $\alpha_i = \{0.15, 0.25, 0.40, 0.20\}$, meaning the central images contribute 65% of the final interpolation while maintaining mathematical validity.

4 Results

Our evaluation strategy aligns directly with our proposed methodological approaches, examining each interpolation configuration through both generated image quality and segmentation performance. This dual analysis approach allows us to assess both technical effectiveness and clinical utility of our methods.

Medical image analysis requires thorough validation through both quantitative metrics and qualitative assessment. We begin by analyzing our three interpolation configurations using visual inspection and Mean Squared Error (MSE) measurements, followed by an evaluation of their impact on downstream segmentation tasks. This comprehensive framework enables us to assess the practical value of our augmentation strategies in medical image processing applications.

4.1 Interpolation Results

We evaluate three interpolation configurations designed for anatomical consistency: sequential slice interpolation, latent space noise injection, and distance-weighted multi-image interpolation. Our analysis employs Mean Squared Error (MSE) not to measure exact reproduction, but rather as an indicator of anatomical plausibility, where lower values suggest successful preservation of biological structures while allowing beneficial variations.

4.1.1 Even and Odd index image interpolation

The even-indexed interpolation approach demonstrates the effectiveness of working with spatially adjacent slices.

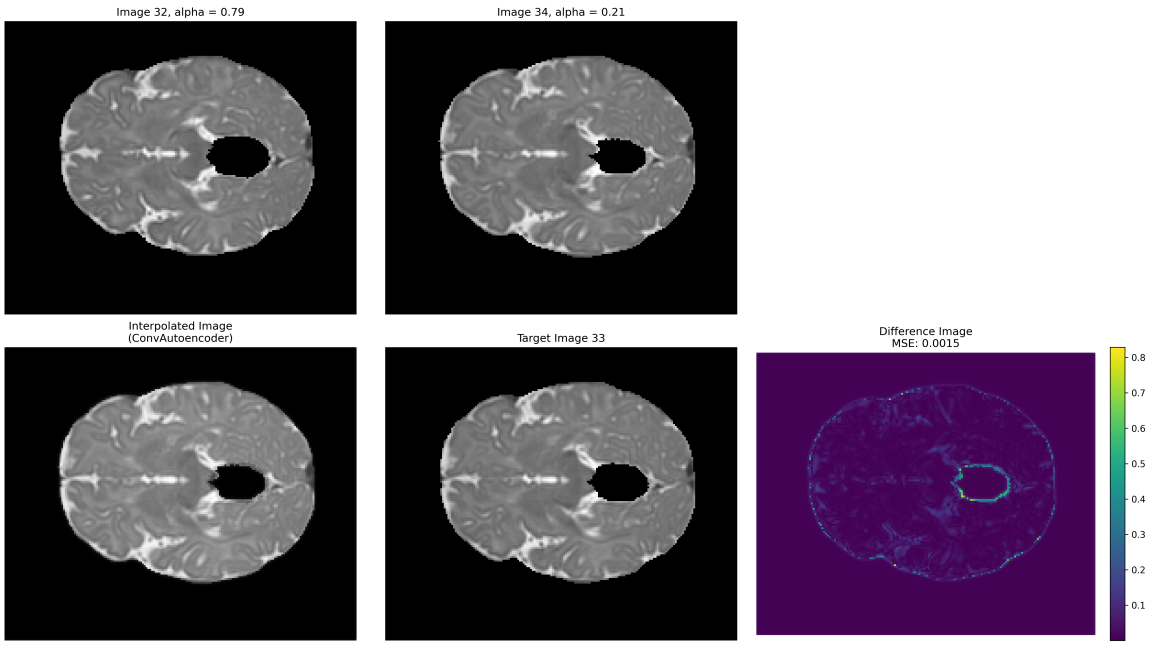


Figure 3: Even-indexed interpolation results demonstrating the effectiveness of our approach. Top row: source images (Image 32, $\alpha_1 = 0.79$ and Image 34, $\alpha_2 = 0.21$) used as input for interpolation. Bottom row: (left) the resulting interpolated image generated by our method, (middle) the target reference image (Image 33) for comparison, and (right) the difference map highlighting structural differences with an MSE of 0.0015. The minimal differences in the difference map, particularly in critical brain regions, demonstrate the method's ability to preserve anatomical structures while generating plausible intermediate states.

As shown in Figure 3, the interpolation between Images 32 and 34 produces a synthetic image that successfully captures the anatomical features of brain structures. The interpolation weights ($\alpha = 0.79$ and 0.21) were automatically determined to optimize the blend between the source images. The difference map reveals that most variations are confined to the brain's peripheral regions and ventricles, while maintaining high fidelity in brain tissue areas. The achieved MSE of 0.0015 indicates strong structural preservation in the generated image. This configuration particularly excels at maintaining the continuity of brain structures across consecutive slices, which is crucial for preserving biological plausibility in the generated images. The results sug-

gest that this approach effectively captures the gradual anatomical transitions that occur between adjacent MRI slices.

4.1.2 Latent space noise injection

The introduction of controlled noise in the latent space increases the diversity of the images generated. By injecting random noise values between -0.01 and 0.01, we aimed to introduce subtle variations while preserving anatomical correctness.

