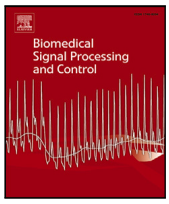
Biomedical Signal Processing and Control 120 (2026) 110037

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Two-step latent diffusion modelling for morphology-guided synthesis of glioma intraoperative ultrasound images

Angelo Lasala [a,*], Maria Chiara Fiorentino [b], Andrea Bandini [c,a], Sara Moccia [d], Stamatia Giannarou [e]

[a] The BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Pisa, Italy
[b] Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy
[c] Health Science Interdisciplinary Research Center, Scuola Superiore Sant'Anna, Pisa, Italy
[d] Department of Innovative Technologies in Medicine and Dentistry Università degli Studi "G. d'Annunzio" Chieti-Pescara, Chieti-Pescara, Italy
[e] Hamlyn Centre for Robotic Surgery, Department of Surgery and Cancer, Imperial College London, London, United Kingdom

ARTICLE INFO

Keywords:
Intraoperative ultrasound
Generative AI
Latent diffusion model
Brain tumour segmentation

ABSTRACT

Intraoperative ultrasound (iUS) is increasingly used in neurosurgery to monitor tumour margins during resection. The adoption of iUS is still limited by low image quality, noise, and heterogeneous echogenicity, which makes surgeons' interpretation of surgical margins challenging. While deep learning can aid automatic margin delineation, the lack of annotated datasets limits the development of robust methods. To address this challenge, we propose a two-step generative framework based on latent diffusion models that consist of (i) an unconditional tumour-mask generator that learns geometric features of real tumours, and (ii) a conditional iUS image generator that synthesizes realistic iUS images by using the generated tumour masks as a prior. Morphological fidelity is assessed through tailored quantitative and qualitative metrics. The performance of automatic tumour margin segmentation algorithms is evaluated through data augmentation experiments to determine whether the inclusion of synthetic data can improve segmentation performance. Compared to state-of-the-art conditional generative models, including diffusion-based approaches (ControlNet) and generative adversarial networks (Pix2Pix), the proposed framework achieves superior qualitative and quantitative performance in representing tumoural and non-tumoural tissue. Performance evaluated using a 5-fold cross-validation protocol yields statistically significant improvements in morphological fidelity (Dice Similarity Coefficient: 0.851; Hausdorff Distance: 16.21). The analysis shows that introducing synthetic data significantly improves boundary delineation performance using nn-UNet, reducing the average Hausdorff Distance from 33.97 to 30.72 in the test set. These results indicate that the proposed framework helps mitigate the scarcity of annotated iUS data by providing realistic samples to support training in neurosurgical image segmentation.

1. Introduction

Malignant gliomas account for nearly 80% of primary malignant brain tumours and diffuse infiltration [1,2]. Surgical resection remains a cornerstone of treatment, but its success depends on the accurate identification of tumour margins to maximize resection while preserving function [3,4]. To support this delicate balance, intraoperative magnetic resonance imaging (MRI) and intraoperative ultrasound (US) are increasingly adopted to visualize tumour margins in real time and monitor resection progression [5]. While intraoperative MRI (iMRI) provides high-quality and easily interpretable images [6], its widespread use is constrained by high costs, technical complexity, and substantial infrastructural requirements. In contrast, intraoperative ultrasound (iUS) has emerged as a valuable alternative, thanks to its portability, affordability, and compatibility with standard surgical instruments, thereby offering a practical solution for real-time intraoperative guidance (Fig. 1) [7].

Currently, iUS interpretation remains challenging because image quality is frequently affected by artefacts, noise, and heterogeneous echogenicity, which complicate the delineation of glioma margins from surrounding functional tissue [8]. As a result, iUS remains highly operator-dependent and subject to substantial variability in interpretation [9], crucial factors that have led surgeons to favour iMRI over iUS in the last decade [7].

* Corresponding author.
E-mail address: angelo.lasala@unich.it (A. Lasala).

https://doi.org/10.1016/j.bspc.2026.110037
Received 5 November 2025; Received in revised form 10 February 2026; Accepted 1 March 2026

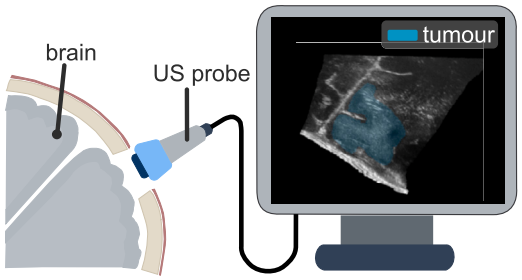


Fig. 1. Intraoperative ultrasound acquisition during neurosurgery. The ultrasound probe is placed directly on the exposed cortical surface during open brain surgery to acquire intraoperative ultrasound (iUS) images.

Deep learning (DL) has emerged as a promising strategy to support iUS interpretation by reducing operator dependence and assisting in intraoperative detection of tumour boundaries. The recent availability of publicly accessible datasets [10–12] has accelerated the development of DL models for tasks such as segmentation of glioma boundaries [13] and of the resection cavity [14]. Yet, collecting large-scale annotated iUS data remains highly challenging due to the complexity of the surgical setting in the operating room, variability in acquisition protocols, and the scarcity of expert annotators [8]. This persistent lack of annotated iUS data is a critical bottleneck, limiting the generalizability and robustness of current DL approaches and their translation into actual surgical practice.

Generative AI has recently emerged as a compelling strategy to overcome this limitation by producing realistic synthetic medical data. In neurosurgical imaging, this has fuelled growing interest in multimodal MRI-US frameworks [15–18]. On the contrary, applications based solely on iUS data remain comparatively underexplored, likely because surgeons have only recently begun to consider relying on iUS to reliably delineate tumour margins [7]. Donnez et al. [19] shows the feasibility of an entirely MRI-free workflow by training a generative adversarial network (GAN) to synthesize resection images directly from iUS, which are used to simulate realistic post-resection scenarios and assist surgeons in interpreting intraoperative findings. In medical image synthesis, latent diffusion models (LDMs) have shown better performance than GANs, offering unprecedented opportunities for high-fidelity medical image synthesis [20]. Yet, their potential within iUS remains largely untapped, representing a promising avenue towards advancing MRI-free surgical guidance. The direct application of generative models to US medical imaging is not straightforward, as the generative process must guarantee robust anatomical fidelity [21,22]. In neurosurgery, this challenge translates into the need for models capable of synthesizing iUS with a faithful representation of gliomas. Moreover, the intrinsic morphological complexity of gliomas poses additional challenges in developing robust pipelines for the generation of synthetic image–mask pairs, which can be reliably leveraged for data augmentations to improve the performance of tumour segmentation models.

To address these challenges, we introduce a LDM framework for synthesizing iUS images of gliomas, with the primary goal of achieving morphologically accurate and clinically faithful representations. To the best of our knowledge, this direction remains unexplored in prior research. To focus on the most informative tumour regions while preserving intra-patient variability, we define a preprocessing step. For each patient, we select the slice containing the largest tumour area and define a spatial window around it. This slice is used as a starting point to sample slices at regular intervals. The framework is organized into two stages: first, an LDM is trained to capture the variability in size and contour geometry of real tumour masks corresponding to glioma regions; second, a conditional LDM generates iUS images using these masks as geometric priors. This two-step design ensures that the synthetic image–mask pairs preserve tumour morphology and variability, ultimately providing high-quality data to enrich downstream segmentation tasks through augmentation. Within the early application of diffusion models to iUS, our study includes a rigorous performance assessment, which remains an open challenge in the field of medical image analysis [23]. We evaluate both the anatomical plausibility of the synthesized data and its effectiveness in downstream segmentation of tumour margins. Thereby contributing to the broader investigation on how the proposed frameworks can be considered a reliable tool for surgical data augmentation.

