Expert Systems With Applications 284 (2025) 127970

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Latent diffusion for arbitrary zoom MRI super-resolution

Jorge Andrés Mármol-Rivera [a,*], José David Fernández-Rodríguez [a,b,c], Beatriz Asenjo [d], Ezequiel López-Rubio [a,b,c]

[a] University of Málaga, Computer Science and Languages Department, Bulevar Louis Pasteur, 35, Málaga, 29071, Andalucía, Spain
[b] University of Málaga, ITIS Software, Bulevar Louis Pasteur, 35, Málaga, 29071, Andalucía, Spain
[c] Biomedical Research Institute of Málaga (IBIMA - Plataforma BIONAND), C/Doctor Miguel Díaz Recio, 28, Málaga, 29010, Andalucía, Spain
[d] Hospital Regional Universitario de Málaga, Avenida de Carlos Haya, 84, Málaga, 29010, Andalucía, Spain

ARTICLE INFO

Keywords:
MRI
Deep learning
Super-resolution
Latent diffusion model

ABSTRACT

In various image processing tasks, enhancing resolution is a fundamental challenge, particularly along specific axes where resolution tends to be lower. This limitation can hinder the performance of models in tasks such as medical image analysis. Traditional approaches often involve interpolation techniques, but they may lead to loss of information or introduce artifacts. Recently, deep learning-based methods, especially those utilizing latent spaces, have shown promise in addressing this issue. Because typical super-resolution methods are designed for 2D images, they can easily be applied to increase resolution in two of the axes in a volumetric MRI, but not the other axis. While volumetric (3D) deep learning models for super-resolution have been proposed, they have very high computational requirements, even if the region of interest to super-resolve does not span the whole volume. In our work, we propose a novel approach that uses a diffusion latent model to increase resolution along an arbitrary axis. Our method involves transforming input images into a latent space, where a U-Net model is employed to capture high-level features. Crucially, just before decoding, we introduce a linear interpolation in the latent space to enhance resolution along the specified axis. This interpolated latent representation is then decoded by the decoder, yielding images with increased resolution, thus achieving a resolution across all axes and, therefore, an increase in resolution of the entire volume, using a 2D deep learning model rather than a fully-fledged 3D model. The proposal has been extensively tested with a wide range of brain lesions and brain tumor images of T1, T2, and FLAIR modes. The experimental comparison with several state-of-the-art methods has consistently shown the advantages of our approach.

1. Introduction

In this modern age, characterized by an extraordinary surge in technological innovations, the advent of generative models represents a significant milestone in the evolution of image processing techniques (Chitradevi & Srimathi, 2014). These sophisticated algorithms have rapidly become one of the most influential tools in the arsenal of digital technology, with their unique ability to create new, varied data that mirrors the intricacy and diversity of original datasets. The applications of generative models span a vast array of industries and disciplines, illustrating their versatility and the depth of their impact.

Deep learning models, and especially generative models, have found many applications in science and technology, and especially in the field of medical diagnostics (Chen, Pawlowski, Rajchl, Glocker, & Konukoglu, 2018; Dimitriadis, Trivizakis, Papanikolaou, Tsiknakis, & Marias, 2022; Matsubara, Tashiro, & Uehara, 2019; Ren & Zhou, 2020; Song et al., 2023). The ability to generate accurate, high-resolution images can be a game-changer in diagnosing and treating a wide range of medical conditions. By producing detailed and realistic representations of human anatomy, generative models assist medical professionals in identifying anomalies and pathologies with greater precision. This can significantly enhance the accuracy of diagnoses (Kermany et al., 2018), the planning of surgical procedures (Khalid, Goldenberg, Grantcharov, Taati, & Rudzicz, 2020), and the personalization of treatment plans (Liefaard et al., 2021), ultimately leading to better patient outcomes.

The rest of the introduction presents preliminary concepts about LDM generative models (Section 1.1), presents previous work on deep learning models for MRI super-resolution (Section 1.2), and outlines the main contributions of this work (Section 1.4).

* Corresponding author.
E-mail addresses: jorgemarmol@uma.es (J.A. Mármol-Rivera), josedavid@uma.es (J.D. Fernández-Rodríguez), asenjob@gmail.com (B. Asenjo), ezeqlr@lcc.uma.es (E. López-Rubio).

https://doi.org/10.1016/j.eswa.2025.127970
Received 21 November 2024; Received in revised form 21 April 2025; Accepted 29 April 2025
Available online 2 May 2025

1.1. Autoencoders and latent space representation in latent diffusion models

Deep generative models have gained substantial attention for their ability to synthesize high-quality data. Among these models, Auto-Encoders, whose acronym is AEs, Bank, Koenigstein, and Giryes (2023); Michelucci (2022) have proven to be a powerful class of neural networks, offering efficient compression and reconstruction of data by learning to encode high-dimensional inputs into a lower-dimensional latent space. The latent space serves as a compact representation of the input, capturing the most salient features while discarding noise or less relevant information. This encoding-decoding architecture enables auto-encoders to reconstruct the original input from its latent representation.

Building upon the foundations of auto-encoders, recent advancements have led to the development of more sophisticated models that operate within this latent space, particularly for data generation. One such innovation is the Latent Diffusion Model, whose acronym is LDM (Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022), a novel approach that combines the power of diffusion processes with the efficiency of latent space representations. Diffusion models, traditionally applied in pixel space, simulate a gradual destruction of data by introducing noise and subsequently learning to reverse this process, generating new samples by removing the noise step by step. While effective in generating high-quality outputs, these models are computationally expensive, particularly when applied to high-dimensional data such as images. The introduction of latent space diffusion addresses this computational bottleneck by shifting the diffusion process from pixel space to a more compact latent space. In Latent Diffusion Models, the auto-encoder first learns to map input data to a latent representation, and the diffusion process is then applied in this compressed space. This approach significantly reduces the computational overhead associated with high-dimensional data while preserving the generative quality of traditional diffusion models. By operating in latent space, LDMs can achieve scalable and efficient data generation with fewer computational resources, making them highly attractive for practical applications in large-scale generative tasks.

At the core of Latent Diffusion Models, there is a process of progressive denoising within the latent space. The model starts with a noisy version of the latent representation of the input data and refines it iteratively over several steps, gradually removing the noise and recovering the original structure. This process is governed by a stochastic differential equation (SDE) (Tang & Zhao, 2024), that models both the forward dynamics, where noise is added and the reverse dynamics, where noise is removed. The goal is to progressively refine the noisy latent representation until it closely resembles the original, clean data, enabling LDMs to generate highly realistic outputs. However, the success of this process relies heavily on an efficient architecture for handling the complex task of denoising at multiple scales. This is where the U-Net plays a crucial role (Ronneberger, Fischer, & Brox, 2015). In LDMs, U-Net serves as the backbone for the denoising step, providing a robust mechanism to process and refine the noisy latent representations. U-Net consists of an encoder-decoder structure connected by skip connections, which allow for the retention of fine-grained details that might otherwise be lost during the compression and expansion stages. The encoder captures hierarchical features of the input, reducing dimensionality, while the decoder reconstructs the data by progressively removing noise. Skip connections between corresponding layers of the encoder and decoder allow the model to combine both high-level and low-level information, making it highly effective for tasks where preserving both fine details and overall structure is essential. In LDMs, this multi-scale processing capability of U-Net ensures that the model can handle intricate textures and patterns during the denoising process, ultimately leading to the generation of highly realistic and coherent data samples.

Despite these strengths, Latent Diffusion Models also pose unique challenges. One critical issue is the design of the latent space itself. The quality and structure of the latent representation learned by the auto-encoder play a crucial role in determining the performance of the diffusion process. A poorly designed latent space may lead to suboptimal results, such as blurred or inaccurate reconstructions. Therefore, significant attention must be paid to the training and optimization of the auto-encoder to ensure that the latent space is sufficiently expressive while maintaining a compact dimensionality. Another challenge is the efficient modeling of the noise injection and denoising processes within the latent space, which requires careful tuning of the diffusion parameters and training schedule. Additionally, a well-regularized latent space can enable meaningful interpolations between data points, allowing for smooth transitions and variations within the latent space, which is highly desirable for generating diverse and coherent samples.

Interpolation in latent space has been a widely explored technique in generative models (Michelis & Becker, 2021), particularly with autoencoders, where it allows for smooth transitions between different data points by interpolating between their corresponding latent representations. This has been useful in generating novel data samples, enabling tasks such as image synthesis and style transfer by manipulating latent variables. However, LDMs offer additional advantages over traditional autoencoders, particularly in terms of high-resolution data generation. LDMs can model complex data distributions through the diffusion process in a latent space, allowing for more accurate and diverse generation. One key advantage of LDMs is their ability to perform super-resolution on latent representations, especially for slices or regions of an image, which enables the generation of fine details and higher-quality outputs. The U-Net architecture plays a crucial role in this process by improving both the resolution and noise suppression. As part of the denoising process, U-Net's encoder-decoder structure, with skip connections, allows it to retain fine-grained details from the original image while progressively refining the latent representation. This helps enhance the resolution of the image slices and removes noise that may arise during the diffusion process, ensuring that the final generated sample is both high-quality and accurate.

1.2. Related work in MRI super-resolution

Magnetic Resonance Imaging (MRI) is a medical imaging modality used to obtain detailed information about the human body's interior. However, MRI images often have limited resolution due to technical limitations, which can make interpretation and diagnosis difficult. Different approaches based on LDM to MRI super-resolution have been proposed to overcome this problem.

In image processing and analysis, the advancement of super-resolution techniques has progressed in tandem with the evolution of generative models, marking a significant leap in the ability to enhance the detail and clarity of images beyond their original resolution (Wang, Chen, & Hoi, 2020; Yang et al., 2019). Particularly in the context of volumetric imaging (MRI), the application of super-resolution methodologies has been transformative. Volumetric images, by their nature, often suffer from lower resolution in certain axes due to the limitations inherent in the imaging process. Super-resolution techniques address this challenge head-on, employing sophisticated algorithms to reconstruct high-resolution images from these lower-resolution datasets. This capability is not merely a technical achievement; it represents a pivotal improvement in the quality and usability of images across a variety of applications.

The significance of super-resolution technologies becomes especially pronounced in the field of medical image analysis (Chen et al., 2018; Dimitriadis et al., 2022; Matsubara et al., 2019; Ren & Zhou, 2020). Here, the quality of the image can directly influence the accuracy of diagnoses and the outcomes of treatments. For conditions where every pixel might hold crucial information about pathological changes, the ability to enhance image resolution can be a game-changer. Super-resolution helps medical professionals detect abnormalities with greater precision and confidence by providing clearer, more detailed views of anatomical structures. This is particularly relevant in MRI analysis,

where enhanced resolution can reveal subtle details that might otherwise go unnoticed in lower-resolution scans. Consequently, super-resolution is not just an advancement in image processing technology; it is a tool that has the potential to impact patient care significantly, offering pathways to earlier detection and more informed treatment decisions. There is, however, an additional hurdle in the case of volumetric data, such as MRIs, where data points are arranged in 3D (not pixels, but voxels): typical deep learning models for super-resolution work over 2D images, not 3D volumes. 2D super-resolution models can still be applied slice by slice for 3D data, but resolution is not increased in the dimension across slices. 2D models can be adapted and re-trained to 3D data, but 3D deep learning models typically require larger datasets (because they have more parameters to train) and even larger amounts of memory and computational power to be used, thus requiring more advanced and expensive hardware than 2D deep learning models.

Many different deep learning models have been used for super-resolution of MRI data, such as SRCNN, a classical CNN that has been adapted to MRI super-resolution both using a regularly spaced shifting mechanism for 3D data (Thurnhofer-Hemsi, López-Rubio, Domínguez, Luque-Baena, & Roé-Vellvé, 2020b) as well as an additional sub-pixel convolution layer for 2D data, i.e., applied slice by slice (Qiu, Zhang, Liu, Zhu, & Zheng, 2020). Other classical CNN architectures also applied for MRI super-resolution include residual networks, both for 2D slices (Shi et al., 2018) and 3D volumetric data (Du et al., 2020), as well as dense learning, also both for 2D slices (He, Hu, Wang, He, & Du, 2021) and 3D data (Du, Wang, Gholipour, He, & Jia, 2018). While classical CNNs treat all information from the input image equally, attention networks are CNNs with architectures crafted to induce them to pay more attention to relevant structures and attributes from the input 2D image, using, for example, feature and spatial attention (Wang, Zhu, He, Jia, & Du, 2022) or global descriptors of spatial information (Zhang, Li, Li, & Fu, 2021b). Transformer architectures are geared towards modeling and prediction of sequential data; they have been so successful (powering most large language models) that they have also been adapted for image processing tasks, such as MRI super-resolution (Forigua, Escobar, and Arbelaez 2022, in this case for 3D data).

While most of the previously described deep learning architectures are specifically trained to solve super-resolution tasks, generative models (already mentioned in the previous subsection) are also trained to mimic training datasets, and they can also be applied to super-resolution tasks by conditioning the generation process to the low-resolution version of the image to super-resolve. GANs were the first widespread paradigm of generative models, trained with a self-improving dynamic process, and have been applied to MRI super-resolution, both for 2D slices and 3D data (Guerreiro, Tomás, Garcia, & Aidos, 2023), although they have to be carefully tuned to avoid mode collapse during training (a failure mode where the generative model only learns a limited subset of the training data).

Diffusion models are another popular paradigm for generative models: they are trained to generate images by an iterative noise removal process and have emerged as a strong alternative to GANs at higher computational costs, with correspondingly high computation times and GPU memory demands. Even with this inconvenience, they have been successfully applied to MRI super-resolution, both for 2D slices (Wu, Chen, Xie, Shen, & Zeng, 2023) and 3D data (Wang et al., 2023). Because diffusion models work by removing noise, they have also been adapted to jointly super-resolve and denoise MRI images (Chung, Lee, & Ye, 2022) in 2D slices. Fortunately, diffusion models can be optimized in various ways, for example adapting the denoising process when using multimodal information to condition the generative process (Yan et al., 2024). As previously described in Section 1.1, one of the methods used to minimize the computational costs of diffusion models is to perform the diffusion process in a latent space, using Latent Diffusion Models. LDMs have also been applied to super-resolution, leveraging the latent space to produce rich conditionings (Wang et al., 2024a), and specifically to MRI super-resolution (Pinaya et al., 2022; Wang et al., 2023), because LDMs allow for the generation of high-resolution images with fine and realistic details. This is achieved by combining diffusion in the latent space with image reconstruction through a neural network, which enables capturing both high and low-frequency features of the image. Our proposal leverages LDMs for MRI super-resolution in a novel way, as described in Section 1.4. In this context, the method proposed in this work, as described in the following sections, uses LDMs for 2D super-resolution, but leverages interpolation in the latent space of the LDM in order to provide fully 3-dimensional super-resolution, but at a fraction of the memory and computational power requirements for full-fledged 3D super-resolution with LDMs.

1.3. Clinical and operational value of the proposed approach

Medical imaging, particularly Magnetic Resonance Imaging (MRI), is essential for accurate diagnostics and effective patient care. However, the practical utility of MRI devices faces significant limitations due to technological aging and operational inefficiencies. A large proportion of currently deployed MRI systems, particularly those exceeding ten years of use, experience degradation in image quality and increased risk of operational failures, mainly due to the limited availability of replacement parts and support (European Society of Radiology (ESR), 2014; ICRP PUBLICATION 154, 2023; Lozano et al., 2018). This technological obsolescence poses critical clinical challenges, as it frequently necessitates substantial financial investments for device replacement or upgrades, thereby affecting healthcare accessibility and cost-effectiveness.

In parallel, healthcare institutions globally struggle with prolonged waiting lists for MRI examinations. These delays adversely affect patient outcomes by postponing essential diagnostic and therapeutic interventions. The magnitude and variability of these waiting times have been well documented, such as significant regional variations reported in Norway, illustrating systemic inefficiencies in imaging workflows (Hofmann, Brandsaeter, & Kjelle, 2023). Moreover, during critical scenarios, such as the COVID-19 pandemic, these inefficiencies were starkly exacerbated, highlighting vulnerabilities in conventional operational models and underscoring the urgent need for more robust imaging solutions (Singer et al., 2025).

In this context, latent diffusion models emerge as a powerful computational strategy capable of reconstructing high-resolution MRI images from undersampled data acquisitions, thereby significantly enhancing both spatial and temporal resolution. This approach not only improves the diagnostic value of images produced by older MRI systems, extending their effective operational lifespan but also directly addresses clinical inefficiencies by accelerating image acquisition and processing times (Brix et al., 2024). Consequently, latent diffusion models provide an economically and clinically advantageous alternative to frequent hardware renewal, enabling healthcare providers to maintain high-quality diagnostic capabilities with existing technological assets. In addition, by enabling higher temporal resolution without increasing acquisition time, our framework contributes to faster patient throughput in clinical MRI workflows. This temporal efficiency has the potential to reduce patient waiting lists, which are a growing concern in many healthcare systems.

Furthermore, deploying advanced image reconstruction techniques, such as latent diffusion models, often requires substantial computational resources. Conventional on-premise computing infrastructures in healthcare institutions frequently struggle to keep pace with the rapidly evolving software requirements of state-of-the-art machine learning frameworks, such as PyTorch. Cloud computing infrastructure thus becomes critically relevant, offering scalable computational capabilities, predictable resource usage, and flexibility in deployment. The stability and efficiency of resource consumption provided by cloud solutions enable hospitals and imaging centers to adopt advanced computational methods without the substantial upfront investment typically associated with high-performance on-premise computing resources (Chatterjee et al., 2025; Zhou et al., 2024). Thus, the integration of latent diffusion MRI reconstruction models within cloud-based computing

environments addresses current clinical challenges comprehensively. This is reinforced by the stability analysis of resource consumption to apply the model presented in Section 3.7 It not only extends the functional utility of aging MRI technology but also optimizes healthcare operations, reduces diagnostic waiting times, and facilitates the broad adoption of advanced computational tools, ultimately improving patient care and operational sustainability.

1.4. Research contributions of this work

In this work, we introduce a novel methodology that leverages the powerful generative capabilities of latent diffusion models to enhance low-resolution images in a unique and comprehensive manner. Our approach addresses two critical challenges simultaneously: generating intermediate slices between low-resolution image slices and significantly increasing the resolution of each slice, leading to an enriched and higher-quality 3D image representation across all axes.

Most deep learning models for super-resolution are 2D, meaning that, in many cases, they can only super-resolve in two dimensions, not in all three at the same time. Of course, many super-resolution models have been adapted to operate in 3D data (using 3D convolutions), enabling their use to super-resolve in all three dimensions for MRI data; see Tables 2 to 5 from Ji et al. (2024) for a comprehensive view of 2D and 3D super-resolution models. However, models using 3D convolutions use significantly more memory than equivalent 2D models, meaning that fully-fledged 3D models are usually shallower than 2D ones (i.e. they have fewer layers and/or fewer features per layer). However, as mentioned in Section 1.4, the contributions of this work leverage the structure of the LDM's latent space to enable super-resolution in all three axes (not only within an MRI slice but also between slices) using a natively 2D super-resolution model. Thus, like other super-resolution models for MRI data, our proposal can improve the accuracy of 3D MRI data to assist in medical diagnostics and surgical planning.

The methodology described in Section 2 begins by slicing the input 3D image volume along one axis (typically the Z-axis, though the method is axis-independent and can be applied to the X and Y axes as well). Between each pair of consecutive low-resolution slices, we generate an arbitrary number of new, intermediate slices. This is achieved by using a denoising diffusion implicit model, Song, Meng, and Ermon (2020) within the latent space (DDIM), where we perform linear interpolation between the latent representations of the original slices. This step allows for smooth and coherent transitions between slices, effectively increasing the density of the image data. At the same time, we apply a resolution-enhancement process to both the original and newly generated slices. To achieve this, a U-Net architecture within the diffusion model is incorpored, which plays a crucial role in refining the resolution of the images. The U-Net's encoder-decoder structure is particularly effective in capturing multi-scale features and spatial information, enabling the model to produce higher-resolution outputs while preserving important details from the low-resolution inputs. This approach works directly from low-resolution images, making it especially useful for applications where high-quality data is not readily available. What sets our methodology apart is its dual advantage: the simultaneous generation of intermediate slices and the enhancement of their resolution. This is a key contribution because few existing approaches manage to tackle both tasks within a unified framework. Most competing methods focus either on interpolation to generate new slices or on super-resolution tasks, but rarely both at the same time. By harnessing the strengths of generative AI, the U-Net architecture, and the latent diffusion process, our approach fills a gap in the field, offering a robust and scalable solution that is capable of improving the quality of image datasets across all dimensions.