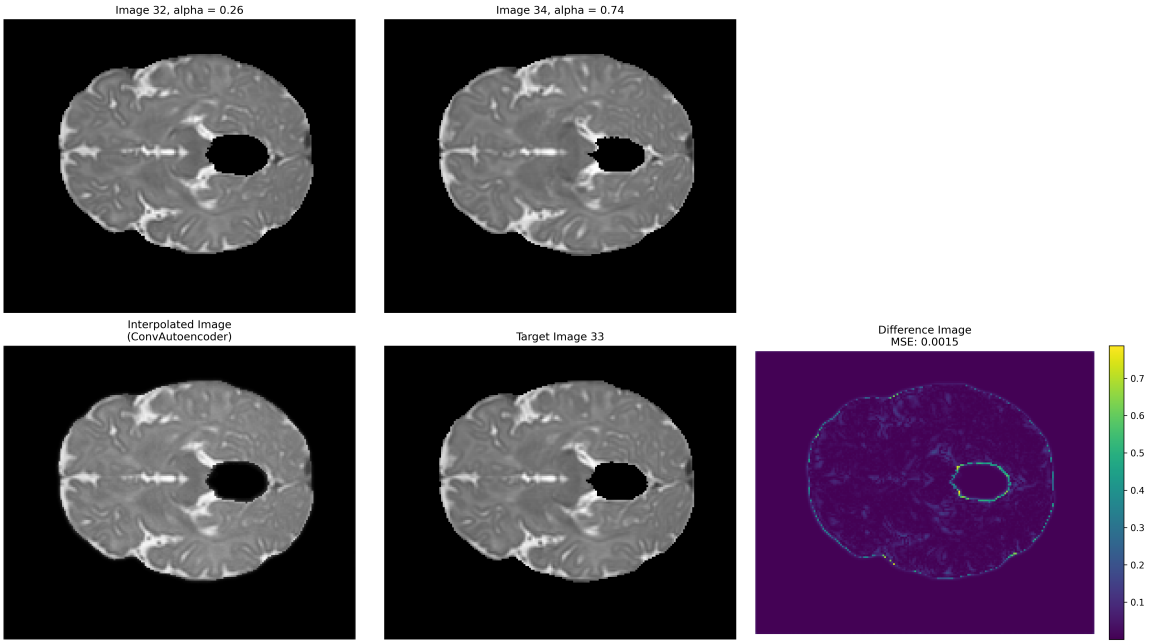


Figure 4: Interpolation results incorporating latent space noise. Top row: source images (Image 32, $\alpha = 0.26$ and Image 34, $\alpha = 0.74$). Bottom row: interpolated result, target image (Image 33), and difference map with MSE = 0.0015. The noise injection introduces subtle variations while maintaining anatomical integrity.

As demonstrated in Figure 4, this approach successfully generates realistic brain MRI images with an MSE of 0.0015. The difference map exhibits a pattern that confirms the preservation of key anatomical structures, with variations primarily in non-critical regions. This configuration's ability to maintain image quality while incorporating noise suggests its potential utility in generating diverse yet anatomically correct synthetic images. The controlled noise injection could prove valuable for improving model robustness in downstream applications.

4.1.3 Multi-image interpolation with distance-based weighting

The multi-image interpolation approach introduces a method for generating synthetic brain MRI images by leveraging information from multiple adjacent slices. This configuration utilizes four source images with carefully calculated weights based on their spatial distance from the target slice.

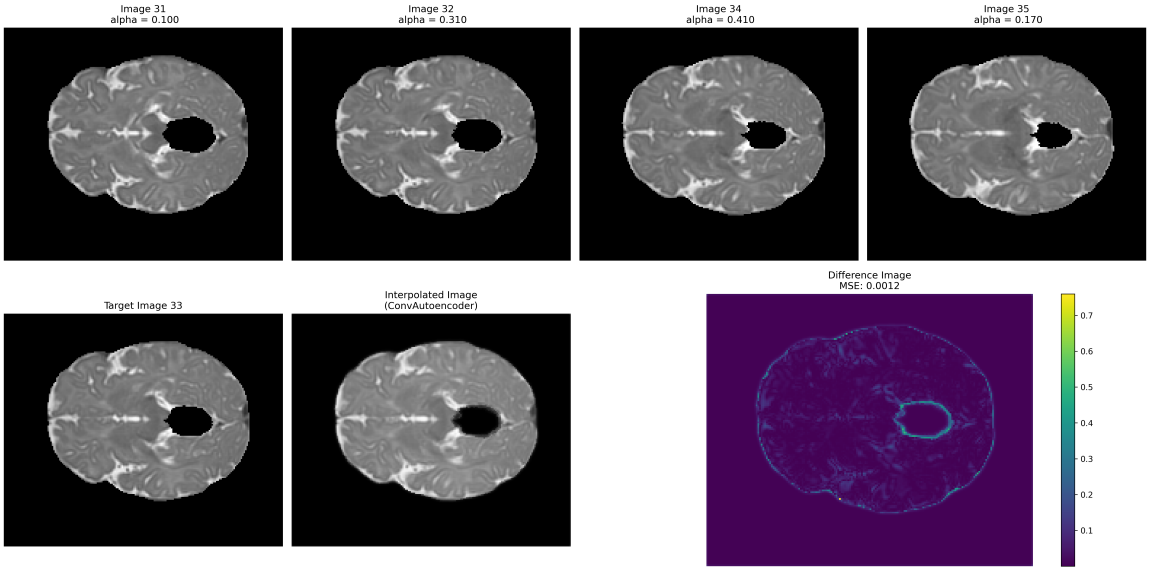


Figure 5: Multi-image interpolation results demonstrating our distance-based weighting approach. Top row: Four source images (Images 31, 32, 34, and 35) with their computed weights ($\alpha = 0.100$, 0.310, 0.410, and 0.170) showing how the weighting scheme favors central images. Bottom row: (left) the target reference image (Image 33), (middle) the synthesized interpolated image, and (right) the difference map with MSE = 0.0012. The low MSE and minimal structural differences in the difference map demonstrate the effectiveness of our multi-image approach in preserving anatomical integrity while incorporating features from multiple slices.

As illustrated in Figure 5, the interpolation between images 31, 32, 34, and 35 produces high-quality results with an MSE of 0.0012. The weighting scheme ($\alpha = 0.100$, 0.310, 0.410, and 0.170) reflects our distance-based strategy, allocating higher importance to slices closer to the target position. The difference map reveals highly consistent preservation of anatomical structures, with minimal deviations across all brain regions. This approach demonstrates particular strength in maintaining the subtle anatomical transitions that occur across multiple consecutive slices.