The contribution of this study can be summarized as follow:

- For the first time, diffusion models are used to synthesize iUS images. We introduce a two-step generative framework based on LDMs to synthesize morphologically faithful iUS images and their corresponding tumour masks. To represent both tumour morphology and the echogenic properties of brain tissue, the framework is structured such that the first LDM captures the distribution of realistic glioma morphologies, while the second conditional LDM translates this geometric prior into synthetic iUS images. This design enables us to generate paired image–mask datasets that can be used directly to train automatic segmentation models.
- We conduct a comprehensive performance evaluation of state-of-the-art generative strategies. First, we compare conditional LDMs with a Pix2Pix GAN baseline to investigate whether diffusion models provide superior performance in iUS image synthesis, as Pix2Pix represents the closest work to ours reported in the literature [19]. Second, we compare the proposed framework with a one-step approach that simultaneously synthesizes image–mask pairs using a unified LDM, in order to assess whether decoupling the generation of tumour masks and iUS images leads to better generated data.
- We propose novel quantitative and qualitative metrics tailored to evaluate the morphological fidelity of the generated iUS data. Furthermore, we analyse the impact of synthetic data on downstream glioma segmentation tasks, showing its potential clinical utility in improving DL model performance for tumour delineation.

2. Related work

In recent years, generative models have gained particular attention in neurosurgical image analysis, especially for enabling the integration of iUS and iMRI data. Dorent et al. [15] propose a Variational Autoencoder (VAE) that jointly synthesizes paired iUS and iMRI images. By leveraging a hierarchical latent space representation and an enhanced posterior parameterization in the variational loss, their model can generate both modalities from an incomplete input set, addressing the frequent challenge in neurosurgical workflows where simultaneous acquisition of iUS and iMRI is not always feasible. Similarly, Singh et al. [16] develop a 3D Pix2Pix framework to generate iMRI volumes directly from corresponding 3D brain iUS scans. More recently, Eker et al. [17] introduce BrainPixGAN, a GAN-based architecture designed to generate iMRI from combined pre-operative MRI and iUS inputs. To compensate the brain shift between pre-operative and intraoperative acquisition, Rahmani et al. [18] propose D2BGAN, an unsupervised Bayesian GAN model for preoperative MR-iUS registration task

While these multimodal approaches have shown promising results, there is a growing shift towards US-only workflows, driven by the high cost, technical complexity, and magnetic constraints of MRI systems. In this context, Donnez et al. [19] propose a two-stage pipeline to simulate post-resection iUS images from pre-resection iUS and a resection cavity mask. In the first stage, a Pix2Pix conditional GAN [24] generates a realistic iUS image containing the resection cavity, using the cavity mask as a prior. In the second stage, the generated cavity is blended into the corresponding real pre-resection iUS scan, ensuring

anatomical consistency by embedding the synthetic cavity within the patient's actual brain structures. This process results in more realistic post-resection iUS images suitable for surgical simulation and training.

Despite the promising results of GAN-based models, LDMs have recently emerged as a more powerful generative paradigm, often surpassing GANs in medical imaging tasks [20]. Building upon this transition from GAN-based approaches, recent research has increasingly focused on diffusion-based generative frameworks for medical image synthesis. These approaches have rapidly emerged as a dominant research direction, as shown by a growing body of recent work in medical imaging. LDMs have been successfully applied across a wide range of imaging modalities, showing improved performances in capturing complex anatomical variability and morphological details [25–29]. Recent applications to ultrasound imaging further support the suitability of LDMs for handling anatomical variability. In particular, conditional diffusion frameworks guided by explicit structural priors have been shown to effectively control lesion geometry and spatial context during ultrasound synthesis. For instance, LDMs conditioned on lesion anatomy have been successfully applied to lung ultrasound to increase the diversity and prevalence of rare pathological patterns [30], while geometry-guided latent diffusion approaches have shown improved anatomical consistency in echocardiographic image synthesis task [22]. Several studies have shown that LDMs maintain high performance even in low-data regimes, making them particularly suitable for medical domains where annotated datasets are limited, such as in endoscopic images [31], fundus photographs [32], and fluorescein angiography scans [33]. In the neurosurgical domain, Kebaili et al. [34] introduce a slice-based LDM architecture to address the challenges of volumetric MRI brain data by operating in a slice-by-slice manner. Their framework employs a 2D VAE to encode MRI slices and their corresponding tumour masks into a latent space, followed by a diffusion model that learns to sample from the joint latent representation. This design enables simultaneous generation of MRI data and tumour masks, effectively capturing their joint distribution even in data-scarce scenarios.

Although the field has matured to the point where LDM-based techniques are increasingly recognized as state-of-the-art in medical image synthesis, their adoption within the iUS domain remains hindered by the inherent variability of images, which often hampers the accurate visualization of tumour boundaries. As a result, an LDM specifically designed for iUS generation without relying on MRI data has not yet been explored. Building on previous works in multimodal synthesis and GAN-based iUS generation, we propose a diffusion-based approach that focuses solely on morphologically faithful iUS image generation.

3. Materials and methods

3.1. Two-step generation framework based on latent diffusion model

Fig. 2 shows the proposed framework. Generating images from ground truth segmentation maps inherently restricts diversity, as it confines the synthesis to the anatomical variability present in existing annotations, limiting their usefulness for data augmentation. In the field of generative models for medical image analysis, recent work [35] addressed this limitation through a novel pipeline in which segmentation maps are no longer tied to existing annotations, but are instead generated by a diffusion model and subsequently used as priors for synthesizing surgical scenes. Inspired by these approach, we design a modular framework composed of two complementary LDM-based models for synthesizing iUS images and their corresponding tumour masks. The rationale for the two-step approach is to disentangle mask generation from iUS synthesis, enabling each model to specialize in complementary aspects of the data. Our hypothesis is that this separation improves control over anatomical variability and supports the coherent and realistic generation of image–mask that can be directly used as synthetic data for augmenting tumour segmentation models.

Step I - Mask generator. In the first step, a *mask generator* (bottom left panel Fig. 2) based on an unconditional LDM is trained to reproduce the distribution of real tumour masks. This allows the model to generate synthetic masks that preserve the morphological variability and geometric plausibility of brain tumours. In this regard, the proposed LDM operates by first encoding the input mask (x^m) into a latent representation (z_0^m) through a variational autoencoder (VAE). A forward diffusion process progressively perturbs z_0^m, adding Gaussian noise over T timesteps, yielding z_T^m. The reverse process, parameterized by a time-conditioned UNet ($\epsilon_\theta^m(z_t^m, t)$), iteratively denoises z_T^m back to an estimate $\tilde{z}_0^m$. Finally, the VAE decoder reconstructs the corresponding synthetic mask ($\tilde{x}_{gen}^m$). Training relies on minimizing the following objective:

$$\mathcal{L}_{\text{LDM}} = \mathbb{E}_{z^m=\mathcal{E}^m(x^m),\epsilon\sim\mathcal{N}(0,1),t}\left[|\epsilon - \epsilon_\theta^m(z_t^m, t)|_2^2\right] \quad (1)$$

where $\mathcal{E}^m$ is the encoder of the VAE used to compute the latent representation of x^m, and ϵ is sampled from normal distribution with null mean and unitary standard deviation ($\mathcal{N}(0,1)$).