The use of linear interpolation (Michelis & Becker, 2021) in the latent space is a particularly efficient and stable solution, avoiding the complexity and potential artifacts. This allows our model to perform well even with very low-resolution input images, ensuring that the generated and enhanced slices remain coherent and free of distortions. Furthermore, by utilizing the latent space in a diffusion model, our approach benefits from the structured, continuous nature of this space, which has been trained to capture meaningful variations in the data.

Overall, our methodology stands out for its unique ability to leverage generative AI in a way that few other approaches have, addressing both slice generation and resolution enhancement simultaneously within a single cohesive framework. This dual capability is what differentiates our work, as it enables the creation of new slices between existing low-resolution ones while also improving the resolution of both the original and generated slices. The integration of latent diffusion models with powerful architectures like the U-Net further amplifies the effectiveness of this approach, as the U-Net's multi-scale feature extraction and spatial preservation capabilities ensure that the resolution enhancement is both accurate and detailed. By working directly from low-resolution images, our method overcomes the challenges associated with sparse or limited data, offering a robust and scalable solution for improving the quality of image datasets in a computationally efficient manner. In Section 3, we demonstrate the effectiveness of our approach by testing it across a variety of MRI datasets. Through both quantitative evaluations using established metrics from the literature and qualitative assessments, we show that our method significantly improves the resolution of slices from 64x64 (a space with very limited information) to 256x256, effectively quadrupling the resolution. Additionally, we generate an arbitrary number of intermediate slices between each pair of consecutive slices, further enriching the data. This makes the methodology particularly valuable for applications requiring high-resolution 3D image reconstructions, such as medical imaging. Furthermore, it leaves an interesting framework open for future research in other fields, given the great potential it has demonstrated in this domain.

2. Methodology

The field of medicine, particularly medical imaging, requires advanced computer vision techniques for super-resolution and image generation to enhance detail and improve diagnostic accuracy. Traditional methods like interpolation in pixel space often compromise information integrity, adding noise that makes no biological sense when trying to approximate an image, such as, for example, smooth gradations in brightness levels between brain tissue and dark areas with cerebrospinal fluid. On the other hand, while autoencoders have proven useful for generating images in the latent space, they are limited because they require having two high-resolution images available beforehand. Our approach proposes using a latent diffusion model, where random noise in the low-resolution latent space, conditioned on the original low-resolution images coming from the slices through one of the axes in a 3D volume, is used to generate a random number of biologically coherent high-resolution images. This is achieved by interpolating the output of the conditioned random noise and using the decoder to reconstruct the high-resolution images.

2.1. Latent difussion models framework

To understand our approach, it is vital to have a general idea of the latent diffusion model framework. A latent diffusion model encodes an input image x into a latent space using an encoder E, uses a U-Net model to perform a Gaussian diffusion process on the image in the latent space, and then decodes the result of the diffusion process to reconstruct the image (Rombach et al., 2022) as the Fig. 1 shows.

Let $x \in \mathbb{R}^{H\times W\times C}$ denote the input image with height $H \in \mathbb{N}$, width $W \in \mathbb{N}$, and $C \in \mathbb{N}$ channels. The encoder E (Bank et al., 2023) maps the input image x to a latent space z:

$$z = E(x) \tag{1}$$

The diffusion process (Croitoru, Hondru, Ionescu, & Shah, 2023) is performed using a U-Net architecture. The latent space representation z

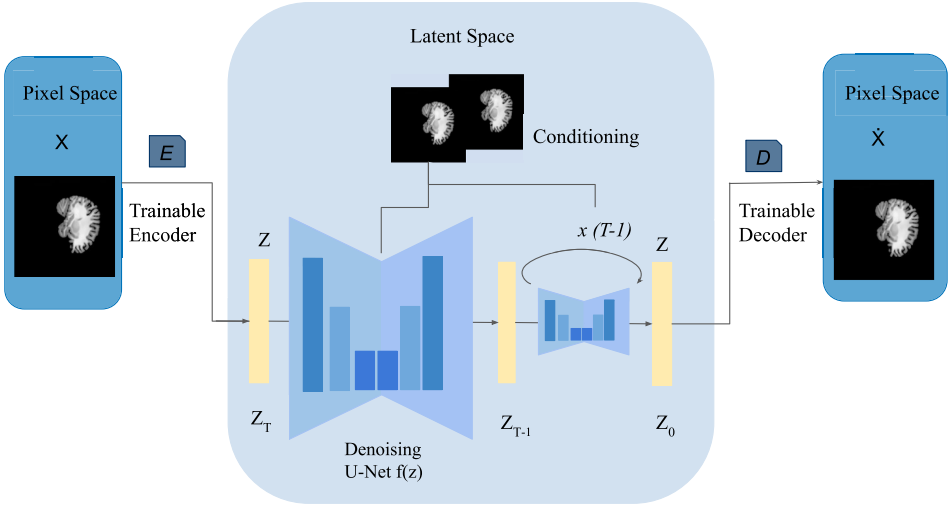

Fig. 1. Training process of a latent diffusion model. An image x is encoded into the latent space by a trainable encoder E. The model trains the U-Net ($f(z)$) by first generating noisy versions of the image and then learning to remove that noise by focusing on the details. Finally, the image is generated by running the output of the diffusion process through a decoder D, translating from the latent space to an image.

is passed through the U-Net, which consists of an encoder-decoder structure with skip connections. At each level of the encoder, the resolution of the latent space is typically downsampled, while the decoder upsamples the latent representation to the original resolution. The downsampled original image is used to condition the generation process, providing valuable contextual information so that the model can better understand the overall structure and content of the original image. This contextual guidance helps the U-Net architecture to generate high-resolution details that are consistent with the content of the original image, resulting in reconstructions that maintain fidelity to the input while enhancing visual quality. Specifically, let $f(z)$ represent the U-Net architecture, the diffused latent representation $\tilde{z}$ is obtained as:

$$\tilde{z} = f(z) \tag{2}$$

Finally, the diffused latent representation y is decoded back to the image space:

$$\hat{x} = D(\tilde{z}) \tag{3}$$

where D is the decoder (Bank et al., 2023) function.

2.2. Latent space

The latent space plays a fundamental role in our model, which is why the following section is dedicated to explaining a series of additional details to clarify the methodology in the upcoming sections and to understand the value that a latent space provides to a methodology for generating comprehensive images. The diffusion process involves a sequence of transformations, adding and removing noise to generate a data sample from an input distribution, possibly conditioned to data from another distribution. This data sample represents a point in a latent space, which is decoded by a decoder network to reconstruct the higher-resolution data. The benefit of using a latent space (Rafailov, Yu, Rajeswaran, & Finn, 2021) in LDM is that it allows for more efficient computation of the diffusion process than in the image space. By modeling the diffusion process in the low-dimensional latent space, the computation can be done much faster than in the high-dimensional image space, which is computationally expensive. This is because the latent space is often of much lower dimensionality than the image space, which reduces the computational cost.

Another advantage of using a latent space is that it can provide a natural regularization mechanism. By imposing a prior distribution over the latent variables, we can encourage the model to generate samples that are more plausible and diverse. For example, we can impose a prior that favors latent variables with small magnitudes, which encourages the model to generate samples that are smoother and more regularized. This can be useful for generating high-quality, realistic images. A latent space is regularized (Pati & Lerch, 2021) when constraints or penalties are applied to the latent representations during the training of a machine learning model. These constraints are introduced to encourage desirable properties in the latent space, such as smoothness, continuity, or sparsity, which can lead to better generalization and interpretability of the model. Regularization techniques help prevent overfitting and improve the robustness of the learned representations.

Finally, the use of a latent space can also provide a more succinct representation of the data. By analyzing the learned latent variables, we can gain insights into the underlying structure of the data and the factors that drive its variation. This can be useful for applications such as data exploration or scientific discovery, where we may want to understand the relationships between different variables in the data.

2.3. Sampling with denoising diffusion implicit models

The way sampling is performed is how the model operates, so we dedicate the next section to explaining an efficient and fast sampling method, as well as the crucial role of the U-Net in this process. The LDM Denoising Diffusion Implicit Models, whose acronym is DDIM (Song et al., 2020), will be used to perform inference in the experiment. The process of sampling from an LDM using DDIM involves a series of steps designed to refine and generate high-quality images from a latent space representation. This section presents the mathematical underpinnings and practical execution of sampling using DDIM within the framework of LDMs.

DDIMs offer a non-Markovian variant of the traditional denoising diffusion probabilistic model, enabling faster and more controlled sampling processes. By employing a deterministic trajectory through the diffusion process, DDIM allows for efficient backward sampling from the noise distribution, thereby facilitating the generation of coherent and high-quality images. Consider a latent space representation $z \in \mathbb{R}^{h \times w \times c}$, where $h, w, c \in \mathbb{N}$ denote the height, width, and number of channels in the latent space, respectively. The DDIM sampling process reverses the diffusion process, transforming a sample of noise into a coherent image representation in the latent space. The sampling process from an LDM using DDIM involves a carefully structured sequence of operations aimed at reconstructing a coherent and high-quality image from an initial noise distribution. Here, we provide a more detailed description of each step involved in the sampling process:

1. **Initialize from Noise:** The process begins by initializing the latent representation z_T by sampling from a standard Gaussian noise distribution, $\mathcal{N}(0, I)$. This step generates a purely random noise vector that serves as the starting point for the reverse diffusion process. Mathematically, this can be represented as:

$$z_T \sim \mathcal{N}(0, I), \tag{4}$$

where z_T is the latent representation at the final timestep T.

2. **Iterative Denoising:** The core of the DDIM sampling process is an iterative denoising procedure that gradually transforms the initial noise vector z_T into a coherent image representation in the latent space. For each timestep t from T down to 1, the latent representation z_t is refined using a deterministic update rule that incorporates the learned denoising function $\epsilon_\theta(z_t, t)$. This learned function is crucial as it predicts the noise that was added at timestep t during the forward diffusion process, allowing for its removal or "denoising" in the reverse process. The deterministic update rule can be formulated as follows:

$$\begin{aligned} z_{t-1} &= f(z_t, \epsilon_\theta(z_t, t)) \\ &= \sqrt{\alpha_{t-1}} \frac{z_t - \sqrt{1-\alpha_t}\epsilon_\theta(z_t, t)}{\sqrt{\alpha_t}} + \sqrt{1-\alpha_{t-1}-\sigma_t^2} \cdot \epsilon_\theta(x_t, t) + \sigma_t \epsilon_t, \end{aligned} \tag{5}$$

where ϵ_t is random noise, $\alpha_t = 1 - \beta_t$, β_t is a pre-defined variance schedule of the diffusion process and

$$\sigma_t = \eta \sqrt{\frac{1-\alpha_{t-1}}{1-\alpha_t}} \sqrt{1 - \frac{\alpha_t}{\alpha_{t-1}}}, \eta > 0.$$

η is used to calculate σ. $\eta = 0$ makes the sampling process deterministic. Each iteration effectively reverses the diffusion step, reconstructing finer details and reducing noise in the latent representation.

3. **Denoising with U-Net:** The core of the reverse diffusion process involves denoising the latent variable at each timestep using a neural network, typically a U-Net architecture. The U-Net plays a crucial role in predicting the noise component $\epsilon_\theta(z_t, t)$ present in the latent representation z_t at each timestep t. Its encoder-decoder structure, with skip connections, allows the model to capture both global and fine-grained features, facilitating the precise removal of noise while preserving important image details. This iterative denoising, guided by the U-Net, is critical for transforming random noise into a coherent image, as the model refines z_t over multiple steps to reconstruct the final high-quality image representation. The effectiveness of DDIM's sampling process largely depends on the ability of the U-Net to perform accurate noise predictions at each step, ensuring efficient and high-quality image generation.

A DDIM tends to be faster (Kong & Ping, 2021) for sampling due to its simplified architecture and direct conditioning on low-resolution images. This streamlined approach reduces computational overhead, enabling quicker generation of samples without sacrificing quality.

It is important to highlight the role of two fundamental components in the sampling process: the parameter η and the timesteps T. The parameter η encompasses the learnable weights of the denoising network, specifically the U-Net, which is trained to accurately estimate the noise $\epsilon_\theta(z_t, t)$ introduced at each timestep. The precision of these noise predictions is essential, as it directly impacts the quality and coherence of the final reconstruction. On the other hand, the timesteps T define the number of iterations through which the reverse diffusion process unfolds. A larger T allows for a finer and more gradual denoising trajectory, potentially leading to higher-quality outputs at the expense of increased computational time. Conversely, reducing T accelerates the sampling but may degrade the fidelity of the reconstruction if insufficient steps are allocated to effectively remove the noise. Therefore, an appropriate balance between η and T is crucial for achieving both efficiency and quality in the image generation process.

For clarity, Algorithm 1 summarizes the full DDIM sampling procedure described in this subsection, providing a step-by-step overview of the iterative denoising process.

In addition to DDIM, we also mention the Pseudo Linear Multistep (PLMS) sampling strategy (Farooq, Abaid, Ullah, & Corcoran, 2024), which further accelerates the sampling process while improving stability. PLMS leverages previous noise estimates across multiple steps to refine the current update, effectively approximating higher-order solvers for the reverse diffusion equation.

Algorithm 1 DDIM Sampling Procedure.

Input: Number of timesteps T, noise scale η, trained U-Net ϵ_θ, conditional input c
Initialize: $z_T \sim \mathcal{N}(0, I)$
for $t = T$ to 1 **do**
  Predict noise: $\hat{\epsilon} = \epsilon_\theta(z_t, t, c)$
  Compute α_t, α_{t-1} from the noise schedule
  Compute $\sigma_t = \eta \sqrt{\frac{1-\alpha_{t-1}}{1-\alpha_t}} \sqrt{1 - \frac{\alpha_t}{\alpha_{t-1}}}$
  if $\eta > 0$ **then**
    Sample noise $\epsilon_t \sim \mathcal{N}(0, I)$
  else
    $\epsilon_t = 0$
  end if
  Update z_{t-1}:

$$z_{t-1} = \sqrt{\alpha_{t-1}} \frac{z_t - \sqrt{1-\alpha_t}\hat{\epsilon}}{\sqrt{\alpha_t}} + \sqrt{1-\alpha_{t-1}-\sigma_t^2} \cdot \hat{\epsilon} + \sigma_t \epsilon_t$$

end for
Output: Denoised latent representation z_0

2.4. Interpolation using latent diffusion models for image generation

In this section, the proposed approach to leverage the generative and resolution-enhancing capabilities of latent diffusion models is outlined. The methodology revolves around the manipulation of image slices obtained along the Z-axis (although it can be replicated for the other two axes because the methodology is axis-independent, relying on the same principles of image slice manipulation; by treating the X-axis and Y-axis similarly, the approach remains consistent, allowing for the same operations to be applied regardless of the axis of slicing.), which is typically characterized by lower resolutions. The proposed generative process of slices between two consecutive slices used in our experiments and shown in Fig. 2 is as follows:

Let n be the multiplicative value of the resolution and x_A, x_B two consecutive low-resolution slices along the Z-axis:

- Initially, random noise images $\tilde{l}_A$, $\tilde{l}_B \in \mathbb{R}^{64} \times \mathbb{R}^{64}$ are generated, which serve as anchor points in the latent space; $\tilde{l}_A, \tilde{l}_B \sim \mathcal{N}(0, I)$.
- The latent vectors $\tilde{l}_A$ and $\tilde{l}_B$ are processed through the denoising diffusion implicit model (Section 2.3), conditioned on the original low-resolution images x_A and x_B. This diffusion process yields l_A and l_B as final results, also in the latent space, i.e., l_A, $l_B \in \mathbb{R}^{64} \times \mathbb{R}^{64}$.
- Within this latent space, a line segment $\overline{AB}$ between l_A and l_B is interpolated. By extracting n equidistant points along this line, l_j, $j = 1, \ldots, n$, the trajectory for generating intermediate images is effectively defined:

$$l_j = l_A + \frac{j}{n+1}(l_B - l_A) \tag{6}$$

- Subsequently, the intermediate latent space points l_j, $j = 1, \ldots, n$ along with l_A and l_B are decoded using the decoder network D, resulting in the generation of n high-resolution images that seamlessly interpolate between the original two low-resolution slices x_A, x_B. A

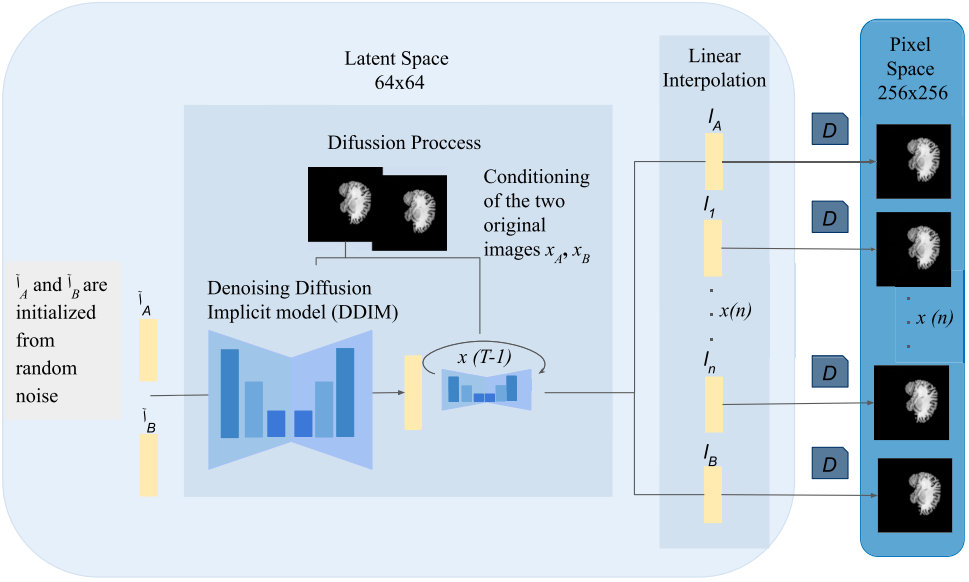

Fig. 2. Proposed neural architecture. In order to super-resolve multiple output images intercalated between two input images across the Z dimension, the denoising diffusion implicit model is initialized with random noise and is conditioned to the low-resolution images, and interpolation is performed in the latent space between their respective latent representations just before decoding them into images.

total of $n+2$ images are obtained, encompassing the original two slices A and B, and the n generated ones, all in higher resolution:

$$\hat{x}_A = D(l_A) \tag{7}$$

$$\hat{x}_j = D(l_j),\ j = 1, \ldots, n \tag{8}$$

$$\hat{x}_B = D(l_B) \tag{9}$$

where $j = 1, \ldots, n$; $\hat{x}_j, \hat{x}_A, \hat{x}_B \in \mathbb{R}^{256} \times \mathbb{R}^{256}$; and D is the decoder. It is highlighted that $\hat{x}_A$ and $\hat{x}_B$ would correspond to the outputs by the model of the original images.

It is important to note that this process deviates from the standard operation of a latent diffusion model, as illustrated in Fig. 1. A denoising diffusion implicit model is used for sampling, conditioned by the original low-resolution images, but the LDM-model is trained according to the architecture shown in Fig. 1. Here, the latent space has the same resolution as the lower-resolution input pixel space since we start from low-resolution slices. The model ultimately manages to increase the resolution across all axes, as it also enhances the resolution of each slice, providing a doubled value.

Linear interpolation in the latent space offers simplicity and efficiency while still producing smooth transitions. Specifically, because the latent space in diffusion models is often well-structured and continuous, linear interpolation can effectively blend high-level features between two points without needing complex or nonlinear transformations. This straightforwardness ensures that the interpolated results remain coherent and realistic, as the latent space has been trained to capture meaningful variations. Additionally, linear interpolation is computationally inexpensive compared to more complex methods, making it practical for generating intermediate steps. A key advantage of linear interpolation compared to more complex interpolation methods (e.g., spline or polynomial interpolation) in the latent space is its simplicity and stability. Linear interpolation ensures a predictable, smooth transition between two points without introducing unwanted distortions or artifacts, which can sometimes occur with higher-order methods due to overfitting or oscillations (Gallagher, 2005; Herman, Rowland, & Yau, 1979).