4.1.4 Quantitative Analysis

Each of these configurations demonstrates the capability to generate high-quality synthetic brain MRI images while maintaining anatomical plausibility. The MSE scores presented in Table 1 provide quantitative validation of each method's effectiveness in preserving structural integrity. It's important to note that our primary objective was not to exactly reproduce the target images, but rather to generate biologically plausible variations that could enrich our dataset. In this context, the low MSE scores are particularly encouraging as they indicate that our generated images maintain anatomical correctness while introducing subtle variations. These controlled differences are actually beneficial for data augmentation purposes, as they help introduce meaningful diversity into the dataset without compromising anatomical validity. These results suggest that our interpolation strategies offer viable approaches for augmenting brain MRI datasets while maintaining the critical anatomical features necessary for medical imaging applications, with the added benefit of introducing controlled variability that can enhance model robustness in downstream tasks.

Metric	Even/Odd index interpolation	Latent space noise injection	Multi-image interpolation
Mean Squared Error	0.0002 ± 0.0043	0.0002 ± 0.0045	0.0005 ± 0.0055

Table 1: MSE scores achieved by different interpolation configurations, showing mean and standard deviation. Lower values indicate better structural preservation, while the presence of standard deviation confirms the generation of diverse yet anatomically plausible variations.

4.2 Segmentation experiments

To evaluate the effectiveness of our interpolation strategies for data augmentation, we conducted extensive segmentation experiments focusing on brain tissue classification. We first established a baseline by training a segmentation model on the original dataset, which served as our control for performance comparison. The model was designed to segment four distinct classes: background (Class 0), cerebrospinal fluid (Class 1), gray matter (Class 2), and white matter (Class 3). Subsequently, we augmented the training data using each of our interpolation configurations and systematically evaluated their impact on segmentation performance across all tissue classes.

4.2.1 Experimental Setup

The experimental framework was designed to rigorously assess the impact of different augmentation strategies on segmentation quality. As shown in Table 2, each

interpolation configuration generated a different number of synthetic images, leading to varying dataset compositions. The even and odd index interpolation methods each produced 360 new images, maintaining spatial consistency within their respective sequences. When combining both even and odd augmentations, this doubled to 720 images, providing a more comprehensive coverage of anatomical variations. The multi-image interpolation approach, leveraging information from four adjacent slices, generated 695 synthetic samples, offering a balance between dataset size and anatomical coherence. These generated images were strategically combined with the original dataset to create augmented training sets of varying sizes. To ensure proper evaluation, we maintained consistent train-validation splits throughout all experiments, with approximately 80% of data allocated for training and 20% for validation. This splitting strategy was carefully designed to preserve the distribution of anatomical features across both sets.

Configuration	Generated dataset size	Augmented dataset size	Train size	Validation size
Even/Odd index interpolation	360	784	627	157
Even + Odd augmented dataset	720	1144	915	229
Latent space noise injection	360	784	627	157
Multi-image interpolation	695	1119	895	223

Table 2: Dataset sizes for different interpolation configurations

For rigorous evaluation, we implemented a comprehensive training protocol:

- Each configuration was tested with weight values ranging from 0.05 to 0.95 (step size 0.05).
- Models were trained three times per configuration to ensure statistical reliability.
- Results were aggregated across runs to obtain robust performance metrics.

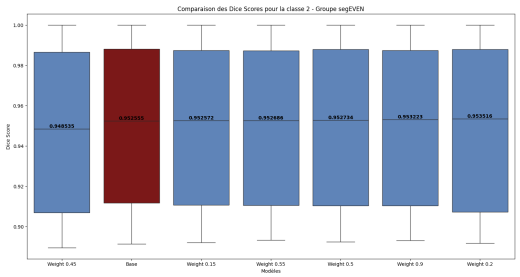
4.2.2 Segmentation Performance Analysis

Quantitative results

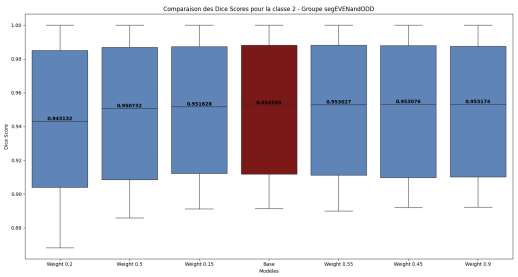
Our comprehensive analysis of segmentation performance revealed varying impacts across different augmentation strategies and tissue classes. While some configurations demonstrated notable improvements, others showed mixed results depending on the weight configurations and tissue types involved. For clarity in visualization, we selected representative weights (α = 0.2, 0.15, 0.9, 0.5, 0.45, 0.55) that best illustrate the range of performance across configurations. The most significant improvements were observed in gray matter (Class 2) and white matter (Class 3) segmentation,

though with different patterns of enhancement across configurations as you can see in (Figure 6 and 7)

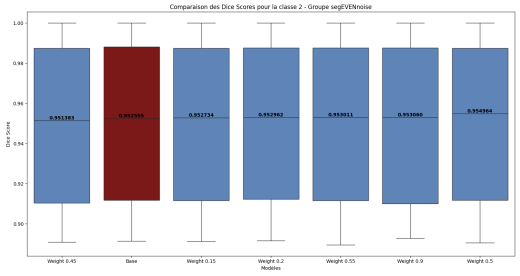
- **Gray Matter (Class 2) Performance:** The box plots for Class 2 segmentation (Figure 6) compare our different configurations against the baseline model (shown in red). The even-indexed configuration shows moderate improvements for some weight values, though not consistently across all configurations. In the even-odd approach, performance is comparable to the baseline, with some weights showing slight improvements and others showing similar performance. The noise-injected configuration demonstrates the most varied performance, with some weights achieving better results than the baseline but with increased variance. The multi-image approach shows the most stable performance across different weights, maintaining performance levels close to or slightly above the baseline with consistent variance.