Step II - iUS generator. Once synthetic tumour masks are generated, in the second step, an *iUS generator* (bottom right panel Fig. 2) is implemented as a conditional LDM that leverages the synthetic masks to guide US synthesis. The proposed LDM design for the iUS generator operates by first encoding the input iUS (x^u) into a latent representation (z_0^u). The forward diffusion process adds Gaussian noise to z_0^u over T timesteps to obtain z_T^u. Then, A time-conditioned UNet ($\epsilon_\theta^u(z_t^u, t)$) then performs the reverse process, iteratively denoising z_T^u to estimate $\tilde{z}_0^u$. Conditioning is integrated into the denoising UNet, yielding the objective:

$$\mathcal{L}_{\text{condLDM}} = \mathbb{E}_{z^u=\mathcal{E}^u(x^u),\epsilon\sim\mathcal{N}(0,1),t}\left[|\epsilon - \epsilon_\theta^u(z_t^u, t, c(\tilde{x}_{gen}^m))|_2^2\right], \quad (2)$$

where $\mathcal{E}^u$ is the encoder used to compute the latent representation of x^u, and $c(\cdot)$ denotes the embedding function of the conditioning input. In our case, the conditional input corresponds to the mask synthesized by the mask generator ($\tilde{x}_{gen}^m$). In our proposed model the mask embedding is directly concatenated with the noisy latent input. Specifically, the mask is processed through a stack of seven 2D convolutional layers with SiLU activations to produce a feature map aligned with the spatial resolution of the latent iUS representation. This embedding is then fused with the noisy latent input and passed through the denoising UNet, which iteratively reconstructs the denoised latent representation of the iUS image. The denoising UNet relies on the architecture of UNet of stable diffusion (SD) [36], and initialized with pretrained weights of the SD model, originally trained on the large-scale LAION-5B dataset [37] for text-to-image generation. We adapted this model to the iUS domain by extensive finetuning, which updates all the model parameters to fully adapt the generative model to the iUS domain. Since our focus is exclusively on spatial conditioning, the CLIP encoder [38] for text prompts is deactivated by providing an empty textual input.

3.2. Dataset

In this study, we employ the publicly available RESECT dataset [10], which includes 23 clinical cases of low-grade gliomas (WHO Grade II [2]) from adult patients who underwent surgery at St. Olavs University Hospital between 2011 and 2016. For each case, 3D iUS scans were acquired at three distinct stages of the surgical procedure: pre-resection, intra-resection, and post-resection. Representative examples are shown in Fig. 3, highlighting both the infiltrative nature of gliomas and the resulting challenges in achieving accurate segmentation.

The resolution ranges from $(0.14 \times 0.14 \times 0.14)$ mm^3 to $(0.24 \times 0.24 \times 0.24)$ mm^3 depending on the probe types and imaging depth. Tumours were manually segmented following the method proposed by Munkvold et al. [39]. We refer to the original publication [11] for a detailed description of the annotation protocol.

In line with our objective of synthesizing iUS images with gliomas, we restrict the analysis to the pre-resection scans from the RESECT

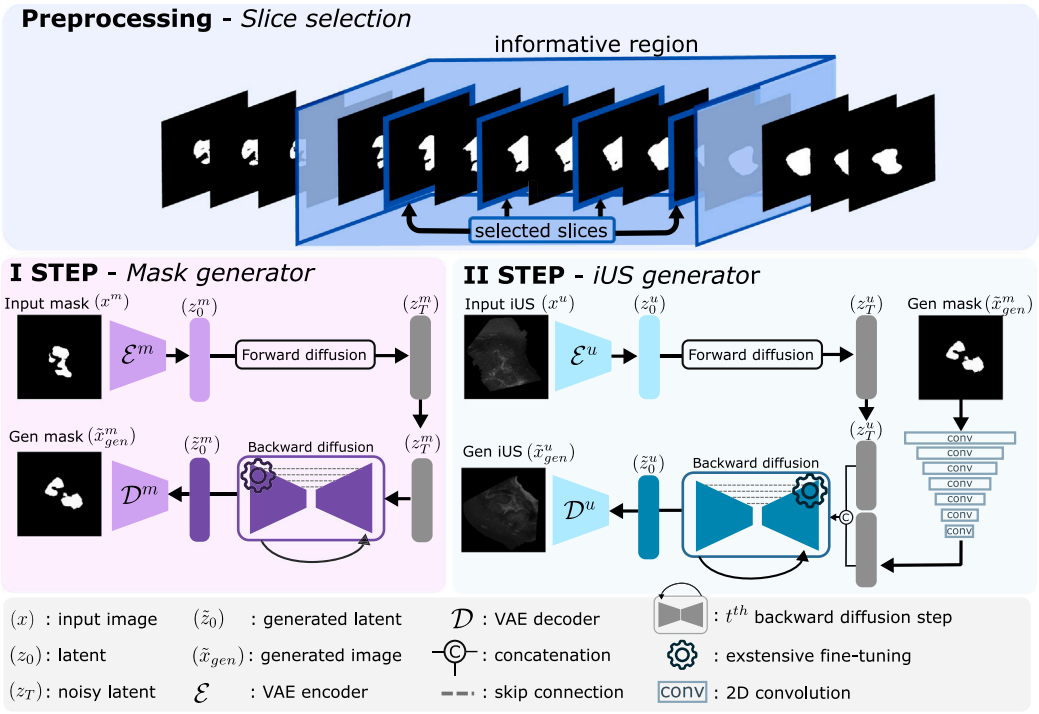


Fig. 2. Overview of the proposed two-step generation framework. (Top panel) Preprocessing — Slice selection: as a preprocessing step before the two-step generation, the slice containing the tumour with the largest surface area is first identified. Informative region is defined as a spatial window covering one-quarter of the scan depth and centred on slice with largest tumour surface. Then, slices are sampled at regular intervals of 1.0 mm the final dataset. (Bottom left panel) I Step — Mask generator: unconditional LDM trained to sample morphologically plausible synthetic tumour masks. (Bottom right panel) II Step – iUS generator: conditional LDM that integrates generated mask as geometrical prior to synthesize realistic iUS images.

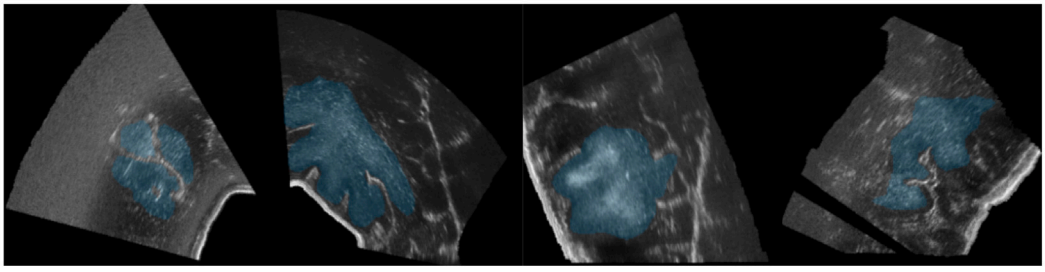


Fig. 3. Examples of annotated data from the RESECT dataset. The infiltrative nature of gliomas (light blue masks), leads to blurred tumour boundaries and blending with surrounding non-tumour tissue, which makes automatic segmentation particularly challenging.

dataset. Following the strategy proposed in recent studies on automatic 2D tumour segmentation in iUS [13], we apply as preprocessing a *slice selection* procedure to focus on the most informative regions while preserving intra-subject variability, as shown in top panel of Fig. 2. For each patient, we first identify the slice containing the tumour with the largest surface area. A spatial window corresponding to one-quarter of the total scan depth is then centred on this slice, defining the informative region. Depending on the subject, this region spanned from 12.0 mm to 23.0 mm. Within this window, slices are sampled at regular intervals of 1.0 mm to construct the final dataset.

To ensure a robust evaluation, we adopted the data splitting strategy proposed by Canalini et al. [14], holding out subjects 24, 25, and 27 as the test set (50 images in total). The remaining subjects are randomly partitioned into five folds using a 90:10 split ratio, ensuring that subjects with the same identification number did not appear in the validation sets across different folds. Table 1 summarizes subject identification numbers and the total number of images for each fold.