The interpolation in the latent space effectively reduces the slice thickness in millimeters by generating n intermediate slices between two consecutive low-resolution slices. Given an original slice spacing of s_{orig} mm, the new slice thickness s_{new} is:

$$s_{\text{new}} = \frac{s_{\text{orig}}}{n+1} \tag{10}$$

For instance, if $s_{\text{orig}} = 5$ mm and $n = 4$, then:

$$s_{\text{new}} = \frac{5}{4+1} = 1 \text{ mm} \tag{11}$$

Thus, the anisotropic resolution is significantly enhanced, bringing it closer to the in-plane resolution and improving volumetric coherence in medical imaging.

The choice of resolutions 64×64 and 256×256 is especially useful because 64×64 is such a low resolution that it allows us to quickly evaluate whether the model can generate coherent and acceptable quality images, even under challenging conditions. If the model performs well at such low resolutions, it suggests that it would likely excel with higher-resolution images.

As a conclusion of the previous three subsections, Fig. 3 summarizes the complete workflow of the proposed method. Starting from low-resolution images, the process first applies DDIM sampling in the latent space conditioned on the low-resolution data. Then, a latent representation is obtained, where linear interpolation is performed to increase the anisotropic resolution. Finally, the latent space is decoded to produce the super-resolved high-resolution image. This workflow allows efficient reconstruction while maintaining anatomical consistency and reducing computational complexity.

2.5. Baseline super-resolution with intermediate slice generation

In magnetic resonance imaging (MRI), the quest for enhancing resolution in multidimensional image data, particularly across the X, Y, and Z axes, necessitates the selection of an appropriate baseline model. A detailed justification for the baseline selection is imperative. It is essential to underscore that the chosen method does not mandate a fixed zoom level but rather facilitates a variable and arbitrarily large zoom range. This flexibility is crucial in the context of MRI, where the requirement for detailed examination across different scales is paramount, but that limits the competitor to also accepting this arbitrary choice of super-resolution along one of the axes and additionally increasing the model's slice resolution across the same axis chosen.

The baseline model under consideration is used to enhance resolution along the Z-axis by generating intermediate images between consecutive slices, denoted as $\hat{x}_A, \hat{x}_B$ for slices in the X and Y dimensions after increasing their resolution by four with the LDM-model (although the same idea can be applied to increase the resolution in the other

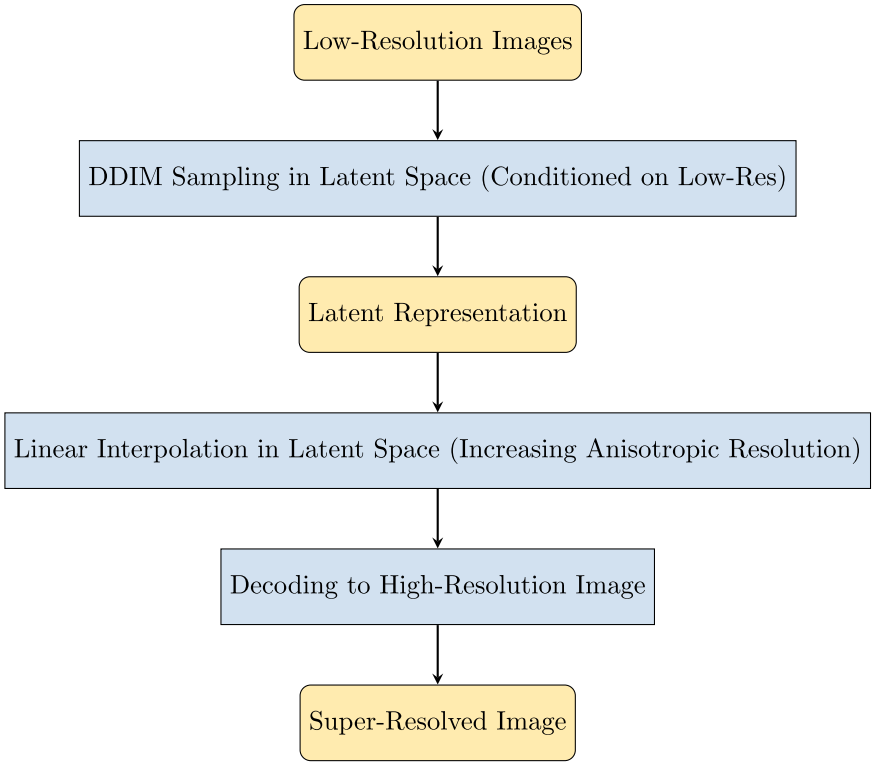

Fig. 3. Workflow of the proposed method for super-resolution with latent space interpolation.

axes). This baseline model can be formulated as follows:

$$I'_j = \frac{j \cdot \hat{x}_A + (j-1) \cdot \hat{x}_B}{n}, \quad j = 1, \ldots, n \tag{12}$$

where n is the multiplicative value of the resolution, and I'_j is the interpolated image. This model embodies a systematic approach to interpolate between adjacent slices, using information from both current $\hat{x}_A$ and subsequent $\hat{x}_B$ images. Notably, this baseline model results in smooth gradations between bright and dark areas in the interpolated slices wherever there are significant differences between pixel values in $\hat{x}_A$ and $\hat{x}_B$, but these color gradations are not realistic, since most features in the MRI slices correspond to body structures that should be imaged as presenting predominantly sharp transitions in pixel values across different slices. Our proposal minimizes this issue. In other words, we will compare generation in the image space with generation in the pixel space.

2.6. Blurriness removal

The images generated by the model may sometimes be somewhat blurry. Therefore, we propose to apply an EDSR (Zhao et al., 2019) network, which is much lighter to train, to remove the blurriness. Let $\hat{x}$ denote a blurred-resolution image generated with an LDM, and x' denote its corresponding high-resolution image. The goal is to use an EDSR to remove blurriness in the predicted image and produce an output image x_{HC} that is as sharp and detailed as possible.

To achieve this, we can formulate the problem as an image restoration task, where we aim to learn a mapping function $F(x)$ that maps the blurred-resolution image to its corresponding high-resolution image. This can be achieved by training an EDSR on a dataset of paired blurry and sharp images, using a loss function that measures the difference between the predicted and ground-truth high-resolution images. The EDSR can be represented as follows:

$$x_{HC} = F(\hat{x}; \theta), \tag{13}$$

where θ represents the learnable parameters of the EDSR, and x_{HC} is the predicted high-contrast image. The goal is to find the optimal values of θ that minimize the difference between the predicted and ground-truth high-resolution images, which can be formulated as a mean squared error (MSE) loss:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} ||x'^{(i)} - x_{\mathrm{HC}}^{(i)}||^2, \tag{14}$$

where N is the number of training samples, and $x_{HC}(i)$ and $x'(i)$ denote the predicted and ground-truth high-resolution images for the i-th training sample, respectively.

During training, the EDSR is optimized to minimize the loss function $L(\theta)$ using backpropagation and stochastic gradient descent (SGD) or one of its variants. The process of using an EDSR to remove blurriness from images generated with an LDM involves training the EDSR on a dataset of paired blurred- and high-resolution images, using a loss function that measures the difference between the predicted and ground-truth high-resolution images. This approach effectively reduces blurriness in the images generated by the LDM, enhancing structural clarity and preserving fine anatomical details. By applying EDSR as a post-processing step, we can refine the super-resolved images while maintaining the smooth transitions achieved by the latent diffusion process. An example of this improvement can be seen in Fig. 4, which illustrates a comparison between an image before and after applying EDSR, demonstrating the reduction of blurriness and the enhancement of sharpness in the reconstructed high-resolution output.

2.7. Competitors

Due to the dual nature of our proposal, which combines slice-wise super-resolution using LDMs and latent space generation, establishing direct competitors is particularly challenging. Existing 3D LDM (Kim & Park, 2024; Nam et al., 2022) approaches are limited by the scarcity of public implementations and datasets, and they differ conceptually from our method, which focuses on sequential slice enhancement rather than full volumetric generation. Therefore, beyond our baseline, we introduce additional competitors inspired by state-of-the-art techniques

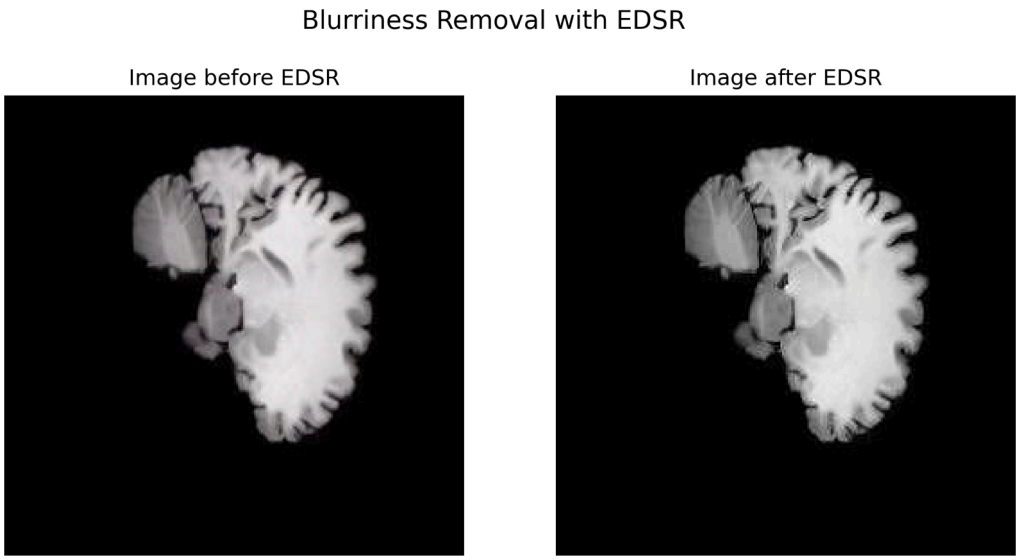

Fig. 4. Comparison of an image of the OASIS dataset presented in Section 3.2 before and after applying EDSR. The left side shows the original LDM-generated image, which exhibits some blurriness, while the right side presents the same image after EDSR refinement, demonstrating improved sharpness and enhanced structural details.

commonly applied in medical image super-resolution. Specifically, we incorporate SRGAN combined with linear interpolation in the pixel domain and WDSR with linear interpolation in the pixel domain, as motivated by recent works demonstrating the effectiveness of GAN-based and residual-based architectures for enhancing medical images (Guerreiro et al., 2023; Madhav, Nandhika, & Devi, 2024; Tan, Zhu, & Lio', 2020; Yoshimura et al., 2023). Additionally, we include bicubic upsampling with latent interpolation via an autoencoder, following the methodology proposed by Sander, de Vos, and Išgum (2022), which specifically addresses the challenge of improving anisotropic MRI resolution through semantically smooth interpolation in the latent space-an objective closely aligned with our own goal of enhancing through-plane resolution while preserving anatomical consistency. These alternatives provide meaningful and complementary competitors for assessing the impact of our LDM-based framework within the context of medical image super-resolution.

3. Experiments and results

In our experiments, we address the task of generating $n \in \mathbb{N}$ consecutive intermediate slices between two existing slices of a 3D brain volume along the Z-axis, while simultaneously enhancing the in-plane resolution of each slice by a factor of 4. This combination of spatial super-resolution and through-plane interpolation results in a final volume with significantly increased resolution, both within slices and across slices, achieving overall factors that are uncommon in the current literature. To enable proper evaluation of this generative task, we construct ground truth volumes following the procedure described in Section 3.1. For the quantitative experiments presented in Section 3.3, we focus on the case of $n = 1$, generating a single intermediate slice between two real slices to enable direct comparison with the ground truth. For the qualitative analysis in Section 3.5, we explore higher values of n, such as $n = 4$, $n = 6$, and $n = 10$, which correspond to unusual and arbitrary anisotropic resolution increases of factors like 5, 7, and 11, providing insight into the model's performance under more challenging and less conventional reconstruction scenarios. In addition to the qualitative study, we incorporate an expert medical evaluation conducted by four radiologists in Section 3.5.5. Furthermore, a final computational efficiency analysis is performed in Section 3.7, where we assess the scalability and feasibility of our approach under realistic and different conditions.

The approach revolves around training a latent space that, when appropriately regularized, effectively captures the semantic features inherent in the images. By using the latent space to interpolate between images, the methodology offers a powerful means to generate a sequence of intermediate slices within the 3D brain image and increase the resolution of the slices in the XY plane, which adds double value by increasing the resolution in all axes. The process of generating intermediate images between two slices along the Z-axis enhances the resolution of a 3D image by increasing the density of data in the Z-dimension. By introducing additional slices between existing ones, fine details that might otherwise be lost due to low sampling density in the Z-dimension are captured. This process essentially interpolates information between adjacent slices, improving the three-dimensional representation by providing a more detailed and accurate visualization of objects and structures in the 3D image. Although slices are taken across the Z-axis, it is worth noting that this same method can be applied to the other two axes. Due to the similarity between brain structures, increasing the resolution in the other two axes produces remarkably similar results, and to demonstrate this, we consider datasets in which MRI scans have been taken from different angles.

3.1. Experiment details

Since the models generate intermediate images, we lack a ground truth, there is no objective way to evaluate the quality of image generation. Therefore, the following procedure through which we will base our quantitative results in Section 3.3 is run in order to conduct the quantitative evaluation of the experiment and obtain an objective ground truth:

- The resolution of the slices is first decreased by dividing the total resolution by four.
- Next, 10 % of the slices are randomly removed from each patient's dataset. These eliminated slices will serve as ground truth to evaluate the performance of our experiment.
- Subsequently, the missing 10 % of the slices are generated by the process described in Section 2.4 taking $n = 1$ and the resolution of all remaining slices is upscaled.
- Finally, the generated slices are compared with the ground truth for validation, as well as with the Baseline model defined in Section 2.5, and the competitor methods described in Section 2.7.

For the qualitative study in Section 3.5, the resolution of the slice image will be initially increased by a factor of 4. We generate an arbitrary number of intermediate slices with different values of n, such as 6, to enhance the resolution along the Z-axis by a factor of 7, or 10 slices to increase the resolution by a factor of eleven. These values, uncommon prime numbers in the literature, demonstrate the flexibility of our method. Simultaneously, we enhance the resolution of each individual slice by a factor of four, transforming images from $\mathbb{R}^{64\times64}$ to $\mathbb{R}^{256\times256}$. A detailed visual analysis is conducted to examine the images generated with different values of n and across multiple patient datasets, comparing the results with those produced by the Baseline method and the competitor methods. This allows us to evaluate the quality of the slices generated. In addition to the qualitative study, a medical evaluation was conducted through a survey completed by four expert radiologists. This survey aimed to provide an objective assessment of the generated images based on three key criteria: realism, consistency between slices, and anatomical correctness. By incorporating this external validation, we ensure that our proposed method is not only quantitatively robust but also meets the perceptual and diagnostic expectations of medical professionals. The results of this evaluation are detailed in Section 3.5.5.

To make the experiments as transparent as possible, the details of the training are outlined below. As for the hardware specifications, we have run our experiments in 32GB NVidia A100 GPUs managed by the Supercomputing and Bioinnovation Center of the University of Málaga.

Next, we detail the architectures and training configurations of the models. The training details for datasets are the same. Initially, the autoencoder VQ-VAE (Razavi, Van den Oord, & Vinyals, 2019) is trained with a base learning rate of 4.5×10^{-6}. The data-dependent configuration includes three channels in the latent space, a resolution of 256×256, and 128 channels in intermediate layers. 2 residual blocks are used, and neither attention mechanisms nor dropout are employed. For the loss function, LPIPS (Wu et al., 2023) with a discriminator is utilized. The discriminator is not conditioned and accepts three input channels, with a weight of 0.75 for discriminator loss and 1.0 for codebook loss. The training process of the LDM spanned over 1000 timesteps, with a linear interpolation scheme starting at 0.0015 and ending at 0.0155. Results were logged every 100 timesteps for monitoring. The loss function employed was reconstruction loss.

The model architecture consisted of three key stages: the initial processing stage (based on the VQ-VAE model trained before), the conditional stage, and the U-Net module, operating on an image size of 64×64 with three channels, featured attention mechanisms at resolutions 16×16 and 8×8, and comprised two residual blocks with channel multipliers. The conditional stage, which was not trainable, played a crucial role in guiding the generation process. It was configured to operate as an identity function.

During the training phase, a batch size of 16 was used for both the autoencoder and the U-Net, ensuring efficient processing while optimizing computational resources. To enhance the robustness and generalization of the model, a comprehensive data augmentation strategy was applied, leveraging the BSRGAN framework (Zhang, Liang, Van Gool, & Timofte, 2021a), which is specifically designed for super-resolution tasks. Within this framework, we employed the BSRGAN Light degradation model, a simplified variant tailored to introduce realistic yet mild degradations commonly observed in MRI acquisitions. This includes subtle blurring, low-level noise, and moderate compression artifacts, helping the model learn to reconstruct high-quality images from degraded inputs. Additionally, a downscaling factor of 4 was applied, along with random cropping, where crop sizes varied between 50 % and 100 % of the original image dimensions. This combination of controlled degradation and diverse spatial augmentations significantly enriched the training data, enabling the model to adapt to a wide range of image conditions while improving generalization to real-world MRI scans.

In our experiments, the sampling process is carried out using DDIM with $\eta = 1$ (resulting in a stochastic sampling trajectory) and a total of $T = 200$ timesteps. This configuration offers a favorable balance between computational efficiency and reconstruction quality, as the stochasticity introduced by $\eta = 1$ allows for greater diversity in the generated samples while maintaining stable and high-fidelity reconstructions over 200 denoising steps. The choice of 200 timesteps ensures that the noise removal process is sufficiently gradual to preserve fine anatomical structures without incurring excessive computational costs. This setting is particularly advantageous in medical imaging scenarios where both variability and detail preservation are important. While DDIM was selected for its ability to balance speed and quality, we also explore the potential of PLMS sampling, which further stabilizes and accelerates the denoising process by leveraging previous noise estimates, especially when working with fewer timesteps. A comprehensive quantitative comparison of these different sampling strategies and hyperparameter configurations can be found in Section 3.6.

The EDSR model used to remove blurriness presented in this paper comprises a convolutional neural network architecture with a depth of 16 residual blocks (He, Zhang, Ren, & Sun, 2016) with Rectified Linear Unit (ReLU) activation functions, trained for 25 epochs using the Adam optimizer with a learning rate schedule. Each epoch involves 1000 steps, and the training process utilizes the mean absolute error as the loss function. The model is subjected to random data augmentation techniques, including random cropping and flipping, to enhance its robustness and generalization capabilities. The net starts with an input layer adaptable to varying image sizes and color channels; the model proceeds with feature extraction through convolutional layers, followed by a cascade of 16 residual blocks. These blocks, characterized by dual convolutional layers with ReLU activations, facilitate feature refinement while residual connections ensure smooth gradient flow during training. Subsequent upsampling operations amplify spatial resolution, culminating in a final convolutional layer to generate high-resolution outputs. Employing normalization and denormalization routines ensures stable training and faithful reconstruction of pixel values.

3.2. Datasets

In the experiments, the analysis was conducted using four different datasets. Specifically, the OASIS dataset (Marcus et al., 2007), the ATLAS dataset (Liew et al., 2018), the OpenNeuro dataset (Markiewicz et al., 2021) and the BraTS dataset (Baid et al., 2021; Bakas et al., 2017; Menze et al., 2014) were selected to define the datasets used for cross-validation, while 3 patients from each dataset were used as the test sets. The rationale behind selecting these datasets lies in their complementary characteristics, which provide a diverse representation of brain images, ensuring a wide range of anatomical variations and pathological conditions. OASIS is focused on aging and Alzheimer's disease, offering valuable data on neurodegenerative conditions and includes T1-weighted images. ATLAS, in contrast, includes data from various neurological disorders with T1-weighted acquisitions, adding diversity to the pathological spectrum. OpenNeuro contributes a broad range of neuroimaging data, including healthy subjects and those with different clinical conditions, and provides FLAIR images, enriching the overall variability. Additionally, to further increase the heterogeneity of the analysis, the BraTS dataset (Menze et al., 2014) was included, which offers T1, T2, and FLAIR images of patients with brain tumors, introducing more diversity in both pathology and imaging modalities.