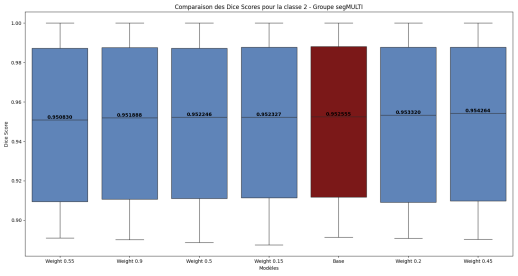
(a) Dice scores comparison for even-indexed interpolation augmentation



(b) Dice scores comparison for combined even and odd interpolation augmentation



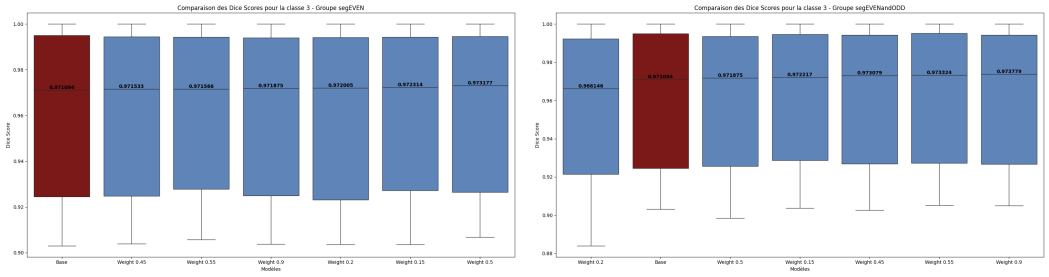
(c) Dice scores comparison for noise-injected interpolation augmentation



(d) Dice scores comparison for multi-image interpolation augmentation

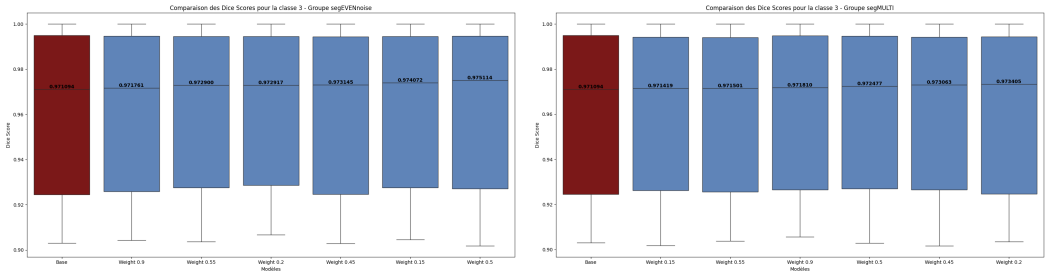
Figure 6: Comparative analysis of gray matter (Class 2) segmentation performance across different augmentation strategies. Each plot compares the baseline model against various weight configurations.

- **White Matter (Class 3) Performance:** For Class 3 (Figure 7), we observed improvements compared to the baseline model across several configurations. The even-indexed and even-odd configurations showed improvements over the baseline for some weight values. The noise-injected approach achieved notable improvements, despite showing slightly higher variance in performance. The multi-image approach demonstrated stable performance with moderate improvements over the baseline for certain weight configurations.



(a) Dice scores comparison for even-indexed interpolation augmentation

(b) Dice scores comparison for combined even and odd interpolation augmentation



(c) Dice scores comparison for noise-injected interpolation augmentation

(d) Dice scores comparison for multi-image interpolation augmentation

Figure 7: Comparative analysis of white matter (Class 3) segmentation performance across different augmentation strategies. Each plot compares the baseline model against various weight configurations.

- **Cross-Class Analysis:** Examining the performance across all tissue classes (Table 3), we observe that while gray and white matter showed improvements, background and cerebrospinal fluid experienced minor performance decreases ($\leq 0.06\%$). The noise-injected configuration achieved the highest improvements for both gray matter ($+0.25\% \pm 0.61\%$) and white matter ($+0.41\% \pm 0.71\%$). The multi-image approach showed the most balanced performance across all classes, with lower standard deviations indicating more stable improvements.

Method	Class			
	0: Background	1: Cerebrospinal Fluid	2: Gray Matter	3: White Matter
segEVEN	$-0.03 \pm 0.09\%$	$-0.03 \pm 0.25\%$	$0.10 \pm 0.52\%$	$0.24 \pm 0.22\%$
segEVENandODD	$-0.04 \pm 0.10\%$	$-0.06 \pm 0.16\%$	$0.07 \pm 1.05\%$	$0.28 \pm 0.79\%$
segEVENnoise	$-0.03 \pm 0.07\%$	$-0.02 \pm 0.26\%$	$0.25 \pm 0.61\%$	$0.41 \pm 0.71\%$
segMULTI	$-0.03 \pm 0.03\%$	$-0.04 \pm 0.15\%$	$0.18 \pm 0.36\%$	$0.24 \pm 0.28\%$

Table 3: Segmentation performance improvements (%) over baseline model (trained without data augmentation) across different data augmentation configurations. Values represent relative changes in Dice coefficient scores, with positive values indicating improvement. Standard deviations ($\pm$) reflect consistency across multiple test runs.

Qualitative results

While the quantitative metrics provide important numerical validation of our methods, the visual assessment of segmentation results offers crucial insights into the anatomical coherence and clinical relevance of our approach. The following qualitative analysis examines specific cases that illustrate both the strengths and limitations of each augmentation strategy. To comprehensively evaluate the effectiveness of our augmentation strategies, we performed a detailed qualitative analysis comparing the baseline model's performance with each augmentation configuration.