3.3. Comparison with alternative strategies

To evaluate the effectiveness of our **proposed** framework described in Section 3.1, we compare it with alternative generative strategies. In line with recent studies on the use of LDMs in medical imaging [31], we incorporate the **ControlNet** model [40] into the two-step framework as an alternative conditional approach to the iUS generator described in Section 3.1. In this setting, the backbone of the reverse diffusion process is a SD model [36] pretrained on the large-scale LAION-5B text–image dataset [37]. To adapt the model to the neurosurgical domain, we first finetune the SD backbone unconditionally on iUS images. Spatial conditioning is then introduced through the ControlNet architecture, which injects tumour mask information into the internal layers of the pretrained SD network. During refinement, the SD backbone is kept frozen, while a trainable copy of the encoder processes the tumour mask input. The encoder outputs are linked to the frozen decoder through zero-initialized 1×1 convolutions, enabling stable learning of spatial conditioning. This design preserves the robustness of the

Table 1
Train and validation splits for each fold. Subject identification numbers are listed along with the total number of images in square brackets. The test set includes subjects 24, 25, and 27, comprising a total of 50 images.

	Train		Val	
Fold 1	1 2 3 4 5 6 7 11 12 13 14 16 17 18 19 21 23 26	[353]	8 15	[35]
Fold 2	1 2 4 5 6 7 8 11 12 13 15 16 17 18 19 21 23 26	[354]	3 14	[34]
Fold 3	1 2 3 4 5 6 7 8 11 12 13 15 16 17 18 21 21 23	[345]	19 26	[43]
Fold 4	1 2 3 5 6 8 11 12 13 14 15 16 17 18 19 21 23 26	[344]	4 7	[44]
Fold 5	1 2 3 4 5 6 7 8 11 12 13 14 15 17 18 19 21 26	[350]	16 23	[38]

pretrained SD backbone while adapting it to iUS synthesis guided by tumour masks.

To benchmark our approach against state-of-the-art methods, we use **Pix2Pix** for translating tumour masks into iUS images, following the original formulation of [24] and its application to iUS synthesis in [19]. For a fair comparison with our framework, inference is performed using the same synthetic tumour masks generated by our mask generator, ensuring that all models are evaluated on an identical input set and allowing a consistent assessment of generative performance.

To further assess our two-step framework, we compare it with a unified alternative (referred to as the **one-step** experiment), inspired by the approach proposed in [34]. In this setting, a single LDM jointly generates iUS images and their corresponding tumour masks, with the goal of learning their joint distribution and preserving spatial coherence between tumour and non-tumour structures. The VAE is provided with the concatenation of an iUS image and its tumour mask to obtain a shared latent representation. The latent diffusion process is then applied unconditionally to this representation, and the VAE decoder reconstructs both the synthetic iUS image and the associated tumour mask from the denoised latent variables.

3.4. Experimental setting

For VAE training, both tumour masks and iUS images are resized to 256 × 256 pixels. Data augmentation is applied on-the-fly, including random rotations between −30° and 30°, random horizontal and vertical translations of up to 25 pixels, and random horizontal flipping. In addition, for iUS images only, we apply brightness, contrast, and gamma adjustments, following the strategy proposed in recent work [13]. The loss function is a combination of pixel-wise MSE, perceptual loss [41], adversarial loss using a PatchGAN discriminator, and the Kullback–Leibler (KL) regularization. As suggested in [36], all components are equally weighted, except for the KL term, which is weakly weighted with a coefficient of 10^{-6}. The weights of VAEs are finetuned of the weight of VAE pretrained on LAION-5B dataset [37]. This training strategy is applied to proposed framework, ControlNet, and the one-step framework. In the one-step approach, the VAE input was the concatenated image–label pair, whereas in proposed framework and ControlNet the input consisted of iUS images only. Training is carried out for 100 epochs using the AdamW optimizer with a learning rate of 10^{-4} and a batch size of 4. The best-performing model was selected based on the lowest validation MSE.

During training of conditional LDMs, we use the classifier-free guidance (CFG) approach [42], where the condition is randomly dropped with a probability of 0.1 for each training sample. This design enables the denosing UNet to learn both conditional and unconditional denoising behaviours, facilitating controlled guidance at inference time. A linear scheduler defines the noise variance, ranging from 0.001 to 0.020 with $T = 1000$, following standard latent diffusion model configurations [36]. The training is performed over about $290k$ steps using AdamW optimizer, with a learning rate of 10^{-5} and a batch size of 32. The best-performing epoch is determined based on the validation Freshet inception distance (FID) score [43]. During the sampling,

During sampling, CFG is applied by combining the conditional and unconditional denoising predictions according to

$$\epsilon_\theta(t) = (1+\omega)\epsilon^u_\theta(z^u_t, t, c(\bar{x}^m_{gen})) - \omega\epsilon^u_\theta(z_t, t, \varnothing) \quad \text{with } t = T, \dots, 0 \tag{3}$$

where the unconditional denoising prediction $\epsilon^u_\theta(z_t, t, \varnothing)$ is obtained by feeding the conditional model with a null mask $\varnothing$. The parameter ω controls the strength of the conditioning and therefore governs the trade-off between sample diversity and adherence to the provided condition. The best value of ω is selected by a grid search within the values of 3.0, 5.0, and 7.0. During the training of unconditional LDMs, ω is set to 0, thereby deactivating the CFG mechanism. All models are implemented using PyTorch, and the experiments are performed using an NVIDIA Ampere A100 GPU.

3.5. Performance assessment

In medical image analysis, the evaluation of generative models remains a challenge, as it must account not only for image quality but also for anatomical fidelity and clinical relevance [21,44,45]. To address this, we define evaluation metrics specifically designed to assess generative models for iUS images with gliomas.

To assess the morphological plausibility of the generated tumour masks, we evaluate features related to both shape and size. Tumour size is used to determine whether the models reproduce the distribution of tumour sizes observed in the real data. For a fair comparison, size is estimated for both real and generated masks as the ratio of tumour area within the resized 256 × 256 binary mask. In addition, the Energy Spectral Distance (ESD) is employed to quantify whether the models captured shape-related characteristics of the tumours. The ESD is defined as:

$$\text{ESD} = \frac{1}{N_f}\sum_{f}^{N_f} \mathcal{F}_f(|d - \bar{d}|) \tag{4}$$

where d is the signal formed by the Euclidean distances between the tumour mask contour points and the centroid, $\bar{d}$ is the mean of this signal, and $\mathcal{F}_f$ denotes the fth component of its Fourier Transform (Fig. 4). Statistical differences between real and generated distributions are assessed using the Wilcoxon rank-sum test. Additionally, we also leveraged the relationship between tumour size (s) and ESD to evaluate the morphological consistency of the mask generator. Linear fit between the logarithm of tumour size and the logarithm of ESD was used to assess whether the model could capture the same size–shape relationship observed in real tumour masks:

$$\log(\text{ESD}) = \beta_0 + \beta_1 \log(s) \tag{5}$$

where β_0 denotes the intercept of the linear model, and β_1 is the slope coefficient that quantifies how ESD scales with tumour size. A t-test is performed to assess whether the β_1 coefficients fitted on generated data are statistically similar to those obtained from the real data.

The quality of synthetic iUS images is quantitatively assessed using the FID, which measures the distance between Gaussian distributions of real and generated image features obtained from InceptionV3, and the Kernel Inception Distance (KID), a similar metric based on polynomial kernel comparisons of InceptionV3 features that is less biased regard dataset size than FID [46]. Additionally, a qualitative inspection is performed to estimate the degree of variability between synthetic and real images. As proposed in [21], pixel-wise square differences (SD) map is performed to reveal the local differences between train data and generated data. This approach enable us to qualitatively assess whether

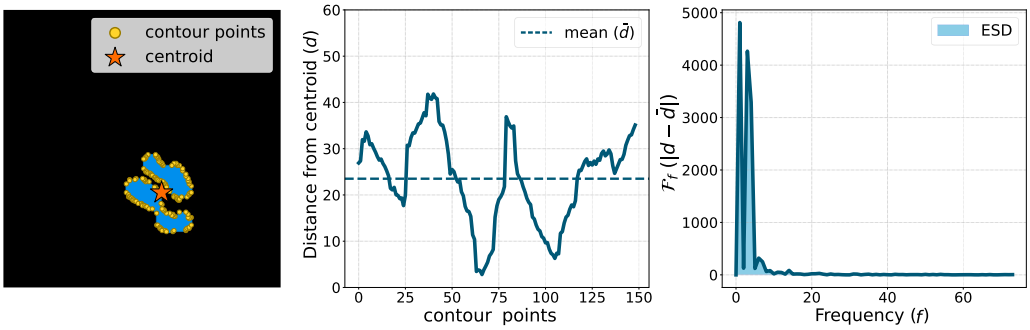


Fig. 4. Energy Spectral Distance (ESD) computation. (Left panel) Contour points (yellow) and centroid (orange) are extracted from the tumour mask. (Middle panel) Radial distances from centroid d are plotted along with mean value $\bar{d}$. (Right panel) Fourier transform of the centred distance signal $|d-\bar{d}|$, with the shaded blue area representing the ESD.

the model can reproduce variability or simple mimicking the training distribution.