Furthermore, since the MRI scans come from different datasets, they are acquired from slightly different angles and orientations, which means the scans are not perfectly aligned across datasets. To address this variability and test the model's robustness across different axes, resolution enhancement was applied in the Z-axis, ensuring that the analysis could capture features not only in the XY plane but also in the depth dimension. This allowed us to verify the performance of the model in all spatial directions.

For uniformity in the analysis, all datasets were preprocessed to ensure the same image dimensions, zero padding was applied to achieve

a final size of 256×256 pixels, ensuring consistent training conditions across all datasets. The number of images is:

- For the OASIS dataset used for cross-validation: $160 \times 17 = 2,720$ images.
- For the ATLAS dataset used for cross-validation: $189 \times 31 = 5,859$ images.
- For the OpenNeuro dataset used for cross-validation: $160 \times 85 = 13,600$ images.
- For the BraTS T1 dataset used for cross-validation: $155 \times 27 = 4,185$ images.
- For the BraTS T2 dataset used for cross-validation: $155 \times 27 = 4,185$ images.
- For the BraTS FLAIR dataset used for cross-validation: $155 \times 27 = 4,185$ images.
- For the OASIS test dataset: $160 \times 3 = 480$ images.
- For the ATLAS test dataset: $189 \times 3 = 567$ images.
- For the OpenNeuro test dataset: $160 \times 3 = 480$ images.
- For the BraTS T1 test dataset: $155 \times 3 = 465$ images.
- For the BraTS T2 test dataset: $155 \times 3 = 465$ images.
- For the BraTS FLAIR test dataset: $155 \times 3 = 465$ images.

The utilization of multiple datasets enriches the analysis by providing a broader spectrum of brain images for training and evaluation. This approach, coupled with cross-validation, enhances the generalizability and robustness of our findings, allowing us to draw meaningful insights into the performance and applicability of our proposed methodology across different cohorts and imaging modalities. In order to ensure the robustness of our model and the reliability of our results, 5-fold cross-validation was employed. This involved partitioning the data into training and validation sets, iteratively training the Latent Diffusion Model on different subsets of the data, and evaluating its performance across multiple folds.

3.3. Quantitative study

A quantitative study will be conducted to rigorously evaluate the performance of the generative capacity of the LDM-model versus the generative capacity of the baseline model. This study aims to assess the model's efficacy in enhancing image resolution along the Z-axis and increasing the resolution of the slices to produce and increasing in the resolution across all the axes. As it is mentioned in Section 3.1, in order to thoroughly assess the generative capability of our algorithm and facilitate meaningful comparisons with real-world images, the following procedure was executed. Within this process, 10 % of the images from the test patient dataset were randomly sampled and systematically removed. Subsequently, using the Latent Diffusion Model, these missing images were reconstructed while simultaneously multiplying the original slice image resolution by four. This approach enables quantification of the algorithm's ability to generate high-quality images from incomplete data and validates its robustness and generalization in scenarios involving missing data and in this way we can evaluate the two great results that we achieved at the same time with the proposed model, increasing the resolution of the slices and also through the axis from which we took the slices, achieving an increase in resolution in all axes.

To evaluate the performance of the proposed baseline model in enhancing image resolution in MRI datasets, a combination of evaluation metrics is employed, including the Structural Similarity Index, acronymed as SSIM, Ahn et al. (2020), Peak Signal-to-Noise Ratio, abbreviated as PSNR, Venu (2023), and Learned Perceptual Image Patch Similarity, acronymed as LPIPS, Wu et al. (2023). SSIM is particularly well-suited for assessing the perceptual quality of the super-resolved images, as it takes into account luminance, contrast, and structure similarities between the generated and ground truth images. SSIM was used to quantify the improvement in image fidelity and structural similarity achieved by a super-resolution algorithm applied to MRI brain scans. Meanwhile, PSNR provides a quantitative measure of reconstruction fidelity by comparing the pixel-wise differences between the super-resolved and ground truth images. Despite its simplicity, PSNR remains a widely used metric in MRI image processing due to its intuitive interpretation and computational efficiency. In addition to SSIM and PSNR, the LPIPS metric is incorporated to assess the perceptual similarity between the generated and ground truth images using learned features from deep neural networks. This metric offers a more sophisticated measure of image quality, capturing subtle perceptual differences that may not be adequately reflected by traditional metrics.

To present the results, we will refer to "SR with SI (ours)" as the abbreviation for "our super-resolution with slice interpolation method" for latent space interpolation, and "SR with SI (baseline)" as the abbreviation for "baseline super-resolution with slice interpolation" for image space interpolation. We refer to the combination of bicubic spatial super-resolution with linear interpolation of images in the latent space as *Bicubic SR with Autoencoder SI*. Similarly, the combination of SRGAN-based spatial super-resolution with linear interpolation of images in the pixel space is denoted as *SRGAN SR with SI*, and the combination of WDSR-based spatial super-resolution with linear interpolation of images in the pixel space is referred to as *WDSR SR with SI*. These naming conventions are maintained consistently throughout the results to facilitate comparison between the proposed method, the baseline, and the selected competitors.

3.3.1. Results for the OASIS dataset

First, we focus on evaluating the generative capability of the model. To do this, we compare the average value of the metrics described in Section 3.3 with the corresponding ground truth, as also detailed in Section 3.3. These averages demonstrate a significant improvement in the quality of the super-resolved images compared to both the Baseline model and the additional competitors, as shown in Table 1. The SSIM metric reveals a remarkable increase in structural similarity for the super-resolved images generated by our model, achieving an average value of 0.833. This is considerably higher than the Baseline (0.331) and surpasses the competitors, which remain around 0.81–0.82. This indicates that our method more effectively preserves anatomical structures across the reconstructed slices. PSNR results further confirm this advantage, with our model achieving an average of 39.13, while the Baseline remains at 30.047 and the competitors do not exceed 32.34. This higher PSNR reflects a significant reduction in reconstruction error and superior fidelity in the final images. Lastly, LPIPS shows an average value of 0.07, indicating minimal perceptual difference from the ground truth. Again, our method outperforms both the Baseline (0.769) and the competitors, which present LPIPS values ranging from 0.11 to 0.83, highlighting the superior perceptual quality of our reconstructions.

Next, we analyze the progression of the metrics across slices, considering those directly upscaled by the LDM, those generated in the latent space and then upscaled (our method), and those interpolated in the pixel space (baseline). As shown in Fig. 5, and consistently observed across all patients, our method maintains stable and superior performance across slices. Despite the natural variations in performance, the metrics demonstrate smooth behavior without abrupt drops in the generated slices. This stability is particularly relevant given the anisotropic nature of the dataset, limited to only 160 slices along the Z-axis, where maintaining consistency between adjacent slices is critical. Throughout

Table 1
Comparison of evaluation metrics with Ground Truth in the OASIS dataset.

Model	SSIM	PSNR	LPIPS
SR with SI (ours)	0.833	39.13	0.07
SR with SI (baseline)	0.331	30.047	0.769
Bicubic SR with Autoencoder SI	0.81	30.18	0.11
SRGAN SR with SI	0.82	32.34	0.36
WDSR SR with SI	0.82	29.05	0.83

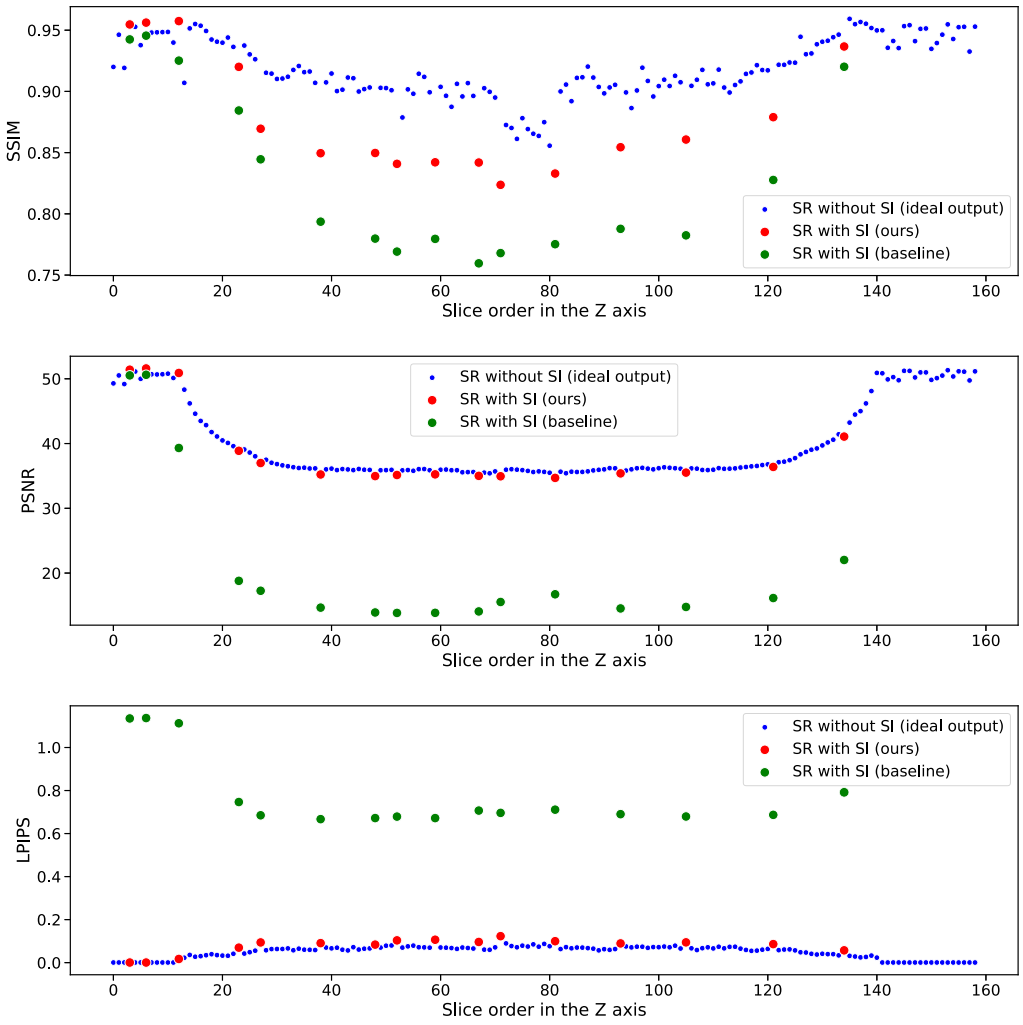

Fig. 5. Comparison of image resolution enhancement techniques against ground truth from the OASIS dataset. The ground truth images have a resolution of 256×256 pixels. Input images with a resolution of 64×64 pixels are upscaled to 256×256 pixels using the Latent Diffusion Model (super-resolution without slice interpolation). The blue dots represent the upscaled images compared to the ground truth on a per-slice basis. This line is broken at points where red and green dots are present to highlight the different methods. Red dots indicate points generated using our proposed method (super-resolution with slice interpolation, ours), where 10 % of the initial low-resolution slices are randomly removed and generated through latent space interpolation. Green dots represent another baseline method combined with LDM (super-resolution with slice interpolation, baseline), where the initial low-resolution slices removed are generated through pixel space interpolation.

the sequence, SSIM values remain high, confirming strong structural consistency between the super-resolved and ground truth images. PSNR values show minimal distortion across slices, even in those generated through latent interpolation, underscoring the robustness of our model in preserving image fidelity. Similarly, LPIPS remains low, demonstrating that perceptual features are successfully maintained throughout the reconstruction, providing high-quality, visually coherent results.

3.3.2. Results for the ATLAS dataset

First, as in the previous section, we focus on evaluating the generative capability of the model by comparing the average value of the metrics described in Section 3.3. The quantitative evaluation against the Ground Truth, shown in Table 2, reveals significant improvements in image quality metrics compared to both the Baseline model and the additional competitors, mirroring the trends observed in the OASIS dataset.

The SSIM values show a substantial enhancement in the structural similarity between the super-resolved images generated by the LDM model and the Ground Truth, achieving an average SSIM of 0.85. This value clearly outperforms not only the Baseline (0.404) but also the competing methods, which remain below 0.83. In terms of PSNR, the LDM model reaches an average of 33.777, demonstrating superior fidelity and a notable reduction in reconstruction error. In contrast, the Baseline achieves 28.876, and the competitors remain below 27.5, highlighting the improved ability of our model to recover fine anatomical details. Furthermore, the LPIPS metric, which assesses perceptual similarity, reinforces the superior performance of the LDM model, with an average LPIPS of 0.129. This result indicates a lower perceptual difference from the Ground Truth compared to both the Baseline (0.62) and the competing methods, which show higher perceptual dissimilarity with values between 0.21 and 0.63. These results confirm that our LDM-based approach not only surpasses the Baseline but also consistently outperforms alternative super-resolution strategies across all evaluated metrics, ensuring a better preservation of structural information, higher fidelity, and improved perceptual quality.

Table 2
Comparison of evaluation metrics with Ground Truth for the ATLAS dataset.

Model	SSIM	PSNR	LPIPS
SR with SI (ours)	0.85	33.777	0.129
SR with SI (baseline)	0.404	28.876	0.62
Bicubic SR with Autoencoder SI	0.82	26.62	0.21
SRGAN SR with Linear SI	0.82	27.12	0.56
WDSR SR with Linear SI	0.83	26.45	0.63

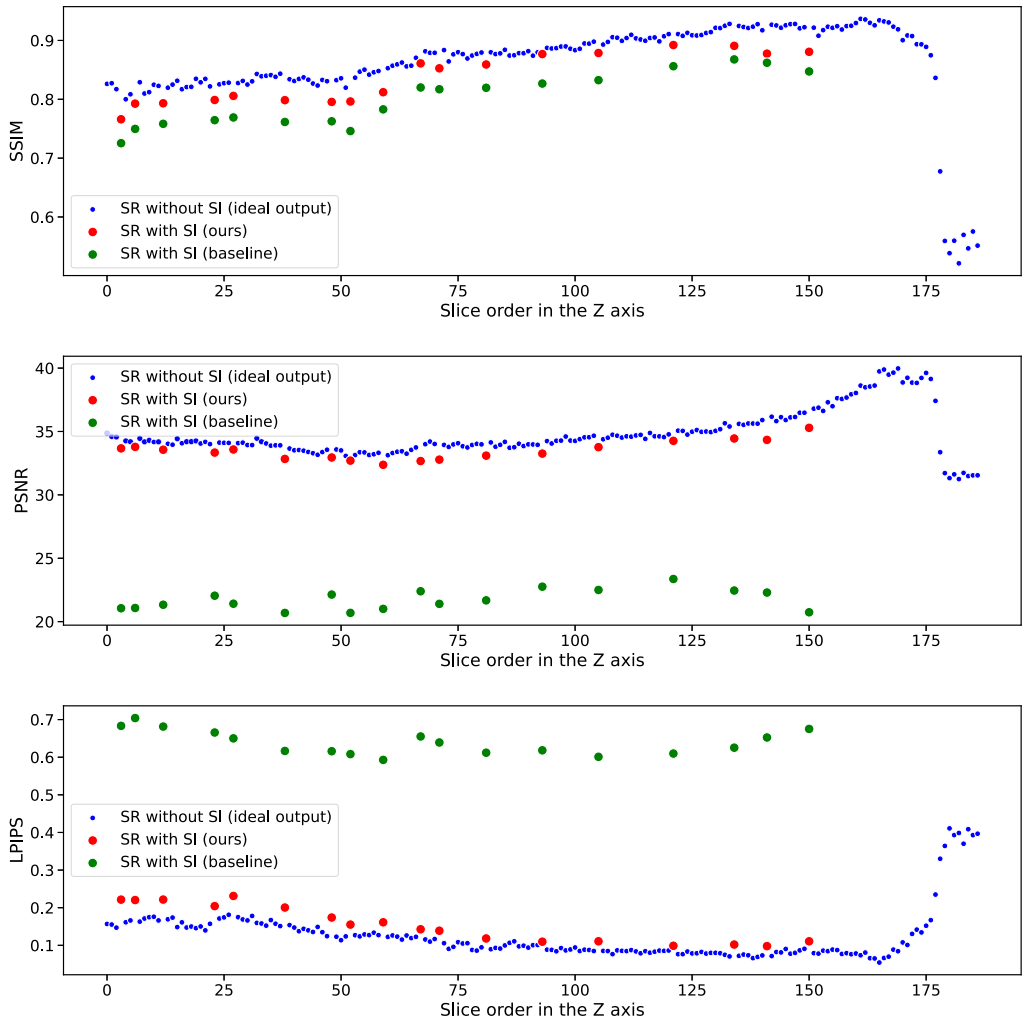

Fig. 6. Comparison of image resolution enhancement techniques against ground truth from the ATLAS test dataset for a specific patient. Ground truth images have a resolution of 256×256 pixels. Input images are downsampled to 64×64 pixels and then upscaled to 256×256 pixels using the Latent Diffusion Model (super-resolution without slice interpolation). The blue line illustrates the upscaled images compared to the ground truth on a per-slice basis. This line is interrupted at points where red and green dots appear, emphasizing the different methods. Red dots represent points generated using our proposed approach (super-resolution with slice interpolation, ours), where 10 % of the initial low-resolution slices are randomly removed and generated via latent space interpolation, then reconstructed by the decoder at 256×256 pixels. Green dots denote another baseline method combined with LDM (super-resolution with slice interpolation, baseline), where the interpolation is made directly in the pixel space. The graphical legend labels the blue line as "SR without SI," the red dots as "SR with SI (ours)", and the green dots as "SR with SI (baseline)".

Once again, a detailed evaluation of the performance across slices for a specific patient from the ATLAS test dataset (with similar trends observed across other patients) is illustrated in Fig. 6. As in Section 3.3.1, the metrics demonstrate stable behavior across slices, likely supported by the regular spacing between consecutive slices, which provides a consistent foundation for evaluation. This stability reinforces the robustness of the LDM in maintaining performance across varying anatomical contexts. The slice-wise analysis highlights the LDM's ability to preserve structural information with consistently high SSIM values, minimal distortion evidenced by stable PSNR scores, and low perceptual dissimilarity according to LPIPS. These findings confirm the effectiveness of our super-resolution model in producing high-quality reconstructions that reliably capture anatomical details while minimizing perceptual artifacts.

3.3.3. Results for the OpenNeuro dataset

For this experiment, we evaluate the performance of the model on the OpenNeuro dataset, which contains FLAIR MRI sequences. This dataset introduces valuable variability to our experiments, as FLAIR images emphasize different tissue contrasts and pathological features compared to the T1-weighted images of previous datasets, thus providing a broader validation of our method across diverse MRI modalities.

As shown in Table 3, our model achieves superior results in all metrics when compared to both the Baseline and the additional competitors. The average SSIM obtained by the LDM model is 0.815, demonstrating strong structural similarity with the Ground Truth. This significantly outperforms the Baseline model, which reaches only 0.617, and also surpasses the competing methods, which achieve SSIM values around 0.79–0.81. In terms of PSNR, although the difference between our model (71.669) and the Baseline (69.895) is smaller than in other datasets,

Table 3
Comparison of evaluation metrics with Ground Truth for the OpenNeuro dataset.

Model	SSIM	PSNR	LPIPS
SR with SI (ours)	0.8153	71.669	0.0952
SR with SI (baseline)	0.617	69.895	0.255
Bicubic SR with Autoencoder SI	0.81	56.62	0.21
SRGAN SR with Autoencoder SI	0.81	57.54	0.54
WDSR SR with Autoencoder SI	0.79	57.87	0.75

Table 4
Quantitative results on BraTS for T1, T2, and FLAIR images.