The baseline model demonstrated strong overall performance, achieving Dice scores of 1.00, 0.97, 0.91, and 0.93 for background, cerebrospinal fluid, gray matter, and white matter classes, as you can see in Figure 8. This robust foundation provided an excellent reference point for assessing the impact of our various augmentation strategies.

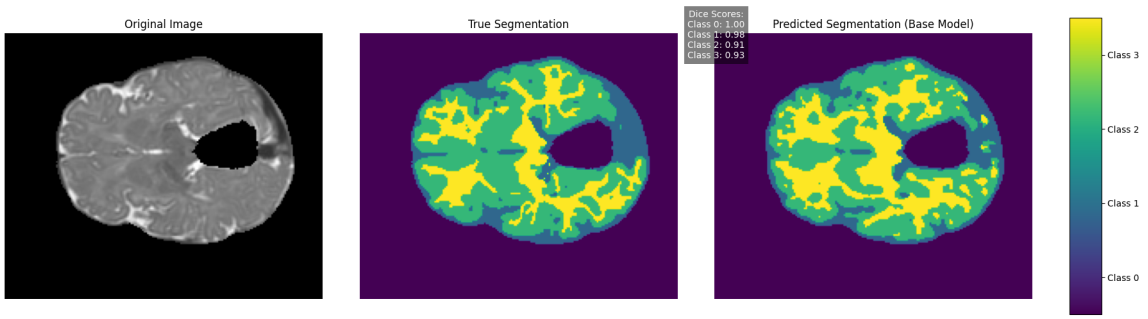


Figure 8: Baseline segmentation comparison. Left: Original T2-weighted brain MRI slice. Middle: Ground truth segmentation showing four distinct tissue classes (Class 0: Background, Class 1: CSF, Class 2: Gray Matter, Class 3: White Matter). Right: Predicted segmentation by the base model, achieving Dice scores of 1.00, 0.97, 0.91, and 0.93 for respective tissue classes.

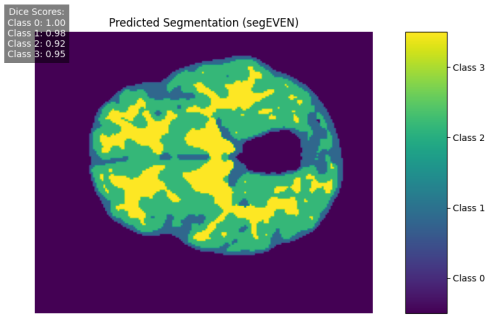


Figure 9: Even-indexed interpolation results (segEVEN).

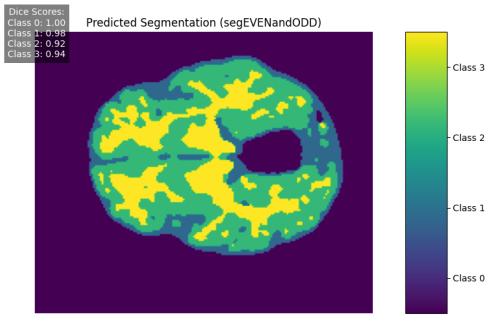


Figure 10: Combined even-odd interpolation results (segEVENandODD).

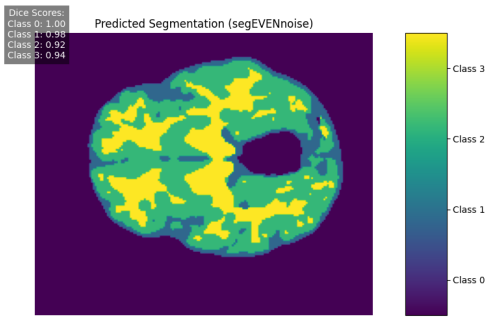


Figure 11: Noise-injected interpolation results (segEVENnoise).

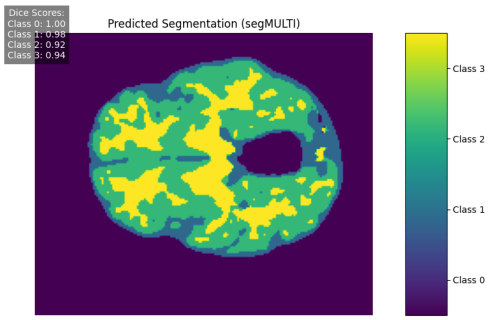


Figure 12: Multi-image interpolation results (segMULTI).

Figure 13: Comparative analysis of segmentation results across different augmentation strategies. (a) Augmented segmentation predictions showing tissue class differentiation (Class 0: Background, Class 1: CSF, Class 2: Gray Matter, Class 3: White Matter). (b) Performance metrics: Dice scores for each configuration displayed in top left corners show quantitative improvements over baseline model (baseline scores: 1.00, 0.97, 0.91, 0.93 for respective tissue classes).

Each augmentation configuration exhibited distinct characteristics in improving upon the baseline performance (Figure 8). The even-indexed configuration (Figure 9) demonstrated enhanced segmentation of gray-white matter boundaries, with difference maps (Figure 15) revealing a reduction in false positive predictions within gray matter regions. This improvement aligned with our quantitative findings of +0.10% and +0.24% increases in gray and white matter segmentation accuracy, respectively.

The combined even-odd configuration showed particular strength in boundary consistency (Figure 10) , especially notable in the cerebrospinal fluid regions where noise

reduction was evident. This configuration achieved modest but meaningful improvements of +0.07% in gray matter and +0.28% in white matter segmentation, with the difference maps (Figure 16) highlighting more precise tissue interface definitions.

Perhaps most notably, the noise-injected configuration (Figure 11) exhibited the most substantial improvements over the baseline model, with difference maps (Figure 17) revealing superior handling of partial volume effects. This configuration's success in achieving +0.25% and +0.41% improvements in gray and white matter segmentation, respectively, suggests that controlled noise injection enhances the model's ability to handle tissue boundary ambiguities.