Beside image quality metrics, we also define measures to assess morphological fidelity, aiming to quantify how well the synthetic iUS images reflect the underlying tumour geometry specified by the conditional input mask. A high level of morphological fidelity may indicate that the generative model accurately preserves the spatial characteristics of the tumour. We use the MedSAM model [47] to obtain tumour mask in synthetic iUS images. Morphological fidelity is then evaluated by comparing the MedSAM-predicted segmentation with the conditional input mask using the Dice Similarity Coefficient (DSC) and the 95th percentile Hausdorff Distance (HD), capturing both the overlap and boundary accuracy between the two masks. We refer to Sec. 1 in Supplementary Notes for additional details about the using of MedSAM for estimating predicted tumour mask.

To evaluate the clinical utility, we conduct a downstream segmentation experiment to assess whether the inclusion of synthetic data could enhance the performance of automatic model for tumour segmentation. We use the data splitting strategy described in Section 3.2 to train nnUNet [48] models as proposed in most recent work about the automatic segmentation of iUS [13]. We use DSC and HD as evaluation metrics between predicted and ground truth mask. The results on the test set obtained by models trained on real data (N_{real}) serve as baseline performance. Then, we incrementally augment the training set for each fold by adding synthetic data (N_{gen}), resulting in training scenarios where synthetic data constituted 50%, 100%, and 150% of the training dataset. Letting $R = \frac{N_{\text{gen}}}{N_{\text{real}}}$ represents the ratio of synthetic to real training images, the baseline experiment is denoted as $R = 0$, while subsequent experiments are labelled as $R = 1$, $R = 2$, and $R = 3$, corresponding to rescale the size of training set by a factor of two, three, and four, respectively. For details about training and testing of nnUNet, we refer to Sec. 2 in Supplementary Notes. This setup allow us to systematically evaluate how varying the ratio between real and synthetic data influences segmentation performance.

4. Results

Table 2 summarizes the quantitative performance of all experiments across five folds, based on image quality and morphological fidelity metrics. In terms of image quality, the one-step framework achieved the best performance, with the lowest average FID (75.49) and KID (0.039), outperforming Pix2Pix, ControlNet, and the proposed framework. Pix2Pix yielded the poorest image quality score, with the highest average FID (238.68) and KID (0.266). Conversely, morphological fidelity metrics revealed that Pix2Pix, and the proposed framework better preserved the spatial characteristics of tumour structures compared to the one-step. Pix2Pix achieved the highest average DSC (0.864) and the lowest average HD (15.97).

Table 2
Quantitative evaluation across all five folds for each generative model. We report image quality metrics FID (↓) and KID (↓), and morphological fidelity metrics DSC (↑) and HD (↓). Best results for each metric are highlighted in bold.

		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Average
Pix2Pix	FID	246.22	239.33	243.43	249.43	225.74	238.68
	KID	0.274	0.257	0.276	0.278	0.244	0.266
	DSC	**0.862**	**0.869**	**0.867**	**0.868**	**0.856**	**0.864**
	HD	**15.61**	**16.17**	**16.71**	**14.48**	16.91	**15.97**
proposed	FID	98.27	99.38	99.71	98.59	100.57	99.30
	KID	0.056	0.063	0.064	0.057	0.065	0.061
	DSC	0.844	0.856	0.843	0.868	0.842	0.851
	HD	16.00	16.66	17.61	14.61	**16.18**	16.21
ControlNet	FID	88.14	84.23	85.28	91.83	91.08	88.11
	KID	0.044	0.042	0.043	0.048	0.049	0.042
	DSC	0.817	0.824	0.822	0.843	0.816	0.824
	HD	17.91	18.39	18.84	16.16	18.61	17.98
one-step	FID	**82.67**	**71.95**	**71.65**	**74.06**	**77.07**	**75.49**
	KID	**0.047**	**0.037**	**0.035**	**0.035**	**0.040**	**0.039**
	DSC	0.850	0.844	0.819	0.841	0.840	0.839
	HD	16.43	17.78	18.05	17.51	17.28	17.41

Further insights are provided through qualitative evaluation via visual inspection. As shown in Fig. 5, which presents real image–mask pairs used during training alongside outputs generated by the one-step model, the lowest FID and KID scores can be explained by the evidence that the generated images closely replicate the corresponding real ones, introducing only limited variability. In line with visual evidences, the SD maps (right columns) show minimal pixel-wise deviations between real and generated samples.

As for the two-step approaches, we first reported the results of the mask generator (first-step model). Table 3 reports the tumour size and ESD statistics for real and generated tumour masks across all five folds. For each metric, the median and interquartile range are reported. Tumour size distributions between real and generated masks were statistically similar across all folds ($p > 0.05$), with the exception of Fold 3 ($p = 0.04$). For the ESD, four out of five folds showed no significant difference between real and generated distributions. Fold 4 was the only case where a significant difference in ESD is observed ($p = 0.03$). To further assess the ability of the model to reproduce realistic morphological properties, we also analysed the size-shape relationship. We compared the linear fits between the logarithm of tumour size and ESD for real and generated masks across all folds. This comparison provided additional insight into the model's capacity to preserve the intrinsic size-shape dependency found in real tumour distributions. Table 4 reports the estimated values of β_0 and β_1 in Eq. (5) for each fold. A t-test is performed to assess whether the β_1 coefficients fitted on generated data are statistically similar to those obtained from the

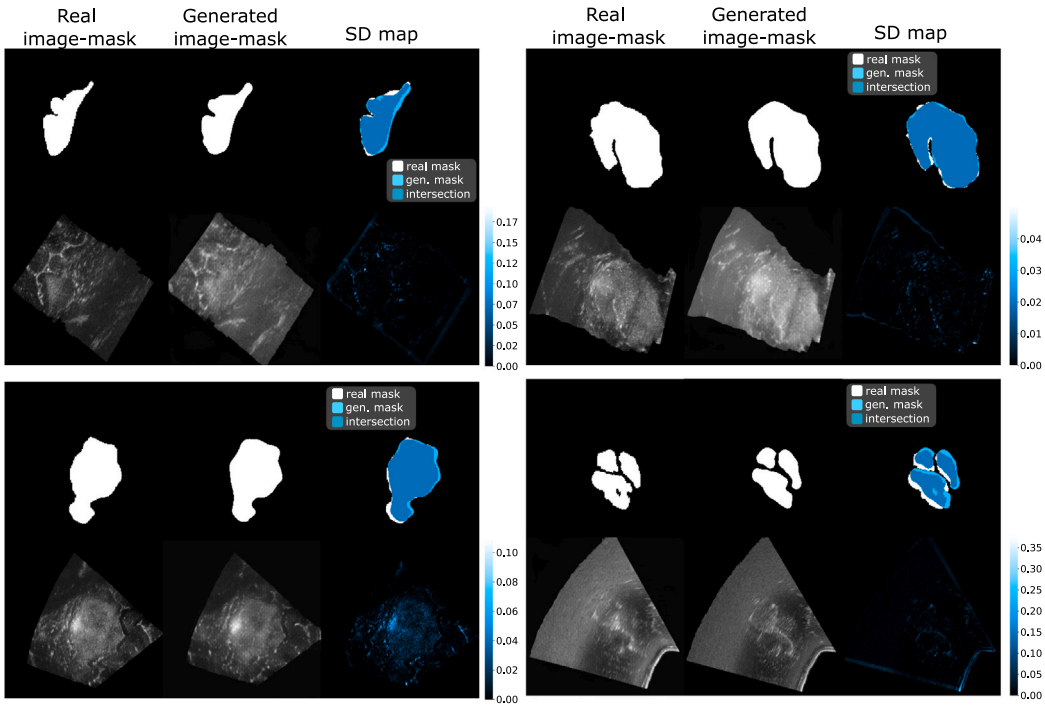


Fig. 5. Qualitative evaluation of one-step approach. Generated image–mask (second column) simply mimics the geometrical shape of real mask and the overall appearance of real iUS (first column), as revealed by SD maps (third column).