Model	PSNR	SSIM	LPIPS
BraTS T1			
SR with SI (ours)	37.11	0.92	0.07
SR with SI (baseline)	23.37	0.81	0.09
Bicubic SR with SI	32.08	0.85	0.14
SRGAN SR with SI	30.78	0.9014	0.44
WDSR SR with SI	31.95	0.90	0.77
BraTS T2			
SR with SI (ours)	38.35	0.91	0.05
SR with SI (baseline)	24.91	0.89	0.07
Bicubic SR with SI	29.41	0.75	0.14
SRGAN SR with SI	27.77	0.90	0.41
WDSR SR with SI	26.45	0.86	0.63
BraTS T2-FLAIR			
SR with SI (ours)	36.16	0.89	0.09
SR with SI (baseline)	18.71	0.7461	0.11
Bicubic SR with SI	29.78	0.80	0.14
SRGAN SR with SI	28.31	0.88	0.41
WDSR SR with SI	29.84	0.89	0.74

this can be attributed to the nature of the FLAIR sequences in OpenNeuro, which exhibit high similarity across consecutive slices with fewer abrupt anatomical variations. This inherent smoothness reduces the reconstruction challenge, resulting in closer PSNR values across models. However, despite this smaller gap, our method consistently delivers the highest PSNR, while the competing approaches remain well below 58, reinforcing the fidelity of our reconstructions. The LPIPS metric further highlights the advantage of our method, with an average value of 0.095, significantly lower than both the Baseline (0.255) and the competing methods, which range from 0.21 to 0.75. This result underscores the ability of our LDM-based approach to better preserve perceptual quality, capturing subtle visual details that are crucial in clinical imaging. Overall, these results clearly demonstrate the superiority of our latent space super-resolution over both the Baseline and alternative approaches. By performing interpolation directly in the latent space, our method achieves enhanced structural similarity (SSIM), reduced information loss (PSNR), and lower perceptual dissimilarity (LPIPS), confirming its robustness and adaptability across different MRI modalities and contrasts.

3.3.4. Results on the BraTS dataset

To enrich our experiments, which mainly consists of T1-weighted images such as OASIS and ATLAS, we incorporate the BraTS dataset due to its diversity of modalities (T1 (native), T2 (weighted), and T2-FLAIR). Table 4 presents the quantitative results obtained on BraTS for each image type, evaluated in terms of PSNR, SSIM, and LPIPS.

From the results, we observe that our model, consistently achieves the best performance across all modalities. Regarding PSNR, which measures the reconstruction fidelity, LDM surpasses the second-best method by a considerable margin in T1 and T2 images, and shows an even greater improvement in FLAIR. In terms of SSIM, which assesses structural similarity, our model achieves the highest values in T1 and FLAIR, while remaining highly competitive in T2. Lastly, LPIPS, which quantifies perceptual similarity, also favors our method, obtaining the lowest values across all modalities. These results highlight the capacity of our approach to adapt to the heterogeneity of brain MRI contrasts, providing robust and high-quality reconstructions regardless of the imaging sequence. By incorporating the BraTS dataset, which includes T1, T2, and FLAIR modalities, we demonstrate that our method generalizes effectively across different clinical MRI formats, extending the evaluation beyond T1-weighted datasets such as OASIS and ATLAS, and reinforcing its potential for application in diverse diagnostic scenarios.

3.4. Deep-significance

Deep-Significance (Ulmer, Hardmeier, & Frellsen, 2022) was used to compare the statistical significance of our results using the previously described metrics (SSIM, PSNR, and LPIPS). Deep-Significance offers a systematic framework for significance testing, which can aid in determining whether the differences observed between the two datasets are statistically significant. An aspect of this methodology that is particularly noteworthy is its orientation towards avoiding the use of p-values. The widespread criticism of p-values in the philosophy of science arises from their application in situations where the underlying assumptions are clearly not met. By providing a rigorous statistical approach, Deep-Significance enhances the reliability of the comparison, ensuring robust conclusions regarding the performance disparities between the OASIS, ATLAS, OpenNeuro and BraTS datasets.

In this context, $\epsilon_{\min}$ is a value returned by Absolute Stochastic Order, abbreviated as ASO, Del Barrio, Cuesta-Albertos, and Matrán (2018), indicating the upper bound on the extent of stochastic order violation. When $\epsilon_{\min} < \tau$ (where $\tau \leq 0.5$), it suggests that one algorithm, denoted as A, is more consistently dominant over another, represented as B. Essentially, smaller values of $\epsilon_{\min}$ signify higher confidence that algorithm A outperforms algorithm B across various scenarios. A value of $\epsilon_{\min} = 0$ was obtained for all three metrics. This indicates that the cumulative distribution function of the baseline is inferior to that of our proposal across the entire domain of definition of both functions. In other words, there is not any point in this domain where the baseline outperforms the proposal.

3.5. Qualitative study

In medical imaging, the quality and fidelity of generated images play a crucial role in diagnostic accuracy and clinical decision-making. Magnetic resonance imaging is a widely used modality for capturing detailed anatomical information. However, image resolution and quality can vary significantly depending on acquisition parameters and imaging protocols. The aim is to compare images generated by interpolation in the latent space of an LDM-model with those produced by the Baseline model and competitor models of Section 2.7. The model proposed in Section 2.4 offers the advantage of producing images with biologically plausible features devoid of pronounced grayscale artifacts. An examination is conducted on how well the generated images preserve anatomical structures and grayscale consistency, as these factors are critical for reliable clinical interpretation. To achieve this, slicing through the Z-axis is performed, and for each pair of images, n images are generated using the model described in Section 2.4. Visual comparisons will be performed in relevant areas of the datasets. We will test the ability to increase the arbitrary resolution capacity by multiplying the resolution by primes such as 11 and 5 that are not as common in the literature as for example 2 and 3 and their products (Chaudhari et al., 2018; Lyu, You, Shan, & Wang, 2018).

Furthermore, we will conduct a qualitative analysis to verify the absence of hallucinations (Sun, Zhu, & Tappen, 2010) in the images generated through the super-resolution model proposed. This entails a thorough examination to ensure that the visual content remains faithful to the original data without introducing any erroneous or misleading elements. Maintaining this standard of accuracy is essential for validating the reliability and effectiveness of our approach in producing high-quality image enhancements. To present the results, we will use the same names for the models as in the quantitative analysis of Section 3.3.

3.5.1. Results for the OASIS dataset

For the qualitative assessment, three slices from two distinct patients within the OASIS dataset were carefully handpicked, which we can observe in Figs. 7, 8 and 9. In each of these figures, the top row corresponds

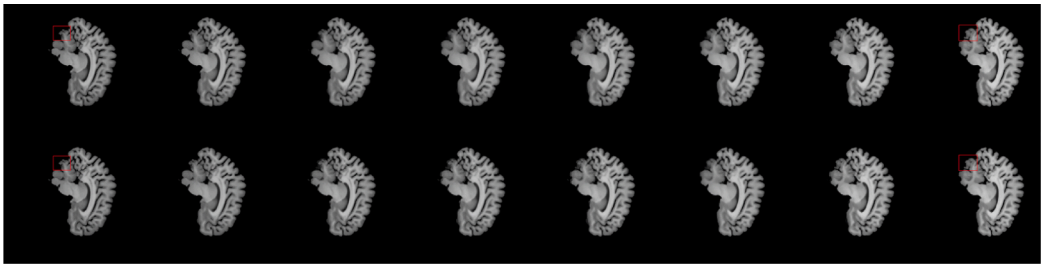

Fig. 7. For one of the test patients, six generated images in a region rich with details are presented. Our results are shown in the top row, while the baseline results are shown in the bottom row. The biological coherence of images generated with the LDM can be observed, highlighting the method's ability to maintain image fidelity across different levels of detail. This ensures that the generated images not only appear realistic but also retain critical biological structures, providing a more accurate representation compared to traditional methods. Additionally, the LDM method shows improved robustness against common artifacts, further enhancing the quality and reliability of the generated images.

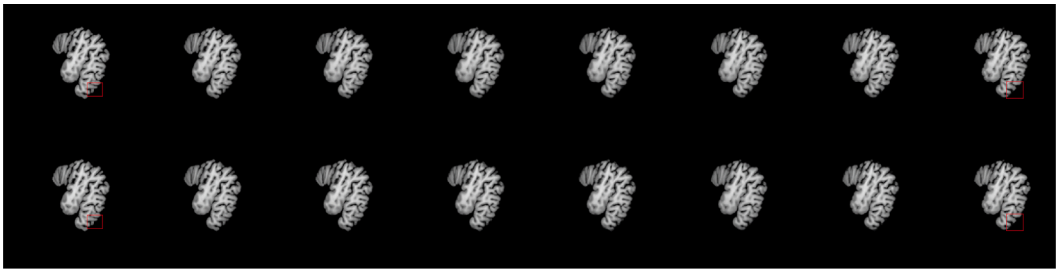

Fig. 8. For the same patient as Fig. 7, six additional generated images are presented in a different region rich with details. Our results are shown in the top row, while the baseline results are shown in the bottom row. The same property is observed. The regions where artifacts appear in the Baseline (pixel-space interpolation) have been highlighted with a red box.

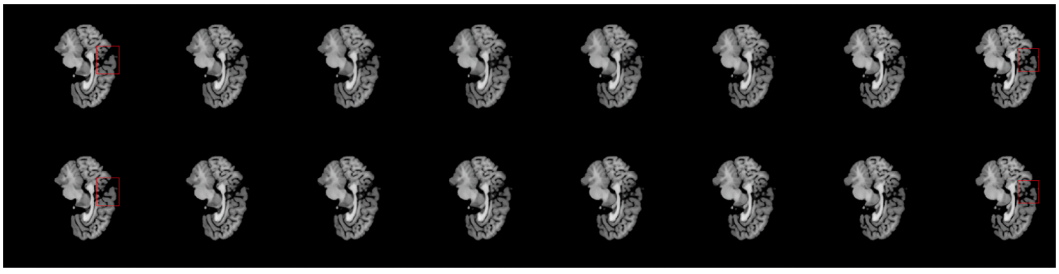

Fig. 9. The generative capability is also demonstrated in another patient from the test set, observing the same consistent generation property. Our results are shown in the top row, while the baseline results are shown in the bottom row. The regions where artifacts appear in the Baseline (pixel-space interpolation) have been highlighted with a red box.

to super-resolution with slice interpolation, while the bottom row corresponds to the baseline method. We can see in the upper row that the super-resolution with slice interpolation methodology demonstrates exceptional consistency in its generation process, effectively addressing the prevalent issue of sharp grayscale transitions. On the contrary, in the bottom row, starting from the third generated image, discernible smooth grayscale transitions emerge, deviating from biologically plausible representations. The images at the endpoints correspond to the ground truth, while the intermediate images represent the generated ones.

Utilizing the model exposed in Section 2.4, we generate images that reduce sharp transitions in grayscale while maintaining the biological coherence of the resonances. This model uses latent representations to facilitate the diffusion process, resulting in images that exhibit both smooth transitions in brightness levels and fidelity to the underlying biological structures present in the resonances, as we can see in Figs. 7, 8 and 9.

To show one of the key strengths of the proposed image generation methodology, we demonstrate the capability to enhance resolution along the Z-axis by prime factors of 5 and 11 in Figs. 11 and 10, respectively. This approach extends our previous example, where the resolution was augmented by a factor of 7. This demonstration underscores the arbitrary nature of resolution enhancement, thereby diverging from conventional practices such as doubling or tripling the resolution, which are prevalent in the existing literature, as well as other multipliers, such as 4 and 6, which are products of these factors. This method highlights the flexibility and superiority of the proposed model in achieving precise and varied resolution enhancements. It should be noted that despite the variety of images that have been used for our experiments, all the exposed properties are preserved.

Notably, the experiments were conducted using low-resolution input images of size 64×64, as shown in Fig. 12, which serve as the basis for generating high-resolution slices. By leveraging the proposed methodology, we successfully generate interpolated slices with a resolution that is four times higher than the original slice resolution, producing an output of 256×256 pixels, as illustrated in Fig. 13. This highlights the model's capacity to enhance spatial resolution while maintaining structural coherence across the generated slices, enabling flexible super-resolution and interpolation with arbitrary resolution increments.

3.5.2. *Results for the ATLAS dataset*

Despite the differences in the image type compared to the previous dataset, the LDM-model has demonstrated the capability to generate images in a coherent manner, in the same way as what was observed in the OASIS dataset. Our proposed LDM has been able to learn a regular latent space in which, similar to the other dataset, proximal images

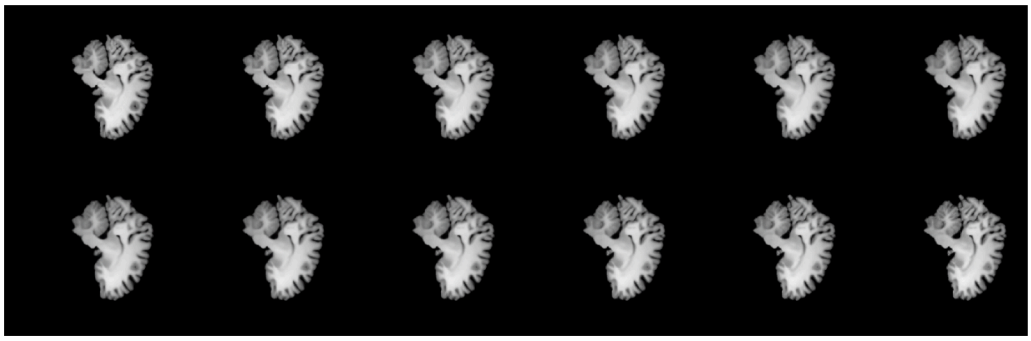

Fig. 10. Increased resolution by a factor of 11 for one of the OASIS patients, highlighting the value of the technique by being able to increase resolution by an arbitrary factor (in this case, a prime number) in contrast to other techniques limited to a single resolution increment. The left limit is on the top line, and the right limit is on the bottom line for a better appreciation of the image. The ten images generated are the 5 in the top row to the right of the left boundary and the 5 in the bottom row to the left of the right boundary.

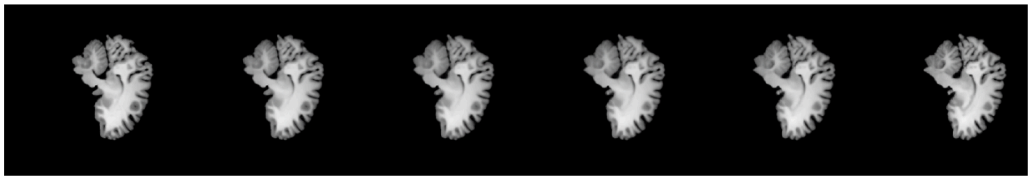

Fig. 11. Increased resolution by a factor of 5 for one of the OASIS patients, highlighting the value of the technique by being able to increase resolution by an arbitrary factor (in this case, a prime number). It can be observed that consistency is preserved in the generation by increasing the resolution for different factors since it is the same patient and the same left and right limits (slices) as in Fig. 10.

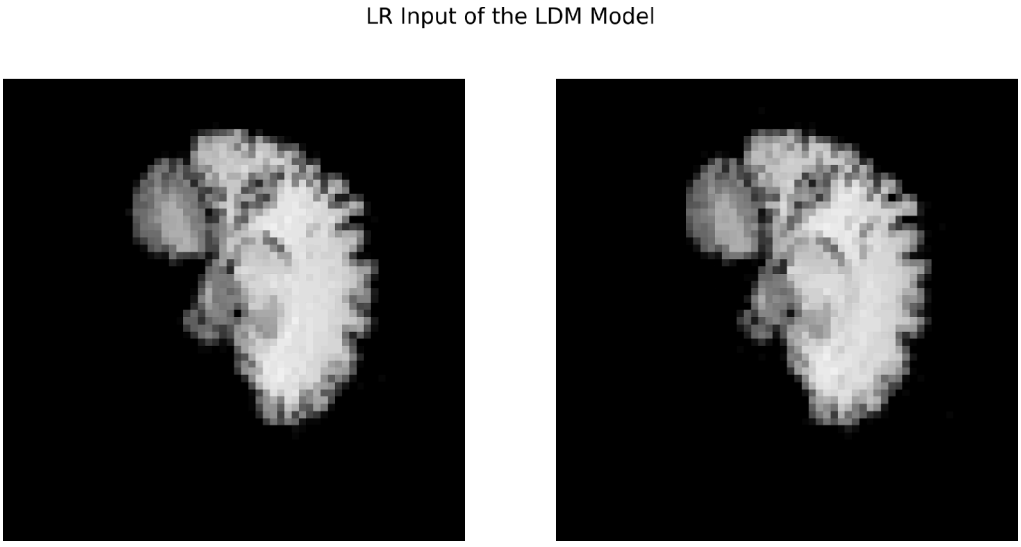

Fig. 12. Example of low-resolution input images used for super-resolution and interpolation. The input slices have a resolution of 64×64 pixels, serving as the foundation for the generative process.

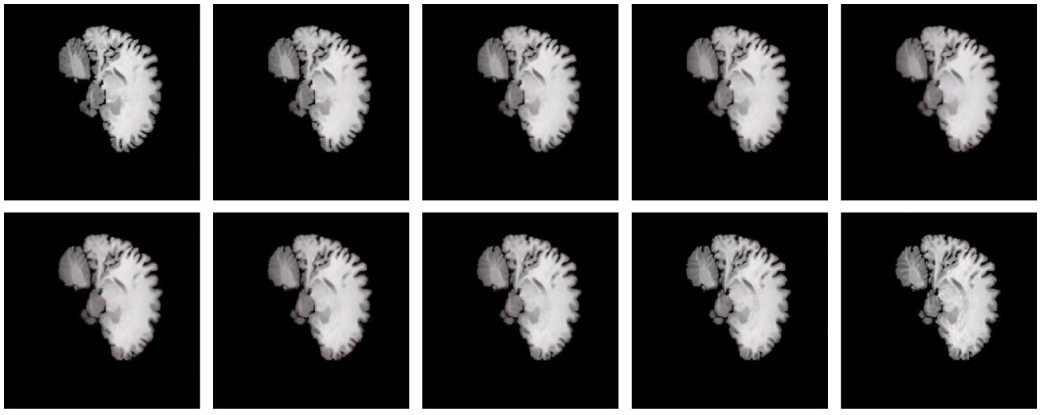

Fig. 13. Example of high-resolution images generated by the model from the input of Fig. 12. The resulting interpolated slices have a resolution of 256×256 pixels, demonstrating the ability of the model to enhance resolution by a factor of 9 while preserving biological coherence in the generation.

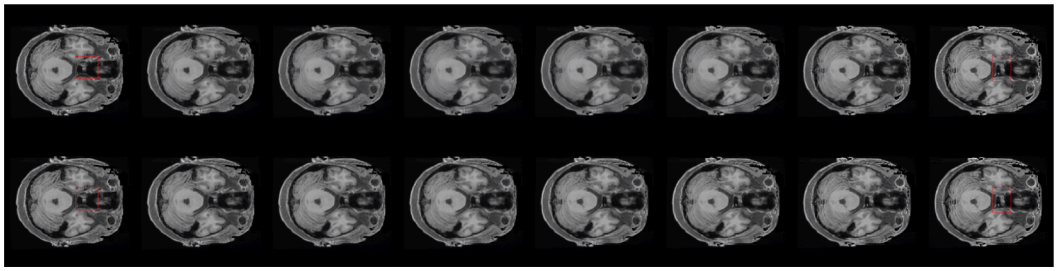

Fig. 14. Comparison with image generation for one of the test patients from the ATLAS dataset. Despite the differences with the images from OASIS, our LDM has been able to learn the images and generate them coherently. Our results are shown in the top row, while the baseline results are shown in the bottom row. In the area marked in red, it can be observed how the image to the left of the baseline gradually fades until it transforms into the image on the right. In contrast, with our method, the images preserve biological coherence, and the degradation effect on the grayscale gradation is reduced.

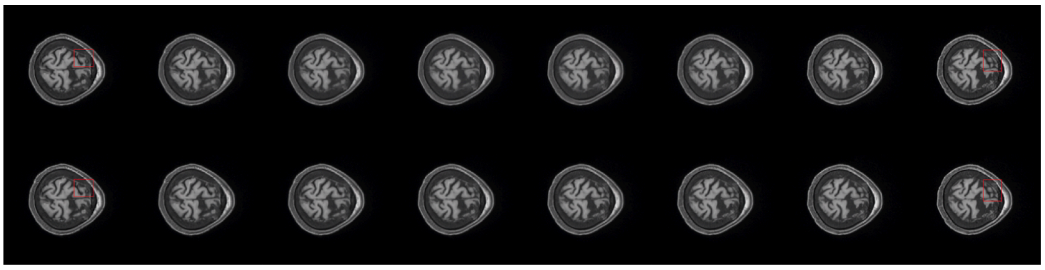

Fig. 15. Images depicting the generation of another part of the brain from the same patient are shown in Fig. 14. Our results are in the top row, while the baseline results are in the bottom row. In the area marked in red, the baseline images show a gradual fade from left to right. In contrast, our method maintains biological coherence and reduces grayscale degradation.

have translated into images with similar features, ideal for conducting interpolation and enhancing resolution through biologically meaningful generation as we can observe in Figs. 14 and 15 (as the previous section, the upper rows correspond to our method as described in Section 2.4, and the lower rows to the Baseline described in Section 2.5).