The multi-image configuration (Figure 12) demonstrated the most balanced performance improvement across all tissue classes. Its difference maps (Figure 18) showed consistent enhancement patterns with notably lower variance compared to other configurations, achieving improvements of +0.18% in gray matter and +0.24% in white matter segmentation. This stability across tissue classes suggests that incorporating information from multiple adjacent slices provides a more robust foundation for segmentation.

These qualitative observations, particularly visible in the difference maps (Figure 15 to 18), complement our quantitative findings and provide insight into how each augmentation strategy influences the segmentation process. The visual analysis reveals that while the baseline model provides strong foundational performance, each augmentation strategy offers unique advantages in handling specific aspects of the segmentation task, from boundary definition to tissue class differentiation.

Particularly noteworthy is the consistent pattern of improvement in gray and white matter segmentation across all configurations, despite the baseline model's already strong performance in background and cerebrospinal fluid classes. This pattern suggests that our augmentation strategies specifically enhance the model's capability to handle more challenging tissue classifications while maintaining high performance in simpler class distinctions.

5 Conclusion

This study presents a comprehensive investigation of interpolation-based data augmentation strategies for brain MRI analysis. Our research demonstrates that carefully designed interpolation methods can effectively generate anatomically plausible synthetic images while improving downstream task performance. The three proposed configurations—even-odd index interpolation, latent space noise injection, and multi-image interpolation with distance-based weighting—each offer unique advantages in preserving anatomical structures while introducing beneficial variations. The quantitative results show significant improvements in segmentation performance, partic-

ularly for challenging tissue classes. These improvements, while modest in absolute terms, are meaningful given the already strong baseline performance and the critical nature of medical image analysis. The low MSE scores ($\leq 0.0005\%$) across all configurations confirm the ability of our methods to maintain anatomical plausibility while introducing controlled variations. Qualitative analysis reveals that our augmentation strategies successfully preserve critical anatomical features while enhancing boundary definition and tissue class differentiation. The multi-image interpolation approach, in particular, demonstrates remarkable stability across all tissue classes, suggesting that incorporating information from multiple adjacent slices provides a more robust foundation for synthetic image generation.

6 Future Work

This work suggests several important directions for future research in brain MRI data augmentation and analysis. First, we propose a deeper investigation into the statistical properties of pixel distributions across different tissue classes. This analysis would help explain the varying performance improvements observed across different tissue types and could lead to more targeted augmentation strategies.

Furthermore, we recognize the importance of validating our methods across diverse datasets. While our current results are promising, testing these augmentation techniques on different MRI protocols, field strengths, and pathological conditions would establish their broader applicability. This cross-validation could include datasets from different institutions, scanner manufacturers, and patient populations, helping to assess the robustness and generalizability of our approach.

Further technical developments could focus on integrating anatomical prior knowledge into the interpolation weights calculation, which may improve biological fidelity of the generated images. The relationship between interpolation parameters and specific clinical applications also warrants investigation, particularly in the context of different pathological conditions. Our noise injection experiments could be expanded to examine a broader range of noise levels and distributions, potentially revealing optimal noise parameters for different tissue types and imaging conditions. Additionally, extending our methods to true 3D volumetric interpolation could better capture the spatial continuity of anatomical structures.

These future investigations would advance not only the technical capabilities of medical image augmentation but also their practical application in clinical settings like medical image compression. Such developments could ultimately lead to more robust diagnostic tools and improved patient care through enhanced medical image analysis.

7 Appendix

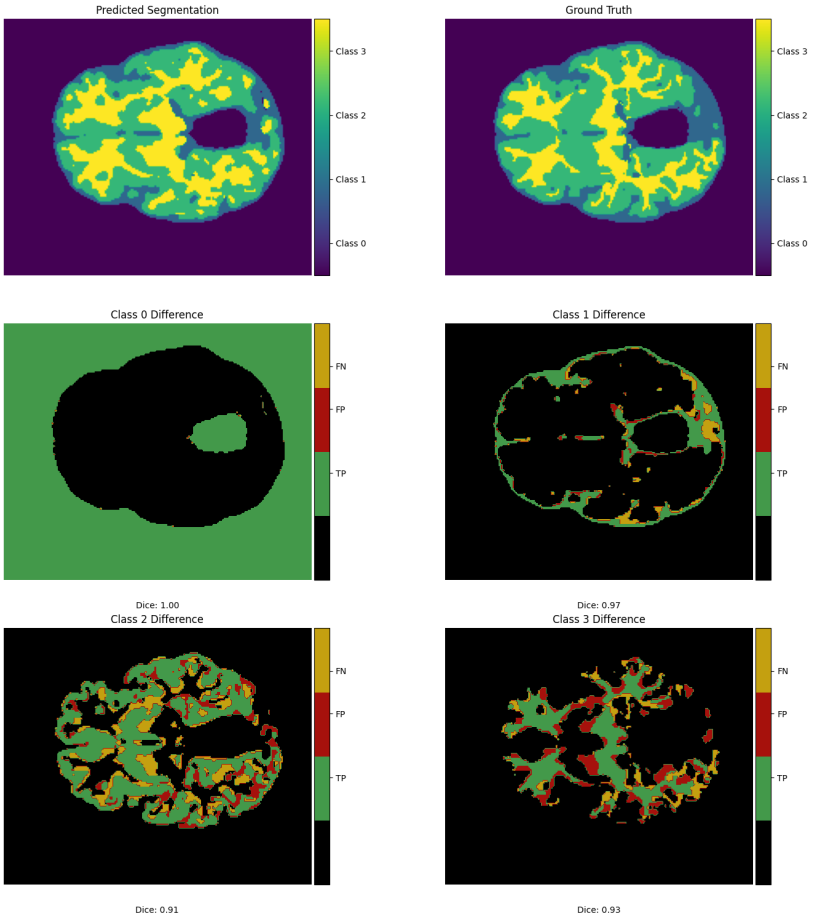


Figure 14: Baseline model segmentation performance analysis. Top row: Comparison between predicted segmentation (left) and ground truth (right) with class labels (0: Background, 1: CSF, 2: Gray Matter, 3: White Matter). Bottom rows: Class-wise difference maps showing True Positives (TP), False Positives (FP), and False Negatives (FN) for each tissue class, with corresponding Dice scores.