Table 3
Quantitative evaluation of the tumour mask generator over five folds. Median and interquartile range, in brackets, of tumour size and ESD are reported for both real and generated (Gen) masks. Statistical significance was tested using the Wilcoxon rank-sum test with a significance level of $\alpha = 0.05$.

	Tumour size			ESD [$\cdot 10^5$]		
	Real	Gen.	p-value	Real	Gen.	p-value
Fold 1	0.12 [0.09]	0.12 [0.07]	0.71	5.09 [6.34]	5.55 [7.49]	0.56
Fold 2	0.12 [0.09]	0.12 [0.09]	0.74	5.27 [6.31]	4.80 [5.55]	0.06
Fold 3	0.10 [0.09]	0.11 [0.09]	0.04	5.28 [7.11]	5.19 [6.65]	0.76
Fold 4	0.12 [0.09]	0.12 [0.08]	0.98	4.75 [6.33]	4.11 [6.12]	0.03
Fold 5	0.12 [0.08]	0.12 [0.07]	0.78	5.58 [7.26]	5.67 [7.91]	0.44

Table 4
Quantitative evaluation across all five folds for mask generator. We report the intercept (β_0) and the slope (β_1) with interval of confidence in bracket.

	Real mask		Generated mask		p-value
	β_1	β_0	β_1	β_0	
Fold 1	0.95 [0.27]	12.85 [0.65]	1.03 [0.30]	13.75 [0.71]	0.40
Fold 2	0.74 [0.32]	12.48 [0.72]	0.85 [0.30]	12.58 [0.67]	0.31
Fold 3	1.06 [0.26]	13.22 [0.64]	0.95 [0.33]	12.90 [0.74]	0.33
Fold 4	1.05 [0.26]	13.07 [0.61]	1.12 [0.27]	13.03 [0.64]	0.51
Fold 5	1.11 [0.26]	13.45 [0.63]	0.97 [0.29]	13.07 [0.69]	0.04

real data. The results suggest that the model replicates the size–shape relationship across all folds, with the exception of Fold 5 (p-value < 0.05), where a statistically significant difference is observed. Fig. 6 shows, for each fold, the distribution of tumour size, ESD, and the corresponding linear fit obtained using Eq. (5).

A visual inspection was provided for qualitative comparison of Pix2Pix, ControlNet, and the proposed framework with Fig. 7 illustrating representative samples. Although the Pix2Pix approach achieved the highest morphological fidelity scores, it performed poorly in terms of overall image realism. In particular, it failed to accurately represent non-tumour tissue and generated tumour regions with unrealistic echogenicity, as reflected in the high FID (238.68) and KID (0.266) values.

The proposed framework and ControlNet both showed better ability to generate realistic tumour tissue and plausible representations of surrounding non-tumour structures (yellow arrows) compared to Pix2Pix approach. Among the two, the proposed framework achieves higher morphological alignment scores, as indicated by better DSC (0.851) and HD (16.21) values. In contrast, ControlNet produced tumours that extend beyond the expected region, with partial overlap of the mask onto non-tumour tissue (yellow boxes), suggesting a limitated ability to enforce precise spatial alignment. This observation is further supported by the statistical analysis in Fig. 8, where the proposed framework showed significantly better morphological alignment compared to ControlNet (Wilcoxon, $p < 0.0001$).

Finally, Fig. 9 shows the results of the downstream segmentation experiments with nnUNet, where synthetic images generated by the proposed model are integrated into the training set, which showed the best overall performance in generating realistic iUS images. As the amount of synthetic data used during training increased, we observed an improvement in the segmentation performance of the nnUNet on the held-out test set. The mean DSC improved from 0.679 (baseline, $R = 0$) to 0.686 at $R = 1$, and to a peak of 0.692 at $R = 2$, while the mean HD decreased from the baseline level of 33.93 to 30.72 at $R = 1$, and to 30.36 at $R = 2$, indicating better spatial accuracy. A statistically significant improvement in HD was observed between $R = 0$ and $R = 1$ (Wilcoxon $p < 0.05$). Increasing the ratio to $R = 3$ did not lead to further improvements, resulting in a slight decrease in performance, with an average HD of 32.62.

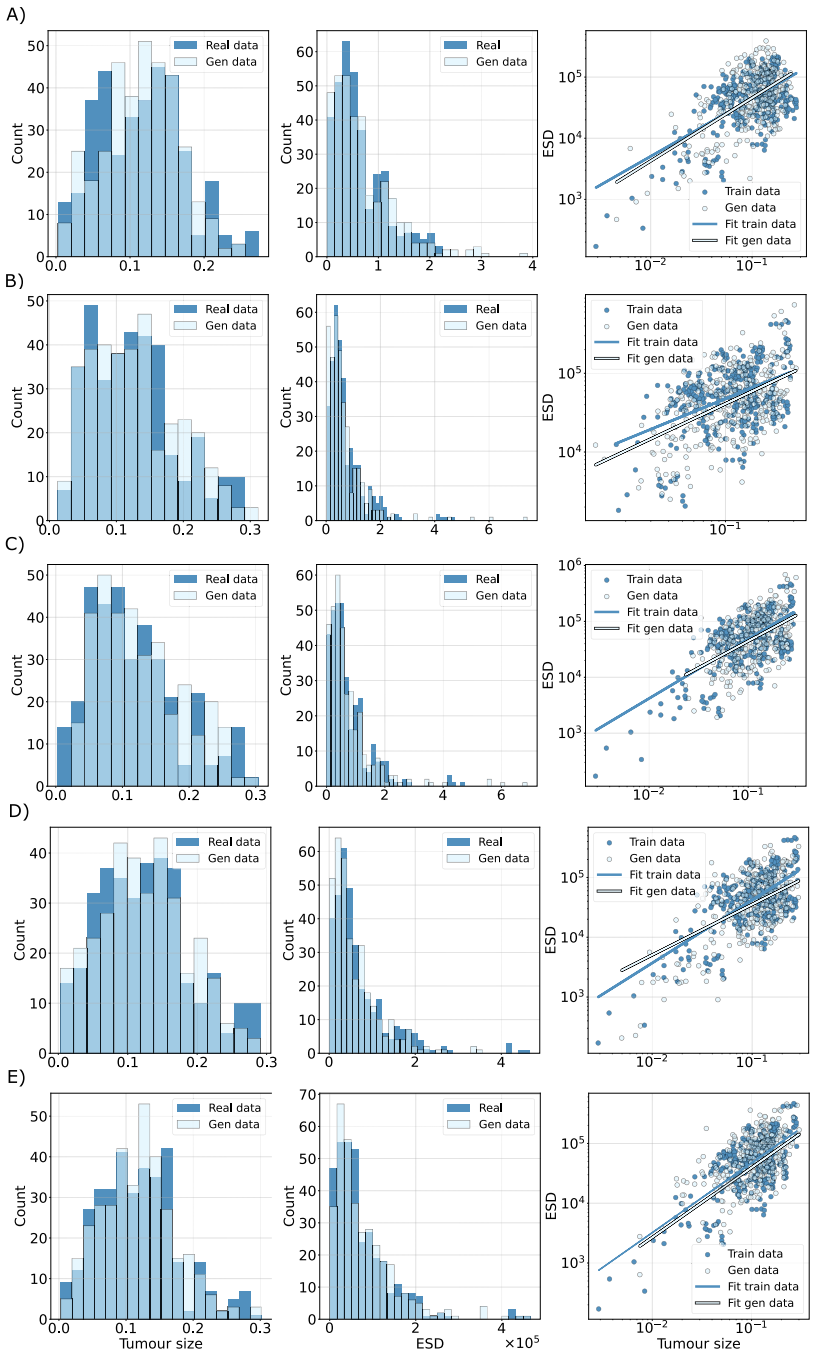


Fig. 6. Evaluation of size-shape relationship. Representation of tumour size (left panel) distribution, ESD (middle panel) distribution, and the size-shape relation (right panel). (A) Fold 1, (B) Fold 2, (C) Fold 3, (D) Fold 4, (E) Fold 5.