For this dataset, despite the considerable diversity among the images, we consistently obtain the same result. This suggests robustness in the inference process, wherein the latent diffusion model effectively captures underlying patterns and features across varied input images, leading to consistent outcomes.

To illustrate the same capability that was shown with OASIS with the ATLAS dataset, we similarly enhanced the resolution along the z-axis by prime factors of 5 and 11, as shown in Figs. 16 and 17, respectively. This method also builds on our previous enhancement by a factor of 7, highlighting the arbitrary nature of resolution increases. By diverging from common practices such as doubling or tripling the resolution and

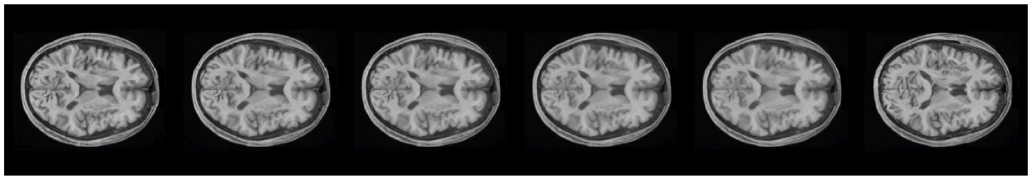

Fig. 16. Increased resolution by a factor of 5 for one of the ATLAS test dataset patients, highlighting the value of the technique by being able to increase resolution by an arbitrary factor (in this case, a prime number) in contrast to other techniques limited to a single resolution increment. It can be seen that biological consistency continues to be maintained in image generation.

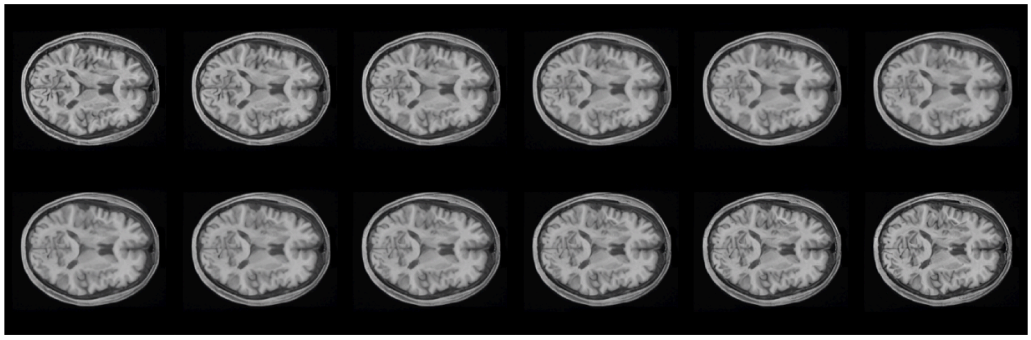

Fig. 17. Increased resolution by a factor of 11 for one of the ATLAS patients, highlighting the value of the technique by being able to increase resolution by an arbitrary factor (in this case, a prime number). It can be observed that consistency is preserved in the generation by increasing the resolution for different factors since it is the same patient and the same left and right limits (slices) as in Fig. 16.

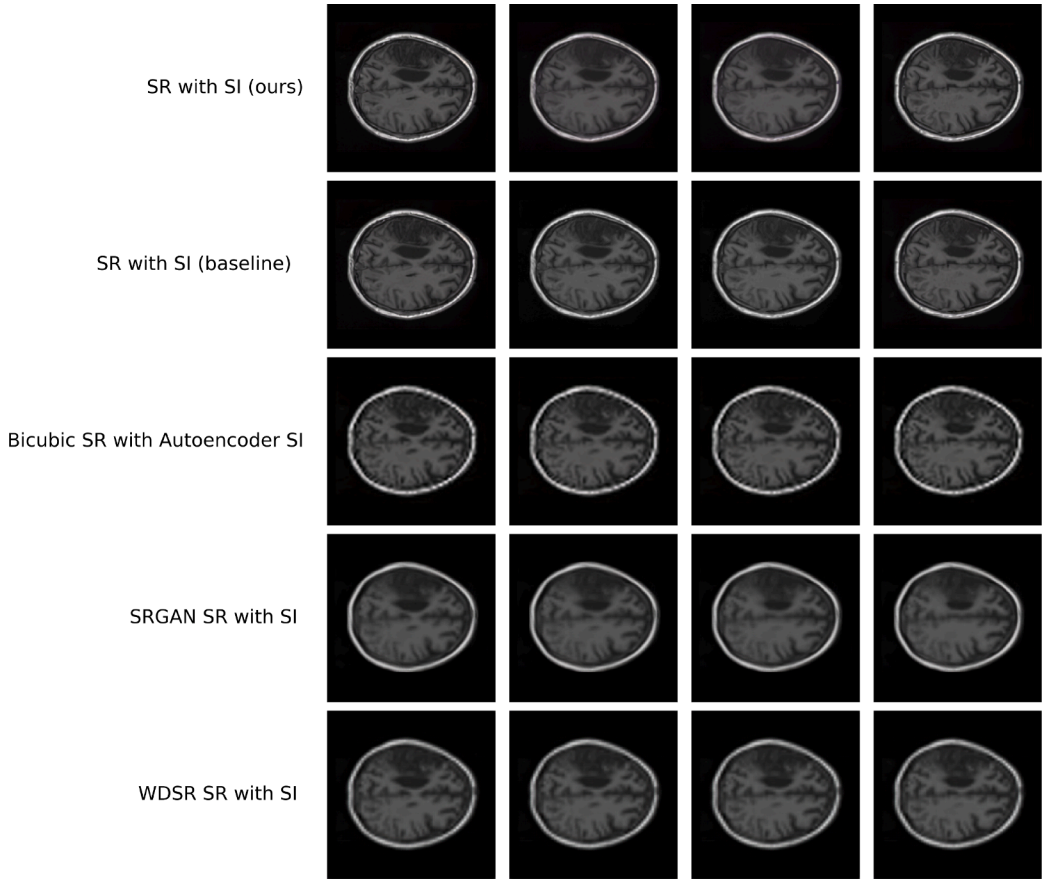

Fig. 18. Comparative analysis of super-resolution with slice interpolation on a T1-weighted MRI from the ATLAS dataset. Our method (SR with SI (ours)) preserves anatomical coherence, maintaining clear gray and white matter separation and cortical structures, outperforming baseline and competing approaches that suffer from blurring and structural inconsistencies.

other multipliers like 4 and 6, this approach demonstrates the flexibility and effectiveness of the proposed model in achieving precise and varied resolution enhancements.

In addition, Fig. 18 presents the comparative analysis for the T1-weighted images from the ATLAS dataset. This modality is particularly relevant for evaluating the preservation of fine anatomical details in healthy brain structures, such as the delineation between gray and white matter and the definition of cortical folds. Our method (SR with SI (ours)) outperforms the baseline and other competing methods by maintaining a higher degree of structural coherence throughout the interpolated slices. Unlike the competing methods, which rely on pixel-space interpolation and tend to introduce blurring and inconsistencies across the interpolated volume, our approach preserves sharp anatomical boundaries and reduces artifacts, ensuring biologically plausible transitions. These results confirm the capacity of our latent space interpolation to maintain anatomical fidelity in T1-weighted brain images, supporting its applicability to structurally detailed datasets like ATLAS.

3.5.3. Results for the OpenNeuro dataset

Although the OpenNeuro dataset differs from the previous ones, the LDM model has once again demonstrated its ability to generate coherent images, as previously observed in the ATLAS and OASIS datasets. The model has learned a structured latent space in which images that are close to each other correspond to images with similar features, making it suitable for interpolation and resolution enhancement through biologically meaningful generation. Figs. 19 and 20 illustrate these results.

In this dataset, despite the diversity in image types, we observe consistent results, indicating the robustness of the inference process. The LDM model effectively captures the underlying features and patterns of the input images, leading to biologically coherent outputs across varied inputs. To demonstrate the flexibility of our approach, we performed resolution enhancement along the Z-axis by prime factors of 7, as shown in Figs. 19 and 20. This builds upon the strategy used in the other datasets, further demonstrating the arbitrary nature of resolution enhancement achieved with our model. The ability to enhance resolution using unconventional prime factors, as opposed to common practices like doubling or tripling, underlines the effectiveness of our method in generating high-quality images.

Fig. 21 shows the resolution enhancement of a different OpenNeuro patient, where the resolution is increased by a factor of 5. This enhancement demonstrates the flexibility of our approach, which allows for arbitrary resolution scaling, not limited to conventional factors like 2 or 3. In this case, the increase by a factor of 5 highlights how the model preserves the biological consistency of the original image while enriching the detail and clarity. Despite the increased resolution, our method ensures that important structural features of the brain are maintained, which is crucial for medical imaging applications. The enhanced image demonstrates a significant improvement in clarity, with finer details becoming visible while avoiding the fading or distortion often seen in traditional resolution enhancement methods.

Fig. 22 presents the comparative analysis for the FLAIR images from the OpenNeuro dataset. This modality is particularly useful for highlighting hyperintense lesions and subtle pathological changes in the brain, making it essential for evaluating the capacity of the model to preserve clinically relevant features during interpolation and super-resolution. As observed in previous datasets, our method (SR with SI (ours)) clearly outperforms the baseline and competing methods by maintaining greater anatomical coherence, reducing blurring, and

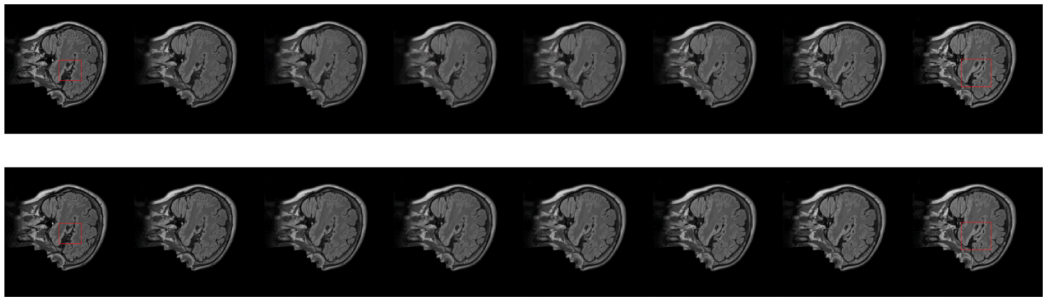

Fig. 19. Resolution enhanced by a factor of 7 for one of the OpenNeuro patients, showing the robustness of our method in maintaining biological consistency across different parts of the brain. Our results are presented in the top row, while the baseline results are in the bottom row. The red-marked region shows how, with the baseline method, the image to the left fades progressively into the image to the right, while our method preserves coherence in structure and intensity.

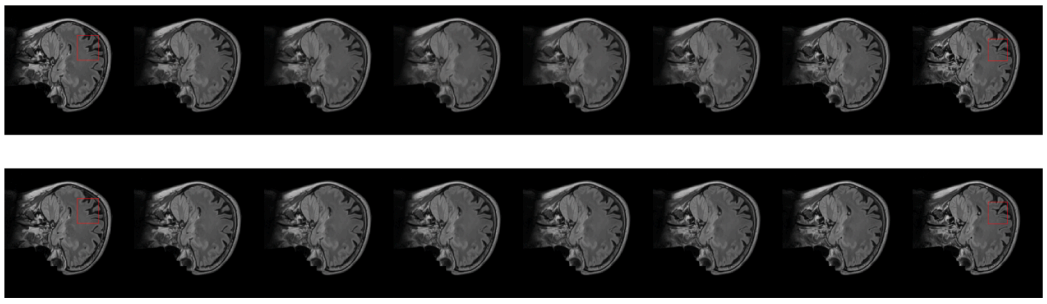

Fig. 20. Resolution enhanced by a factor of 7 for another OpenNeuro patient. The top row shows our results, while the bottom row shows the baseline results. The region marked in red highlights how our method maintains biological consistency and reduces the fading effect seen in the baseline approach, even with an increased resolution factor.

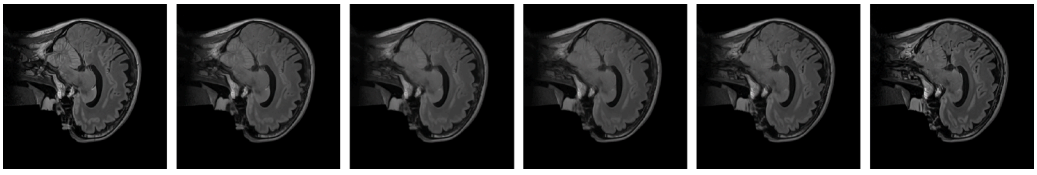

Fig. 21. Resolution enhanced by a factor of 5 for another OpenNeuro patient. We can appreciate that biological coherence is maintained despite changing the super-resolution factor and the dataset. We obtain results that are very comparable to those of OASIS and ATLAS despite the variety in the datasets.

preserving the contrast of hyperintense regions throughout the interpolated slices. Unlike the competing methods, which tend to introduce noticeable artifacts and loss of structural integrity due to pixel-space interpolation, our approach, through latent space interpolation, ensures smooth transitions and the preservation of critical lesion information. This confirms the robustness of our method when dealing with the particular challenges of FLAIR imaging, ensuring that both tissue integrity and pathological areas are accurately maintained across the interpolated volume.

3.5.4. Qualitative evaluation on BraTS

In this section, we perform a qualitative analysis of the proposed method on the BraTS dataset, with a particular focus on assessing its ability to preserve structural coherence in tumor lesions. Given the clinical relevance of accurately maintaining the morphology and boundaries of pathological regions, this evaluation aims to visually verify that the super-resolved and interpolated slices remain consistent and anatomically plausible in areas affected by tumors. To introduce variability into the analysis and ensure a comprehensive assessment, we focus specifically on two distinct MRI modalities within the BraTS dataset: T2-weighted and FLAIR images. These modalities provide complementary information about tissue characteristics and lesion visibility, allowing us to explore the robustness of the method across different contrast profiles and pathological presentations.

In the T2-weighted sequences, we observe that the model successfully preserves both the hyperintense regions corresponding to tumor areas and the surrounding white matter, maintaining continuity and coherence across the interpolated slices. In particular, the model accurately reconstructs the characteristic hyperintense signals of peritumoral edema and infiltrative regions, while also respecting the typical contrast between the white matter and gray matter, which is essential for preserving the anatomical integrity of the brain. Moreover, the transition of tumor margins across the interpolated slices remains smooth and consistent, preventing the appearance of artificial discontinuities or distortions between adjacent slices. This is particularly relevant in BraTS data, where accurate representation of tumor boundaries, lesion morphology, and surrounding white matter structures is essential for clinical interpretation.

These observations are visually confirmed in Figs. 23 and 24, which show examples of interpolation in T2-weighted images with $n = 11$ and $n = 5$, respectively. In both cases, the generated slices exhibit clear preservation of tumor regions, including their shape and intensity, as well as stable reconstruction of the white matter across the interpolated volume. This highlights the model's ability to generate high-quality anisotropic reconstructions while respecting clinically relevant features.

It is important to highlight that Fig. 25 presents the comparative analysis specifically for the T2-weighted images from the BraTS dataset. This modality is particularly relevant for evaluating the preservation of fine anatomical structures and tumor-related abnormalities. In this context, our method (SR with SI (ours)) achieves superior performance by maintaining the continuity of the white matter and the integrity of hyperintense tumor regions across the interpolated slices. In contrast, the competing methods, which rely on pixel-space interpolation, exhibit

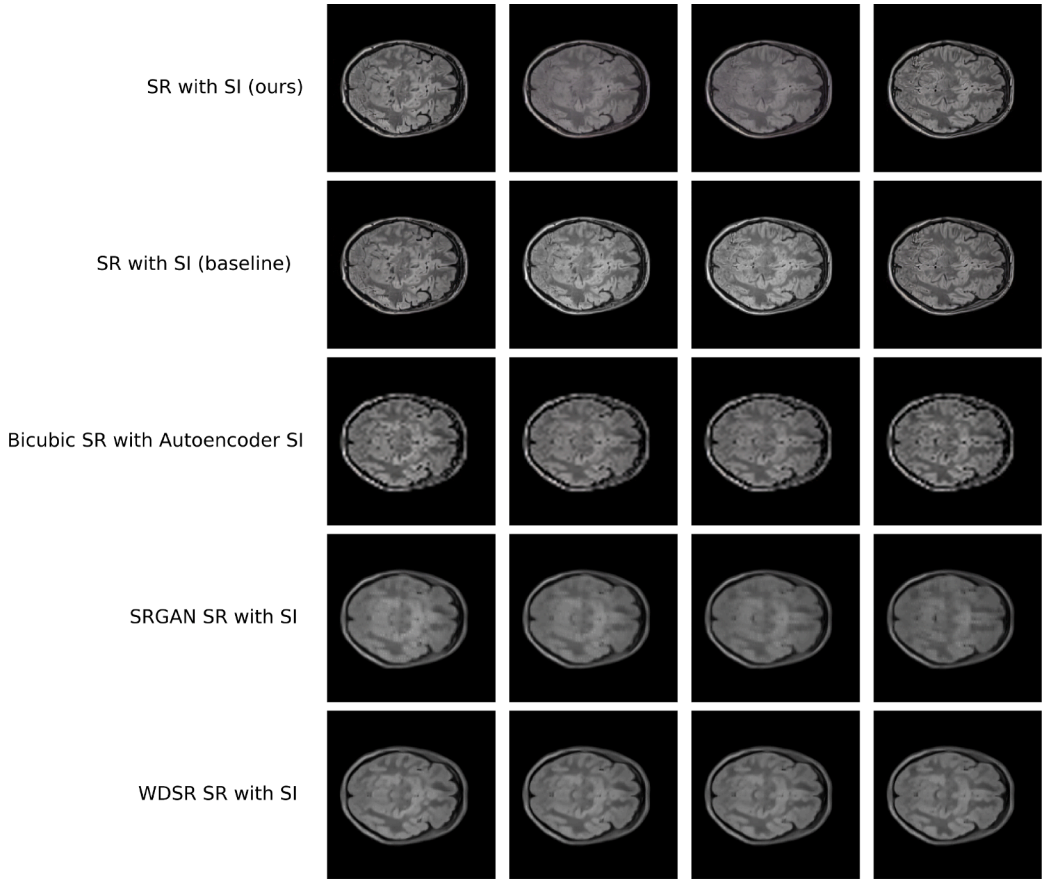

Fig. 22. Comparative analysis of super-resolution with slice interpolation on a FLAIR-weighted MRI from the OpenNeuro dataset. Our method (SR with SI (ours)) preserves the coherence of hyperintense lesions and surrounding brain structures more effectively than the baseline and competing approaches, which suffer from blurring and loss of detail due to pixel-space interpolation.

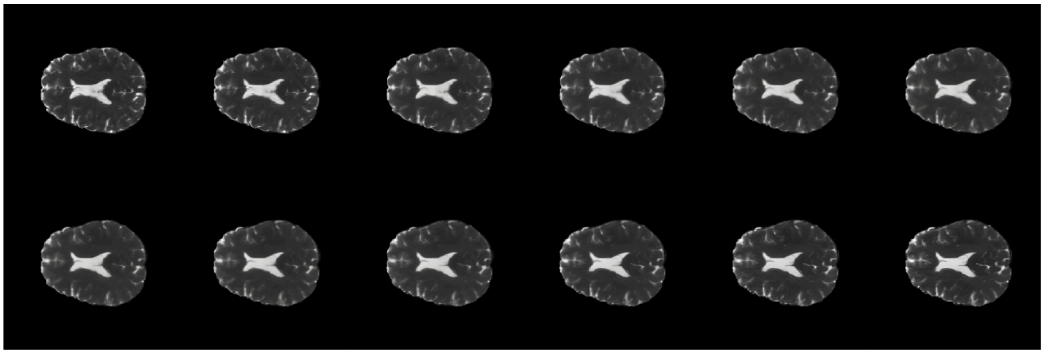

Fig. 23. Example of T2-weighted MRI slice interpolation with $n = 11$. The model preserves tumor regions and surrounding white matter with high structural consistency.