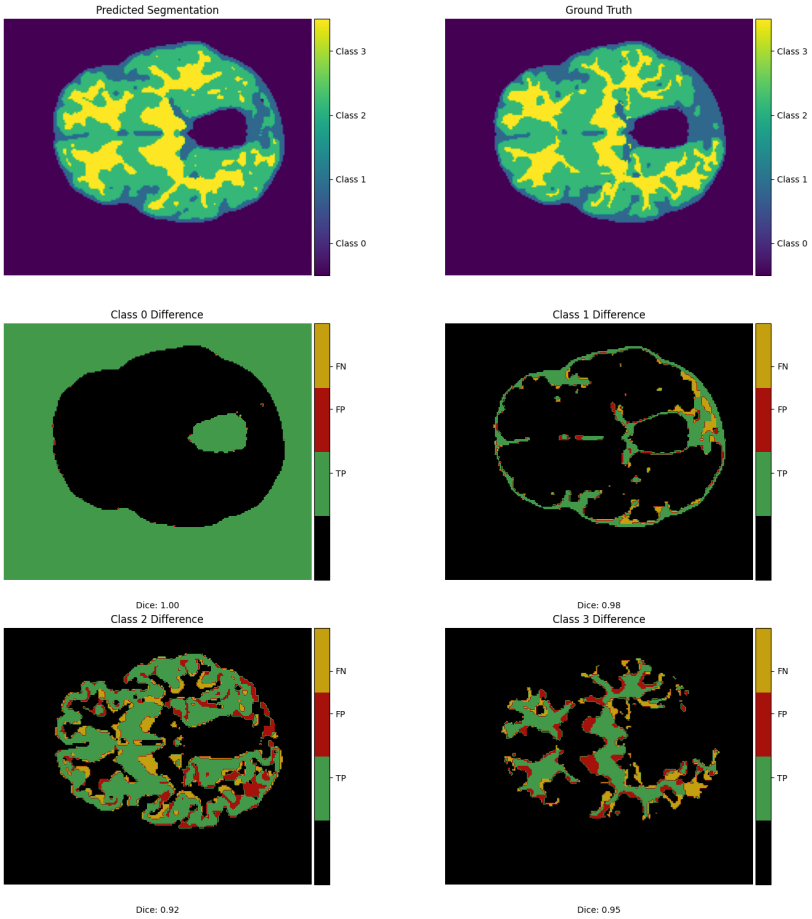


Figure 15: Even-indexed interpolation augmentation results showing segmentation predictions and class-wise difference maps. Top row: Comparison between predicted segmentation (left) and ground truth (right) with class labels (0: Background, 1: CSF, 2: Gray Matter, 3: White Matter). Enhanced boundary delineation is evident in gray-white matter interfaces, with reduced false positives in gray matter regions.

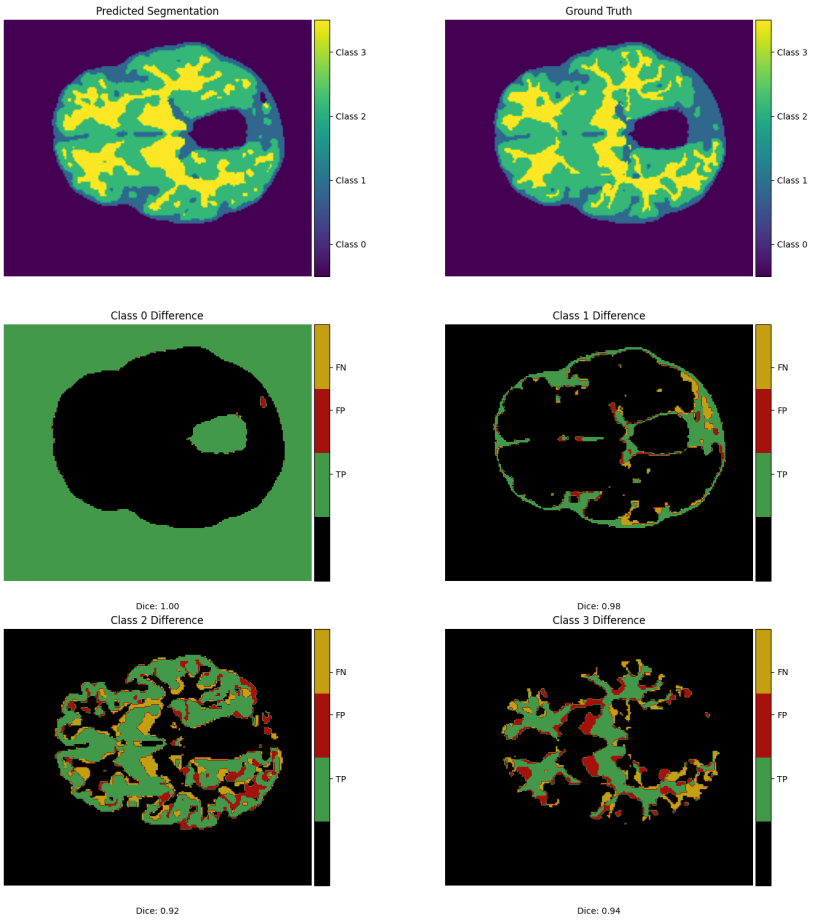


Figure 16: Combined even-odd interpolation augmentation results demonstrating improved boundary consistency and reduced noise in CSF regions. Top row: Comparison between predicted segmentation (left) and ground truth (right) with class labels (0: Background, 1: CSF, 2: Gray Matter, 3: White Matter)