5. Discussions

This study investigated the effectiveness of LDMs for generating iUS images of brain tumours. Despite the growing interest in generative models within neurosurgical imaging, most existing efforts have focused on multimodal frameworks combining MRI and iUS, leaving the potential of uni-modal iUS generation with LDMs largely unexplored. To fill this gap, we introduced a fully iUS-based generative framework specifically designed to synthesize morphologically consistent image–mask pairs. The framework adopts a two-step design: first, a mask generator learns the distribution of tumour shapes and generates morphologically plausible masks; second, an iUS generator conditions on these masks to synthesize realistic iUS images that preserve both tumour and surrounding tissue appearance.

To comprehensively evaluate our approach, we compared it against alternative generative strategies, to identify the solution that best balances morphological fidelity and data variability. Quantitative results revealed noteworthy patterns. As shown in Table 2, there seem to be an inverse relationship between image quality metrics (FID and KID) and morphological fidelity metrics (DSC and HD). This trade-off is clearly apparent in the one-step framework, which achieved the best image quality scores but showed lower morphological fidelity compared to Pix2Pix, ControlNet, and the proposed framework. Conversely, the Pix2Pix model shown the highest morphological alignment, while producing the lowest quality images. These behaviours highlight a critical pitfall in relying solely on quantitative metrics for evaluating medical image synthesis. In complex medical domains such as iUS, qualitative

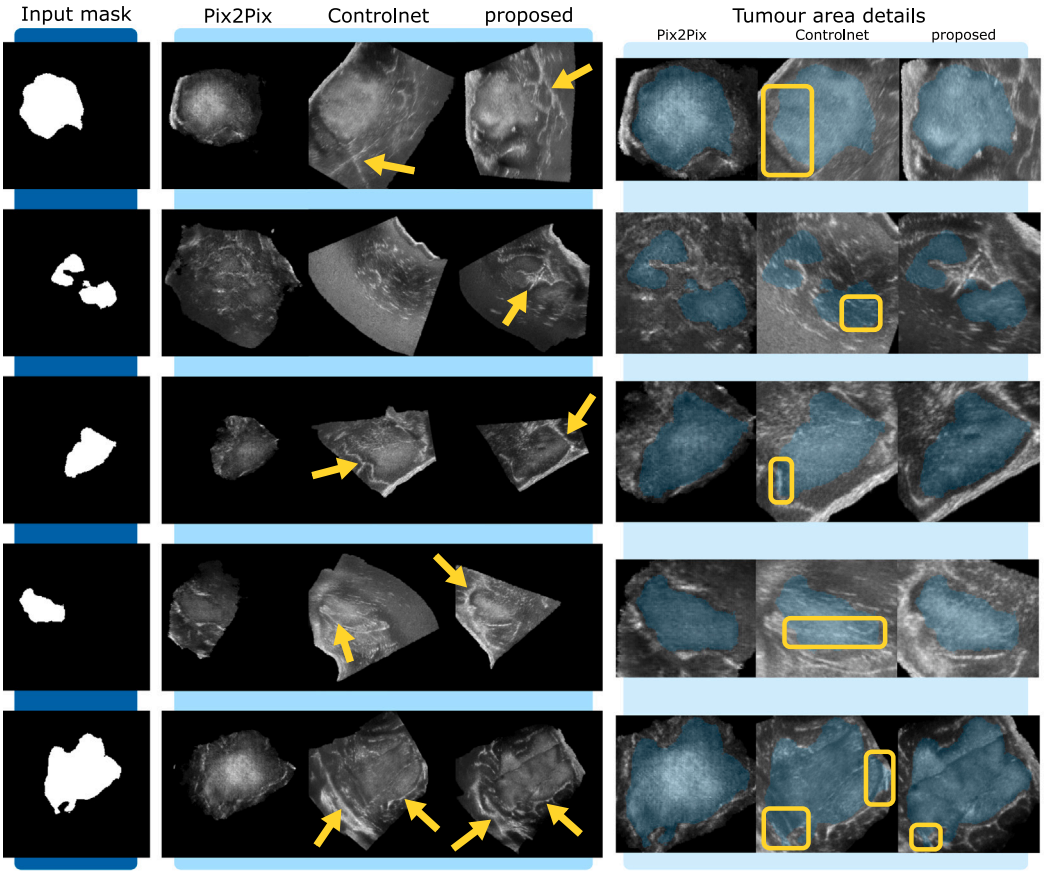


Fig. 7. Qualitative comparison of generated images through different two-step approaches. (First column) Generated input mask. (Second column) Generated iUS images, ControlNet and proposed framework show better ability to reproduce non-tumour tissue (yellow arrow) compared to Pix2Pix approach. (Third column) A details visualization of tumour area, ControlNet produces tumours that extend beyond the expected region, with partial overlap of the mask onto non-tumour tissue (yellow boxes).

evaluation plays a crucial role in revealing strengths and limitations that are not captured by global scores.

In this regard, the visual assessment presented in Fig. 5 shows how the **one-step** framework fails to generalize in terms of tumour shape diversity and iUS image variability. The generated outputs closely mirror samples from the training set, suggesting overfitting. Minimal pixel-wise deviations between real and generated sample indicating that the model tends to replicate training data distribution instead to produce meaningful variability. A plausible explanation is that the task of learning the joint distribution of image and label pairs in a single step is inherently complex and, when constrained by limited training data, may encourage the model to rely on memorization rather than true generalization.

These findings lead to the necessity of decoupling the generative task to better approximate the joint distribution of image–mask pairs. This is addressed by the two-step framework, which separates the problem into two distinct generative sub-tasks: (i) a mask generator that models the unconditional distribution of tumour shapes, and (ii) an iUS image generator that learns the conditional distribution of iUS images given a tumour mask. The underling rational of this modular design allows each model to specialize in a more tractable sub-problem.

For the mask generator model, accurately modelling the geometric characteristics of gliomas is a fundamental requirement. As reported in Table 2, the tumour mask generator showed robust performance in this regard. The distributions of tumour size and ESD are statistically similar between real and generated masks across the majority of folds. Beyond marginal distributions, the model's ability to preserve geometric features is further assessed by analysing the relationship between tumour size and shape complexity. As detailed in Table 4 and illustrated in Fig. 6, the exponential size–shape relationship estimated on generated data closely matches that observed in real tumours in four out of five folds, with no statistically significant differences in the fitted β_1 (Eq. (5)) coefficients. This indicates that the model does not merely reproduce isolated size or shape statistics, but also preserves their joint dependency structure. These results are particularly important as they provide confidence in the use of synthetic masks as priors for conditional iUS image generation. Even though these masks are not derived from real data, their statistical similarity to real tumour masks suggests that they can serve as a reliable starting point for conditional generation.