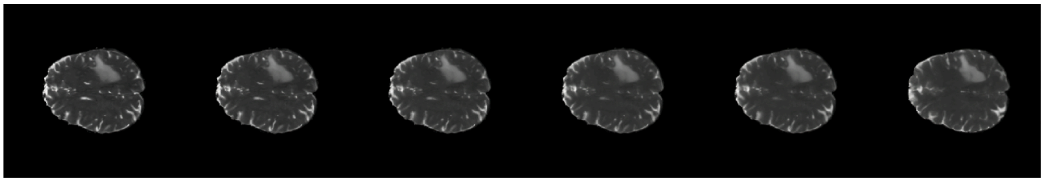

Fig. 24. Example of T2-weighted MRI slice interpolation with $n = 5$. Tumor boundaries and white matter remain well-defined and anatomically plausible.

noticeable artifacts such as blurring, loss of structural definition, and inconsistencies in the representation of critical regions. These differences emphasize the effectiveness of latent space interpolation in preserving biologically meaningful details and ensuring coherent transitions between slices, which are essential for accurate clinical interpretation.

As can be observed in (Fig. 26), in the FLAIR sequences, the model demonstrates a strong ability to reserve the hyperintense regions typically associated with tumor-related edema and necrotic cores, which are particularly well-visualized in this modality. FLAIR images are essential for highlighting fluid-attenuated abnormalities, and our method maintains the integrity of these pathological areas throughout the inter-

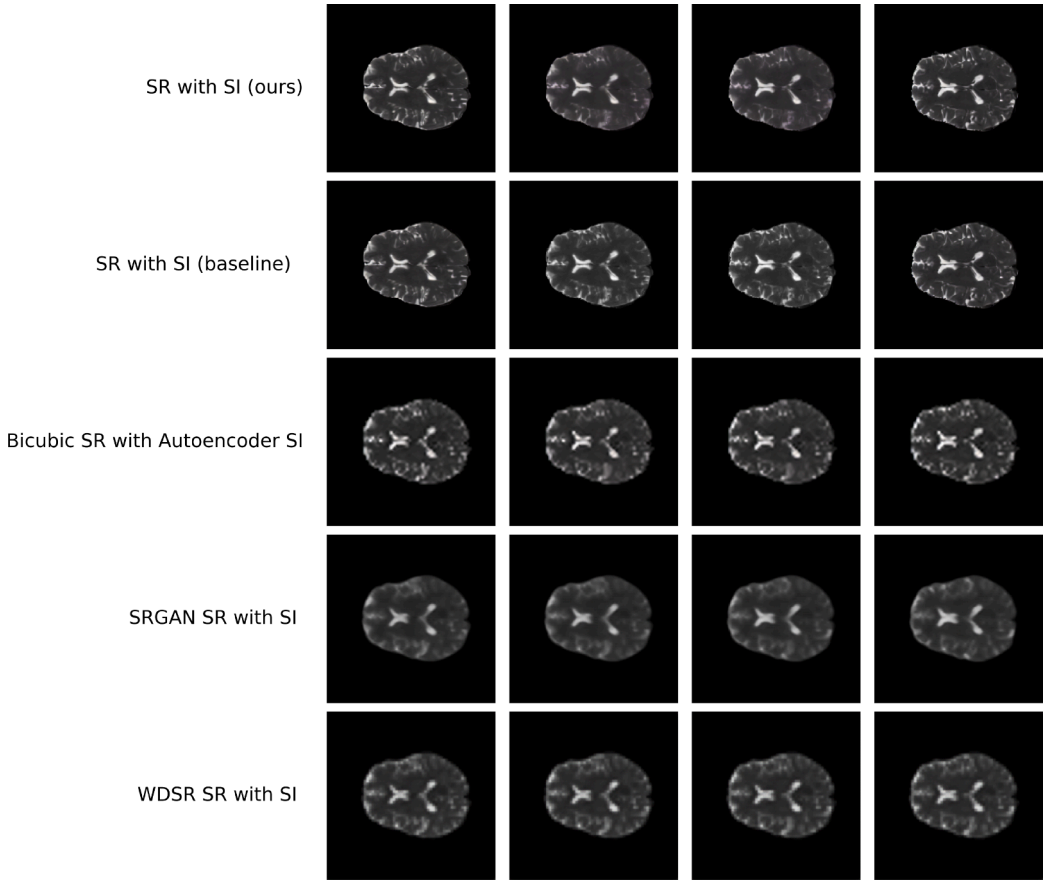

Fig. 25. Comparative analysis of different super-resolution algorithms with slice interpolation methods on a T2-weighted MRI from the BraTS dataset with $n = 3$. Our method (SR with SI (ours)) achieves superior anatomical coherence, preserving both tumor boundaries and white matter structures with greater continuity and biological plausibility than competing approaches, which suffer from blurring and loss of structural definition due to pixel-space interpolation.

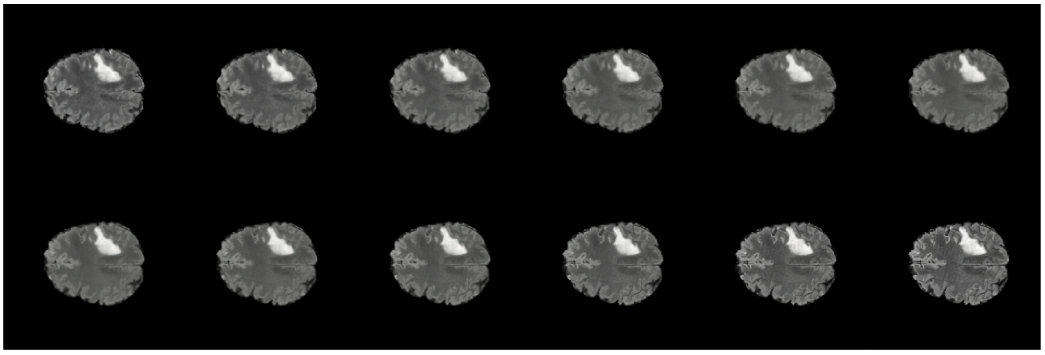

Fig. 26. Example of FLAIR MRI slice interpolation with $n = 11$. The model preserves hyperintense tumor areas and surrounding tissue contrast, ensuring consistent representation of edema and necrotic regions across interpolated slices.

polated slices. Specifically, we observe that the interpolated slices accurately reconstruct the extent and intensity of the peritumoral edema, ensuring that the lesion remains spatially coherent and morphologically consistent across the generated volume. Additionally, the surrounding brain tissue retains its structural fidelity, with no evidence of artificial blurring or deformation introduced during the interpolation process.

3.5.5. Medical validation through radiologist assessment

To ensure the clinical relevance and anatomical correctness of our proposed method, four expert radiologists conducted a qualitative evaluation. The evaluation consisted of a blind survey, conducted in a similar fashion as the blind survey shown in Table 1 of the study by Khader et al. (2023), where radiologists assessed different sets of generated images without prior knowledge of the method used to generate them. In the assessment of our proposal, each dataset consisted of a sequence of 10 interpolated slices for one sample of the ATLAS dataset to evaluate the realism and consistency of the generated images. Each dataset was evaluated based on the following three criteria, and each question was scored using four levels: *Very Bad, Bad, Good*, and *Very Good*.

- Q1 - Realistic image appearance: Assesses whether the generated images resemble realistic MRI scans.
- Q2 - Consistency between slices: Evaluates the degree of coherence between consecutive slices.
- Q3 - Anatomic correctness: Determines whether the anatomical structures are realistically preserved.

Table 5
Most common radiologist evaluation scores for different super-resolution methods.

Method	Q1 Realism	Q2 Slice Consistency	Q3 Anatomic Correctness
SR with SI (ours)	**Good**	**Good**	**Good**
SR with SI (baseline)	Good	Good	Good
Bicubic SR with Autoencoder	Bad, Good	Bad	Bad, Good
SRGAN SR with Linear SI	Bad	Good	Bad
WDSR SR with Linear SI	Bad	Bad, Good	Bad

The results of the survey are shown in Table 5, highlighting several key advantages of our LDM-based approach:

- **Strong realism:** Our method consistently achieved high ratings in image realism (Q1), reinforcing the ability of latent diffusion models to generate MRI scans with convincing textures and contrast. This suggests that our approach effectively reconstructs anatomical structures with a natural appearance, closely resembling real medical images.
- **Excellent slice consistency:** LDM received one of the highest ratings in slice consistency (Q2), significantly outperformed WDSR, which exhibited notable inconsistencies between slices. This result confirms the effectiveness of latent space interpolation in generating smooth and coherent transitions across consecutive slices, which is essential for maintaining anatomical continuity in 3D medical imaging.
- **Good anatomical correctness with room for refinement:** While the baseline approach received high ratings in anatomic correctness (Q3) as well, our method remains highly competitive. The ability of LDM to generate detailed, structurally coherent images makes it a strong alternative for super-resolution in MRI, as highlighted in Section 3.5. However, a slight drawback is that, while LDM effectively reduces the gradient artifacts common in pixel-space interpolation, it may introduce subtle blurring or minor losses in fine anatomical details. Future refinements could enhance the trade-off between smooth transitions and high-fidelity anatomical preservation.

Although our LDM-based approach already demonstrates strong performance, future work will focus on improving anatomical accuracy to match or even exceed the baseline. Refinements may include adjusting the loss function to emphasize structure-aware reconstructions, integrating additional medical priors, and optimizing model conditioning techniques to reinforce anatomical fidelity. The validation results confirm that LDM-based super-resolution is a highly promising strategy for medical imaging, offering an optimal balance between realism, coherence, and resolution enhancement. With further improvements, this approach has the potential to become a leading solution for high-quality MRI reconstruction, enabling clinicians to access sharper, more reliable images for improved diagnostic accuracy.

3.6. Comparison of sampling hyperparameters

To better understand the trade-offs between sampling speed and reconstruction quality, we conducted an extensive evaluation of different sampling strategies, timesteps (T), and η values on the BraTS dataset, which contains a wide variety of different MRI modalities. Table 6 summarizes the results obtained with both DDIM and PLMS samplers under varying configurations, providing insight into how these hyperparameters impact both performance and computational cost in a clinically relevant setting involving heterogeneous MRI contrasts.

As shown in Table 6, we selected DDIM with $\eta = 1$ and $T = 200$ for our experiments, as this configuration provides results that indicate that this setting consistently yields high PSNR and SSIM values while keeping LPIPS low, ensuring perceptual similarity with ground truth images.

Table 6
Comparison of sampling strategies with different timesteps (t_{steps}), η values for DDIM, and corresponding performance metrics.

Sampler	t_{steps}	η	PSNR	SSIM	LPIPS	Sampling Time (s)
PLMS	100	–	27.4515	0.7192	0.0819	23.13
PLMS	150	–	28.0942	0.7394	0.0782	37.40
PLMS	200	–	28.0118	0.7251	0.0727	44.73
PLMS	250	–	29.2369	0.7455	0.0544	55.90
DDIM	100	0	27.9646	0.7217	0.0782	22.26
DDIM	100	0.25	30.2048	0.7790	0.0529	22.26
DDIM	100	0.5	32.4182	0.8358	0.0536	22.23
DDIM	100	0.75	32.7358	0.8518	0.0595	22.24
DDIM	100	1	32.8052	0.8559	0.0615	22.23
DDIM	150	0	29.0160	0.7410	0.0580	37.16
DDIM	150	0.25	30.4254	0.7857	0.0523	37.14
DDIM	150	0.5	32.4636	0.8374	0.0518	37.15
DDIM	150	0.75	32.7680	0.8530	0.0590	37.15
DDIM	150	1	32.7918	0.8561	0.0613	37.17
DDIM	200	0	28.4125	0.7326	0.0722	44.49
DDIM	200	0.25	29.9471	0.7727	0.0539	44.47
DDIM	200	0.5	32.3648	0.8348	0.0516	44.52
DDIM	200	0.75	32.7229	0.8527	0.0578	44.50
DDIM	200	1	32.8488	0.8568	0.0610	44.50
DDIM	250	0	28.0914	0.7280	0.0770	55.63
DDIM	250	0.25	30.8325	0.7874	0.0473	55.63
DDIM	250	0.5	32.4049	0.8359	0.0517	55.63
DDIM	250	0.75	32.6989	0.8512	0.0586	55.65
DDIM	250	1	32.7807	0.8563	0.0601	55.63

Specifically, setting $\eta = 1$ introduces stochasticity into the sampling process, which helps improve the diversity and robustness of the generated outputs while maintaining stable and high-quality reconstructions. Additionally, using $T = 200$ timesteps ensures a sufficiently gradual denoising trajectory to preserve fine anatomical details without incurring excessive computational costs. This combination is particularly advantageous in medical imaging scenarios where both structural detail and variability across samples are important considerations.

We opted for DDIM over PLMS as it provides greater flexibility through the η parameter, allowing control over the stochasticity of the sampling process. This is particularly advantageous when aiming to balance diversity and reconstruction quality, whereas PLMS, although effective for fast and stable sampling with fewer steps, is fully deterministic and less adaptable to different noise schedules and clinical scenarios requiring variability in outputs.

3.7. Computational efficiency

Under the experimental conditions described in Section 3.1, we evaluate the computational efficiency of our method using the BraTS dataset. One of the key advantages of our method lies in its linear computational complexity with respect to the number of interpolated slices when increasing anisotropic resolution. This is because the DDIM sampling process, which transforms random noise into a latent representation, operates with constant complexity regardless of the interpolation factor. Consequently, the only component that scales with the number of interpolated slices is the linear interpolation step itself, making the overall process highly efficient and predictable as resolution requirements grow.

For the super-resolution process, the total sampling time varies according to the isotropic resolution multiplication factor $n + 1$, where n is the number of interpolated slices. To illustrate this, we analyze the case of a representative patient from the dataset, where the original MRI volume dimensions are $64 \times 64 \times 155$.

For this specific BraTS patient, the total sampling times obtained were:

- $n = 10$: Total sampling time of 95.92 seconds.
- $n = 4$: Total sampling time of 72.58 seconds.

- $n = 2$: Total sampling time of 69.16 seconds.

It is important to note that the computational cost of the method is mainly concentrated in the sampling process, as this phase involves the iterative denoising required to generate the latent representation from noise. Despite this, the GPU memory usage has remained stable across all generations, with a consistent consumption of 5.82 GB as the max usage. This stability further highlights the scalability and efficiency of our approach, making it particularly suitable for applications involving large volumetric medical datasets and high-resolution anisotropic reconstructions.

4. Discussion

The proposed methodology establishes a way to learn a latent space within a deep-learning diffusion model that is topologically meaningful. That is, small distances in the latent space translate to small variations in the represented image. More importantly, the images represented by the intermediate points in the latent space are valid, realistic magnetic resonance images. Critically, this intrinsic topological property of the learned latent space enables the employment of interpolation in such a space. Consequently, continuous and smooth interpolation is effectively attained, which enables arbitrarily large or fractional zoom levels, with stable results. This capability is out of reach for most previous super-resolution deep learning models, which must be trained for a specific zoom factor. The range of application of the proposed methodology extends well beyond the MRI domain because it can be applied to other kinds of images, medical or not, which form manifolds within the space of all possible images.

From a clinical perspective, the proposed method offers a set of practical and impactful contributions that extend beyond its computational and algorithmic strengths. In real-world diagnostic settings, particularly within public healthcare systems or regions with limited access to high-end imaging infrastructure, professionals often work with outdated MRI scanners, limited hardware acceleration, and increasingly saturated patient schedules. Our latent diffusion model addresses these challenges by providing a means to enhance the resolution of volumetric MRI data without increasing scan times or imposing additional technological requirements.

The proposed method allows for spatial resolution enhancement in all three axes by generating inter-slice content, while simultaneously increasing intra-slice detail. This dual improvement makes it possible to recover fine-grained anatomical structures (such as subtle lesions, boundary transitions, and tissue variations) that are frequently compromised in fast or low-resolution acquisitions. Such reconstruction quality is of particular importance in clinical disciplines like neurology, oncology, and radiology, where early and accurate identification of structural anomalies has a direct impact on diagnosis and treatment decisions. Moreover, the model's computational efficiency and reduced memory requirements enable deployment not only on high-performance GPU clusters but also on mid-tier hospital workstations or cloud-based systems with scalable resources. This is a key feature for enabling wider adoption, especially in hospitals where upgrading imaging hardware is logistically or financially unfeasible. In such contexts, our approach can be positioned as a software-level enhancement tool that extends the lifespan of existing MRI equipment, maintaining or even improving diagnostic performance over time.

An equally important clinical consideration is the preservation of anatomical fidelity. Our qualitative assessments show that the model maintains biological plausibility and avoids common artifacts introduced by some generative techniques, such as oversmoothing or hallucinated detail. This characteristic is crucial for building clinical trust and ensuring that the enhanced images support, not hinder, medical decision-making. Preliminary feedback from radiologists has also confirmed the clinical plausibility of the reconstructed images, reinforcing the model's potential for real-world diagnostic use.

Beyond static MRI use cases, the general framework of this method opens possibilities for adapting the same principles to dynamic or multimodal clinical scenarios. For instance, the interpolation strategy in the latent space could be extended to improve resolution in time-dependent sequences, such as functional MRI (Ota et al., 2022) or videofluoroscopy (Kim, Choi, Kim, & Shin, 2024). Similarly, the method may be adapted for integration with radiomics pipelines (Xing et al., 2024), where improved spatial granularity can increase the predictive power of extracted features. While these results are promising, further validation using private clinical datasets that reflect a wider range of real-world variabilities could help strengthen the method's clinical applicability and generalizability. Additionally, this work focused on MRI; the application of the method to CT, ultrasound, or even PET images remains an exciting avenue to explore. Future work will concentrate on expanding the clinical validation through direct integration in radiological workflows.

5. Conclusions

The research presented demonstrates a significant advancement in the field of medical imaging, particularly in enhancing the resolution of 3D brain images using a Latent Diffusion Model (LDM). By generating between slices along one of the image axis, this method effectively increases the density of data points in the image, capturing fine details that are often missed in traditional imaging techniques, additionally, it has the added value of enhancing the resolution of each individual slice, thereby increasing the resolution of the entire volume across all axes. This breakthrough has the potential to revolutionize the way medical professionals visualize and analyze brain structures, offering a more detailed and accurate representation than ever before. In addition, the resolution can be increased by any factor, which is the most notable development. The utilization of several datasets underscores the versatility and robustness of the LDM, proving its efficacy across different cohorts and imaging modalities.

Quantitative and qualitative evaluations of the proposed framework against baseline and competitor models reveal its superior performance in generating high-quality, biologically plausible images with minimized sharp transitions in grayscale and enhanced structural fidelity. The remarkable consistency observed in the generated images across both datasets highlights the model's ability to learn and replicate complex anatomical features, ensuring that the enhanced images remain true to the original data.

The model has been validated quantitatively with metrics such as SSIM, PSNR, and LPIPS. Additionally, the comprehensive qualitative analysis reveals that the images generated by our method exhibit an extraordinary fidelity to biological structures alongside a noticeable decrease in abrupt grayscale transitions. This aspect is of critical importance in medical imaging, where the precise depiction of anatomical details can directly impact the diagnostic process and subsequent clinical decisions.

From the clinical and operational perspectives, the proposed Latent Diffusion Model addresses critical issues in medical imaging practice by significantly extending the functional lifespan of existing MRI equipment through enhanced spatial and temporal resolutions. This model effectively mitigates the need for frequent hardware upgrades, providing healthcare institutions with an economically viable strategy to maintain diagnostic excellence while optimizing resource use. Additionally, by dramatically reducing MRI acquisition times and consequently patient waiting lists, it directly contributes to improved patient outcomes through more timely diagnoses. Moreover, the compatibility of this advanced image reconstruction technique with cloud-based computing infrastructures offers an attractive solution to overcome limitations posed by legacy on-premise computing systems. Cloud deployment not only provides the scalability and flexibility necessary for computationally demanding tasks inherent to modern AI frameworks, such as PyTorch, but also ensures predictable resource utilization and cost-efficiency. Thus,

the clinical integration of latent diffusion models represents a practical, sustainable advancement, enabling healthcare providers to deliver superior diagnostic imaging capabilities without substantial infrastructural overhaul.