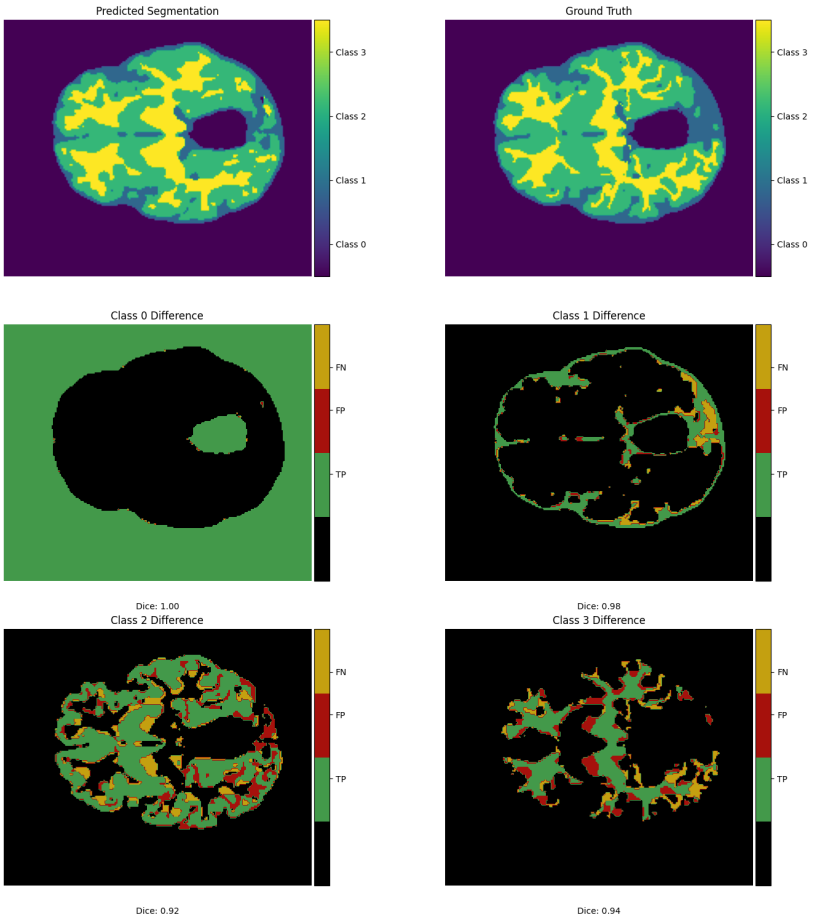


Figure 17: Noise-injected interpolation augmentation results highlighting superior handling of partial volume effects and improved tissue boundary detection. Top row: Comparison between predicted segmentation (left) and ground truth (right) with class labels (0: Background, 1: CSF, 2: Gray Matter, 3: White Matter)

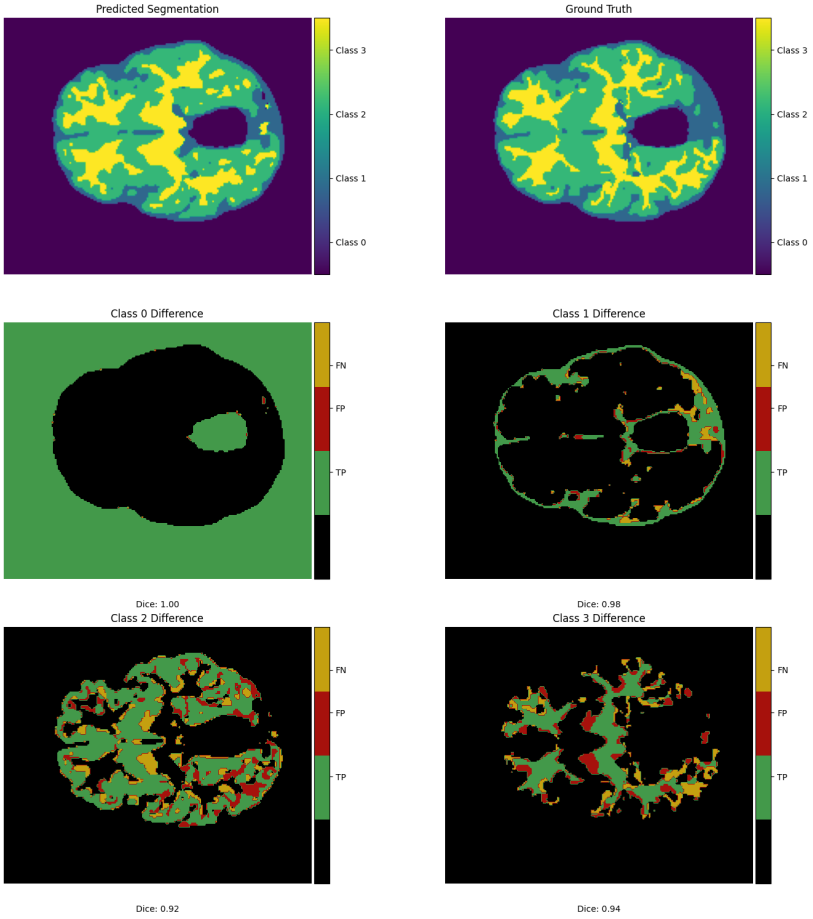


Figure 18: Multi-image interpolation augmentation results showing balanced performance improvements across all tissue classes. The difference maps demonstrate consistent enhancement patterns with notably lower variance compared to other configurations.. Top row: Comparison between predicted segmentation (left) and ground truth (right) with class labels (0: Background, 1: CSF, 2: Gray Matter, 3: White Matter)

8 Bibliography

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

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