Despite these positive results, certain limitations open the door to further enhancements. While the four datasets used offer diversity in modality and acquisition parameters, future validations should include private clinical data from ongoing hospital collaborations to capture variability typical of real diagnostic cases.

5.1. Future work

The successful implementation of this method heralds new possibilities for research and practical applications in sectors where image resolution acts as a critical bottleneck. In the context of medical imaging, for example, our approach can significantly enhance diagnostic accuracy by providing healthcare professionals with images of superior resolution, thereby enabling the identification of minute details crucial for detecting abnormalities. However, there are still many possible avenues for future work, such as modeling median noise (Thurnhofer-Hemsi, López-Rubio, Roé-Vellvé, & Deka, 2020a) and Rician noise (Maza-Quiroga, Thurnhofer-Hemsi, López-Rodríguez, & López-Rubio, 2021), frequent in MRI data, in order to minimize noise artifacts. Another avenue for future research is the optimization of the training process, for example studying the application of progressive learning techniques to streamline the training process and thus improve the resulting topology of the latent space, scheduling across training time the resolution of the spatial features in the training images (Wang et al., 2024b) and the number of layers to be trained (Li et al., 2024, 2022). Our proposal may also be applied to other types of volumetric medical data, such as 3D ultrasound imaging or the various kinds of Computed Tomography scans. More generally, it may be applied not only to volumetric data but also to medical 2D video sequences by interpolating 2D video frames along the time domain, for example in videofluoroscopy tests.

CRediT authorship contribution statement

Jorge Andrés Mármol-Rivera: Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization; **José David Fernández-Rodríguez:** Methodology, Writing - original draft, Writing - review & editing, Supervision; **Beatriz Asenjo:** Data curation; **Ezequiel López-Rubio:** Methodology, Writing - original draft, Writing - review & editing, Supervision.

Data availability

The datasets used in this work are available at:

- https://fcon_1000-projects-nitrc-org.translate.goog/indi/retro/atlas.html?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es-419&_x_tr_pto=sc
- https://www.oasis-brains.org/
- https://openneuro.org/datasets/ds004199/versions/1.0.5
- https://www.cancerimagingarchive.net/analysis-result/rsna-asnr-miccai-brats-2021/

The source code is available in https://github.com/icai-uma/Latent-diffusion-for-arbitrary-zoom-MRI-super-resolution..

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ezequiel Lopez-Rubio reports financial support was provided by Autonomous Government of Andalusia. Ezequiel Lopez-Rubio, Jose David Fernandez-Rodriguez reports financial support was provided by Spain Ministry of Science and Innovation. Ezequiel Lopez-Rubio reports financial support was provided by University of Malaga. Ezequiel Lopez-Rubio, Jose David Fernandez-Rodriguez reports financial support was provided by Fundacion Unicaja. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is partially supported by the Autonomous Government of Andalusia (Spain) under project UMA20-FEDERJA-108, project name Detection, characterization and prognosis value of the non-obstructive coronary disease with deep learning, and also by the Ministry of Science and Innovation of Spain, grant number PID2022-136764OA-I00, project name Automated Detection of Non Lesional Focal Epilepsy by Probabilistic Diffusion Deep Neural Models. It includes funds from the European Regional Development Fund (ERDF). It is also partially supported by the University of Málaga (Spain) under grants B1-2021_20, project name Detection of coronary stenosis using deep learning applied to coronary angiography; B4-2022, project name Intelligent Clinical Decision Support System for Non-Obstructive Coronary Artery Disease in Coronarographies; B1-2022_14, project name Detección de trayectorias anómalas de vehículos en cámaras de tráfico; and, by the Fundación Unicaja under project PUNI-003_2023, project name Intelligent System to Help the Clinical Diagnosis of Non-Obstructive Coronary Artery Disease in Coronary Angiography. The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the SCBI (Supercomputing and Bioinformatics) center of the University of Málaga. They also gratefully acknowledge the support of NVIDIA Corporation with the donation of an RTX A6000 GPU with 48Gb. The authors also thankfully acknowledge the grant of the Universidad de Málaga and the Instituto de Investigación Biomédica de Málaga y Plataforma en Nanomedicina-IBIMA Plataforma BIONAND.

References

Ahn, S., Menini, A., McKinnon, G., Gray, E., Trzasko, J., Huston, J., Bernstein, M., Costello, J., Foo, T., & Hardy, C. (2020). Contrast-weighted SSIM loss function for deep learning-based undersampled MRI reconstruction. In *International society for magnetic resonance in medicine 28th annual meeting*.

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S. et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., & Davatzikos, C. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, *4*(1), 1–13.

Bank, D., Koenigstein, N., & Giryes, R. (2023). Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook*, (pp. 353–374).

Brix, M. A. K., Järvinen, J., Bode, M. K., Nevalainen, M., Nikki, M., Niinimäki, J., & Lammentausta, E. (2024). Financial impact of incorporating deep learning reconstruction into magnetic resonance imaging routine. *European Journal of Radiology*, *175*, 111434.

Chatterjee, N., Duda, J., Gee, J., Elahi, A., Martin, K., Doan, V., Liu, H., Maclean, M., Rader, D., Borthakur, A. et al. (2025). A cloud-based system for automated ai image analysis and reporting. *Journal of Imaging Informatics in Medicine*, *38*(1), 368–379.

Chaudhari, A. S., Fang, Z., Kogan, F., Wood, J., Stevens, K. J., Gibbons, E. K., Lee, J. H., Gold, G. E., & Hargreaves, B. A. (2018). Super-resolution musculoskeletal MRI using deep learning. *Magnetic Resonance in Medicine*, *80*(5), 2139–2154.

Chen, X., Pawlowski, N., Rajchl, M., Glocker, B., & Konukoglu, E. (2018). Deep generative models in the real-world: An open challenge from medical imaging. *arXiv preprint arXiv:1806.05452*.

Chitradevi, B., & Srimathi, P. (2014). An overview on image processing techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, *2*(11), 6466–6472.

Chung, H., Lee, E. S., & Ye, J. C. (2022). MR Image denoising and super-resolution using regularized reverse diffusion. *IEEE Transactions on Medical Imaging*, *42*(4), 922–934.

Croitoru, F.-A., Hondru, V., Ionescu, R. T., & Shah, M. (2023). Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *45*(9), 10850–10869.

Del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2018). An optimal transportation approach for assessing almost stochastic order. *The Mathematics of the Uncertain: A Tribute to Pedro Gil*, (pp. 33–44).

Dimitriadis, A., Trivizakis, E., Papanikolaou, N., Tsiknakis, M., & Marias, K. (2022). Enhancing cancer differentiation with synthetic MRI examinations via generative models: A systematic review. *Insights into Imaging, 13*(1), 188.

Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., & Jia, Y. (2020). Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing*, *392*, 209–220.

Du, J., Wang, L., Gholipour, A., He, Z., & Jia, Y. (2018). Accelerated super-resolution MR image reconstruction via a 3D densely connected deep convolutional neural network. In *2018 IEEE International conference on bioinformatics and biomedicine (BIBM)* (pp. 349–355). IEEE.

European Society of Radiology (ESR) (2014). Renewal of radiological equipment. *Insights into Imaging*, *5*, 543–546.

Farooq, M. A., Abaid, A., Ullah, I., & Corcoran, P. (2024). A comparative study on diffusion sampling methods across diverse medical imaging modalities. In *Proceedings of the Asian conference on computer vision* (pp. 193–206).

Forigua, C., Escobar, M., & Arbelaez, P. (2022). SuperFormer: Volumetric transformer architectures for MRI super-resolution. In *International workshop on simulation and synthesis in medical imaging* (pp. 132–141). Springer.

Gallagher, A. C. (2005). Detection of linear and cubic interpolation in JPEG compressed images. In *The 2nd Canadian conference on computer and robot vision (CRV'05)* (pp. 65–72). IEEE.

Guerreiro, J., Tomás, P., Garcia, N., & Aidos, H. (2023). Super-resolution of magnetic resonance images using generative adversarial networks. *Computerized Medical Imaging and Graphics*, *108*, 102280.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

He, W., Hu, Y., Wang, L., He, Z., & Du, J. (2021). Gating feature dense network for single anisotropic MR image super-resolution. In *ICASSP 2021 - 2021 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 1670–1674). https://doi.org/10.1109/ICASSP39728.2021.9414646

Herman, G. T., Rowland, S. W., & Yau, M.-m. (1979). A comparative study of the use of linear and modified cubic spline interpolation for image reconstruction. *IEEE Transactions on Nuclear Science*, *26*(2), 2879–2894.

Hofmann, B., Brandsaeter, I. Ø., & Kjelle, E. (2023). Variations in wait times for imaging services: A register-based study of self-reported wait times for specific examinations in norway. *BMC Health Services Research*, *23*(1), 1287.

ICRP PUBLICATION 154 (2023). Optimisation of radiological protection in digital radiology techniques for medical imaging. *Annals of the ICRP*, *52*(3), 11–145.

Ji, Z., Zou, B., Kui, X., Liu, J., Zhao, W., Zhu, C., Dai, P., & Dai, Y. (2024). Deep learning-based magnetic resonance image super-resolution: A survey. *Neural Computing and Applications*, *36*(21), 12725–12752.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F. et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122–1131.

Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S. et al. (2023). Denoising diffusion probabilistic models for 3D medical image generation. *Scientific Reports*, *13*(1), 7303.

Khalid, S., Goldenberg, M., Grantcharov, T., Taati, B., & Rudzicz, F. (2020). Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Network Open*, *3*(3), e201664.

Kim, H., Choi, I., Kim, D. S., & Shin, C. W. (2024). Single image super resolution on dynamic x-ray radiography based on a vision transformer. In *Medical imaging 2024: Image processing* (pp. 733–741). SPIE (*vol. 12926*).

Kim, J., & Park, H. (2024). Adaptive latent diffusion model for 3D medical image to image translation: Multi-modal magnetic resonance imaging study. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 7604–7613).

Kong, Z., & Ping, W. (2021). On fast sampling of diffusion probabilistic models. *arXiv preprint arXiv:2106.00132*.

Li, C., Zhang, J., Lin, S., Yang, Z., Liang, J., Liang, X., & Chang, X. (2024). Efficient training of large vision models via advanced automated progressive learning. arXiv preprint arXiv:2410.00350. Retrieved from https://arxiv.org/abs/2410.00350

Li, C., Zhuang, B., Wang, G., Liang, X., Chang, X., & Yang, Y. (2022). Automated progressive learning for efficient training of vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12486–12496).

Liefaard, M. C., Lips, E. H., Wesseling, J., Hylton, N. M., Lou, B., Mansi, T., & Pusztai, L. (2021). The way of the future: Personalizing treatment plans through technology. *American Society of Clinical Oncology Educational Book*, *41*, 12–23.

Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N. et al. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific Data*, *5*(1), 1–11.

Lozano, I. F., Osinalde, E. P., Bolao, I. G., Pineda, S. O., Padial, L. R., & Romo, A. Í. (2018). Criteria for the management of technological assets in cardiovascular imaging. *Revista Española de Cardiología (English Edition)*, *71*(8), 643–655.

Lyu, Q., You, C., Shan, H., & Wang, G. (2018). Super-resolution MRI through deep learning. *arXiv preprint arXiv:1810.06776*.

Madhav, S., Nandhika, T. M., & Devi, M. K. K. (2024). Super resolution of medical images using SRGAN. In *2024 Second international conference on emerging trends in information technology and engineering (ICETITE)* (pp. 1–6). IEEE.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, *19*(9), 1498–1507.

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncalves, M. et al. (2021). OpenNeuro: An open resource for sharing of neuroimaging data. *bioRxiv*. https://doi.org/10.1101/2021.06.28.450168.

Matsubara, T., Tashiro, T., & Uehara, K. (2019). Deep neural generative model of functional MRI images for psychiatric disorder diagnosis. *IEEE Transactions on Biomedical Engineering*, *66*(10), 2768–2779.

Maza-Quiroga, R., Thurnhofer-Hemsi, K., López-Rodríguez, D., & López-Rubio, E. (2021). Rician noise estimation for 3D magnetic resonance images based on Benford's law. In *International conference on medical image computing and computer-assisted intervention* (pp. 340–349). Springer.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024.

Michelis, M. Y., & Becker, Q. (2021). On linear interpolation in the latent space of deep generative models. *arXiv preprint arXiv:2105.03663*.

Michelucci, U. (2022). An introduction to autoencoders. *arXiv preprint arXiv:2201.03898*.

Nam, G., Khlifi, M., Rodriguez, A., Tono, A., Zhou, L., & Guerrero, P. (2022). 3D-LDM: Neural implicit 3D shape generation with latent diffusion models. *arXiv preprint arXiv:2212.00842*.

Ota, J., Umehara, K., Kershaw, J., Kishimoto, R., Hirano, Y., Tachibana, Y., Ohba, H., & Obata, T. (2022). Super-resolution generative adversarial networks with static T2* WI-based subject-specific learning to improve spatial difference sensitivity in fMRI activation. *Scientific reports*, *12*(1), 10319.

Pati, A., & Lerch, A. (2021). Attribute-based regularization of latent spaces for variational auto-encoders. *Neural Computing and Applications*, *33*, 4429–4444.

Pinaya, W. H. L., Tudosiu, P.-D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., Ourselin, S., & Cardoso, M. J. (2022). Brain imaging generation with latent diffusion models. In *MICCAI workshop on deep generative models* (pp. 117–126). Springer.

Qiu, D., Zhang, S., Liu, Y., Zhu, J., & Zheng, L. (2020). Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Computer Methods and Programs in Biomedicine*, *187*, 105059. https://doi.org/10.1016/j.cmpb.2019.105059

Rafailov, R., Yu, T., Rajeswaran, A., & Finn, C. (2021). Offline reinforcement learning from images with latent space models. In *Learning for dynamics and control* (pp. 1154–1168). PMLR.

Razavi, A., Van den Oord, A., & Vinyals, O. (2019). Generating diverse high-fidelity images with VQ-VAE-2. *Advances in Neural Information Processing Systems*, *32*, 14837–14847.

Ren, F., & Zhou, Y. (2020). CGMVQA: A new classification and generative model for medical visual question answering. *IEEE Access*, *8*, 50626–50636.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684–10695).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18* (pp. 234–241). Springer.

Sander, J., de Vos, B. D., & Išgum, I. (2022). Autoencoding low-resolution MRI for semantically smooth interpolation of anisotropic MRI. *Medical Image Analysis*, *78*, 102393.

Shi, J., Liu, Q., Wang, C., Zhang, Q., Ying, S., & Xu, H. (2018). Super-resolution reconstruction of MR image with a novel residual learning network algorithm. *Physics in Medicine & Biology*, *63*(8), 085011.

Singer, C., Boldor, N., Vaknin, S., Olmer, L., Wilf-Miron, R., & Myers, V. (2025). Scheduling an appointment for MRI: Patient perception of wait time and difficulty. *Israel Journal of Health Policy Research*, *14*, 14.

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Song, Y., Wang, H., Frøseth, G., Nåvik, P., Liu, Z., & Rønnquist, A. (2023). Surrogate modelling of railway pantograph-catenary interaction using deep long-short-term-memory neural networks. *Mechanism and Machine Theory*, *187*, 105386.

Sun, J., Zhu, J., & Tappen, M. F. (2010). Context-constrained hallucination for image super-resolution. In *2010 IEEE Computer society conference on computer vision and pattern recognition* (pp. 231–238). IEEE.

Tan, C., Zhu, J., & Lio', P. (2020). Arbitrary scale super-resolution for brain MRI images. In *IFIP International conference on artificial intelligence applications and innovations* (pp. 165–176). Springer.

Tang, W., & Zhao, H. (2024). Score-based diffusion models via stochastic differential equations–a technical tutorial. *arXiv preprint arXiv:2402.07487*.

Thurnhofer-Hemsi, K., López-Rubio, E., Roé-Vellvé, N., & Deka, L. (2020a). Super-resolution of 3D MRI corrupted by heavy noise with the median filter transform. In *2020 IEEE International conference on image processing (ICIP)* (pp. 3015–3019). IEEE.

Thurnhofer-Hemsi, K., López-Rubio, E., Domínguez, E., Luque-Baena, R. M., & Roé-Vellvé, N. (2020b). Deep learning-based super-resolution of 3D magnetic resonance images by regularly spaced shifting. *Neurocomputing*, *398*, 314–327. https://doi.org/10.1016/j.neucom.2019.05.107

Ulmer, D., Hardmeier, C., & Frellsen, J. (2022). Deep-significance-easy and meaningful statistical significance testing in the age of neural networks. *arXiv preprint 2204.06815*.

Venu, D. N. (2023). PSNR based evalution of spatial Guassian kernals for FCM algorithm with mean and median filtering based denoising for MRI segmentation. *IJFANS International Journal of Food and Nutritional Sciences*, *12*(1), 928–939.

Wang, J., Levman, J., Pinaya, W. H. L., Tudosiu, P.-D., Cardoso, M. J., & Marinescu, R. (2023). InverseSR: 3D brain MRI super-resolution using a latent diffusion model. In *International conference on medical image computing and computer-assisted intervention* (pp. 438–447). Springer.

Wang, L., Zhu, H., He, Z., Jia, Y., & Du, J. (2022). Adjacent slices feature transformer network for single anisotropic 3D brain MRI image super-resolution. *Biomedical Signal Processing and Control*, *72*, 103339.

Wang, X., Yan, J.-K., Cai, J.-Y., Deng, J.-H., Qin, Q., & Cheng, Y. (2024a). Super-resolution reconstruction of single image for latent features. *Computational Visual Media*, *10*(6), 1219–1239. https://doi.org/10.1007/s41095-023-0387-8

Wang, Y., Yue, Y., Lu, R., Han, Y., Song, S., & Huang, G. (2024b). EfficientTrain + +: Generalized curriculum learning for efficient visual backbone training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *46*, 8036–8055.

Wang, Z., Chen, J., & Hoi, S. C. H. (2020). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(10), 3365–3387.

Wu, Z., Chen, X., Xie, S., Shen, J., & Zeng, Y. (2023). Super-resolution of brain MRI images based on denoising diffusion probabilistic model. *Biomedical Signal Processing and Control*, *85*, 104901.

Xing, X., Li, L., Sun, M., Yang, J., Zhu, X., Peng, F., Du, J., & Feng, Y. (2024). Deep-learning-based 3D super-resolution CT radiomics model: Predict the possibility of the micropapillary/solid component of lung adenocarcinoma. *Heliyon*, *10*(13),e34163.

Yan, J., Wang, Q., Cheng, Y., Su, Z., Zhang, F., Zhong, M., Liu, L., Jin, B., & Zhang, W. (2024). Optimized single-image super-resolution reconstruction: A multimodal approach based on reversible guidance and cyclical knowledge distillation. *Engineering Applications of Artificial Intelligence*, *133*, 108496.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, *21*(12), 3106–3121.

Yoshimura, T., Nishioka, K., Hashimoto, T., Mori, T., Kogame, S., Seki, K., Sugimori, H., Yamashina, H., Nomura, Y., Kato, F. et al. (2023). Prostatic urinary tract visualization with super-resolution deep learning models. *PloS one*, *18*(1), e0280076.

Zhang, K., Liang, J., Van Gool, L., & Timofte, R. (2021a). Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4791–4800).

Zhang, Y., Li, K., Li, K., & Fu, Y. (2021b). MR Image super-resolution with squeeze and excitation reasoning attention network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13425–13434).

Zhao, C., Shao, M., Carass, A., Li, H., Dewey, B. E., Ellingsen, L. M., Woo, J., Guttman, M. A., Blitz, A. M., Stone, M. et al. (2019). Applications of a deep learning method for anti-aliasing and super-resolution in MRI. *Magnetic Resonance Imaging*, *64*, 132–141.

Zhou, Y., Qian, C., Li, J., Wang, Z., Hu, Y., Qu, B., Zhu, L., Zhou, J., Kang, T., Lin, J. et al. (2024). CloudBrain-ReconAI: A cloud computing platform for MRI reconstruction and radiologists' image quality evaluation. *IEEE Transactions on Cloud Computing*. *12*(4), 1359–1371.