The comparison between Pix2Pix, ControlNet and the proposed framework provides a broad perspective on which generative model is best suited for morphologically faithful iUS synthesis. As qualitatively illustrated in Fig. 7, each model exhibits distinct strengths and limitations. Generated images by **Pix2Pix** fail to replicate the overall characteristics of real iUS, particularly lacking a coherent representation of the surrounding non-tumour tissue. This limitation is consistent with its poor image quality scores presented in Table 2. A closer inspection of gliomas (third column of Fig. 7) reveals that Pix2Pix overemphasizes the area indicated by the conditioning prior, leading to locally accurate but globally unrealistic outputs. While this strong focus results in high morphologically alignment scores, it also causes the model to neglect the broader anatomical context. Consequently, the generated iUS images appear artificially constrained and fail to capture the structural complexity of real images.

In contrast, both ControlNet and the proposed framework do not suffer from this limitation. As shown in the second column of Fig.

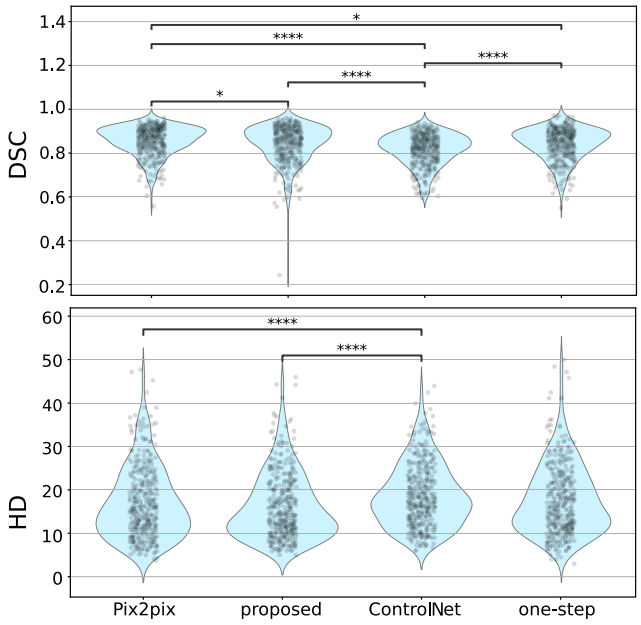


Fig. 8. Statistical evaluation of morphological fidelity metrics. DSC and HD values are collected across all folds. Friedman test reveals statistical differences in DSC ($\chi^2 = 619$, $p < 0.0001$), and HD ($\chi^2 = 147$, $p < 0.0001$) among different experiment. Statistical differences between each experiment are assessed using the Wilcoxon signed-rank statistical test. $*$ denotes the $0.01 < p\text{-value} < 0.05$, $**$ denotes the $0.001 < p\text{-value} \leq 0.01$, $***$ denotes the $0.0001 < p\text{-value} \leq 0.001$, $****$ denotes the $p\text{-value} \leq 0.0001$.

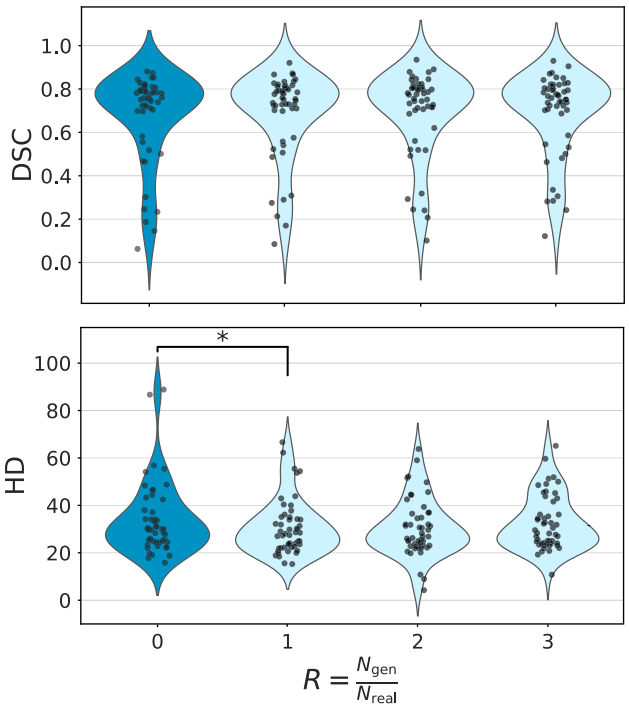


Fig. 9. Statistical results of the clinical utility analysis. Synthetic data (N_{gen}) are progressively integrated with real training data (N_{real}) to evaluate the improvement in segmentation performance as a function of the synthetic-to-real data ratio ($R = N_{gen}/N_{real}$). Friedman test does not reveal any statistical differences in DSC among different R ($\chi^2 = 7.20$, $p = 0.06$), while it reveals statistical differences in HD ($\chi^2 = 11.50$, $p < 0.01$). Statistical differences between the HD distributions between $R = 0$ and $R > 0$ are assessed using the Wilcoxon signed-rank statistical test, with a non-significant result indicating morphological agreement between them. $*$ denotes the p-value < 0.05.

7, these models are able to produce morphologically plausible representations of non-tumour regions, better reflecting the structural heterogeneity seen in real iUS images (highlighted by yellow arrows). These results suggest that LDM-based models outperform GAN-based approaches for iUS image generation, consistent with findings reported in other medical imaging domains [20].

For **Controlnet** and **proposed** framework, the performance across both image quality and morphological fidelity metrics underscores their potential as more clinically meaningful generative tools. However, a key difference emerges upon detailed visual inspection of the tumour regions. iUS images generated by ControlNet often exhibit misalignment between the synthesized tumour and the corresponding tumour masks. As highlighted by the yellow boxes in the third column of Fig. 7, the tumour mask occasionally overlaps with non-tumour regions that clearly exhibit different echogenicity, suggesting inaccurate spatial correspondence. This limitation is less apparent in images generated by proposed framework, where the tumour mask shows minimal overlap with clearly non-tumour regions. The superior alignment achieved by the proposed framework is further supported by the morphological fidelity scores reported in Table 2 and by the statistical analysis presented in Fig. 8, which showed significantly better performance compared to ControlNet.

A plausible explanation for the discrepancy in morphological fidelity performance lies in the different conditioning strategies used in the two models. In ControlNet, the denoising UNet receives only the noisy latent representation as input, while the conditioning information is injected through a separate learnable encoder branch. This auxiliary branch processes the mask and modulates the feature maps of the frozen diffusion backbone via cross-attention and residual connections. This conditional mechanism influencing generation through higher-level geometric features rather than enforcing strict spatial correspondence. In contrast, the proposed framework adopts a more direct conditioning strategy. The model concatenates the tumour mask embedding directly with the noisy latent representation before feeding it into the denoising UNet. This approach injects the spatial information more explicitly into the generation process, effectively anchoring the structure of the tumour in the latent space. As a result, the proposed framework enforces a stronger spatial alignment between the generated iUS image and the corresponding prior.

Overall, the proposed framework outperformed all other generative frameworks in producing high-quality iUS images. In addition to the quantitative evaluation, further qualitative results are provided to visually assess the consistency and diversity of the proposed generative framework. Fig. 10 presents representative examples of generated tumour masks and the corresponding synthetic intraoperative ultrasound images across different tumour size regimes. The ability of the proposed model to account for variability in tumour size includes representative generations of small tumours (left), intermediate tumours with clear depiction of surrounding brain anatomy (centre), and large tumours extending over a wide image area (right). These results illustrate how variations in tumour size and morphology at the mask level are coherently reflected in the generated iUS images, while preserving anatomical consistency across different scenarios.

Building upon these evidences, clinical utility is further evaluated through a data augmentation experiment. By progressively increasing the proportion of synthetic data in the training set, we observed a consistent improvement in segmentation performance on the held-out test set, both in terms of DSC and HD (Fig. 9). These results reveal the effectiveness of synthetic data in enhancing the generalization capabilities of DL models for downstream segmentation tasks. Statistically significant improvement is observed between the baseline ($R = 0$) and the first augmentation setting ($R = 1$) in terms of HD, where the average HD decreased from 33.93 to 30.72. This highlights the clinical relevance

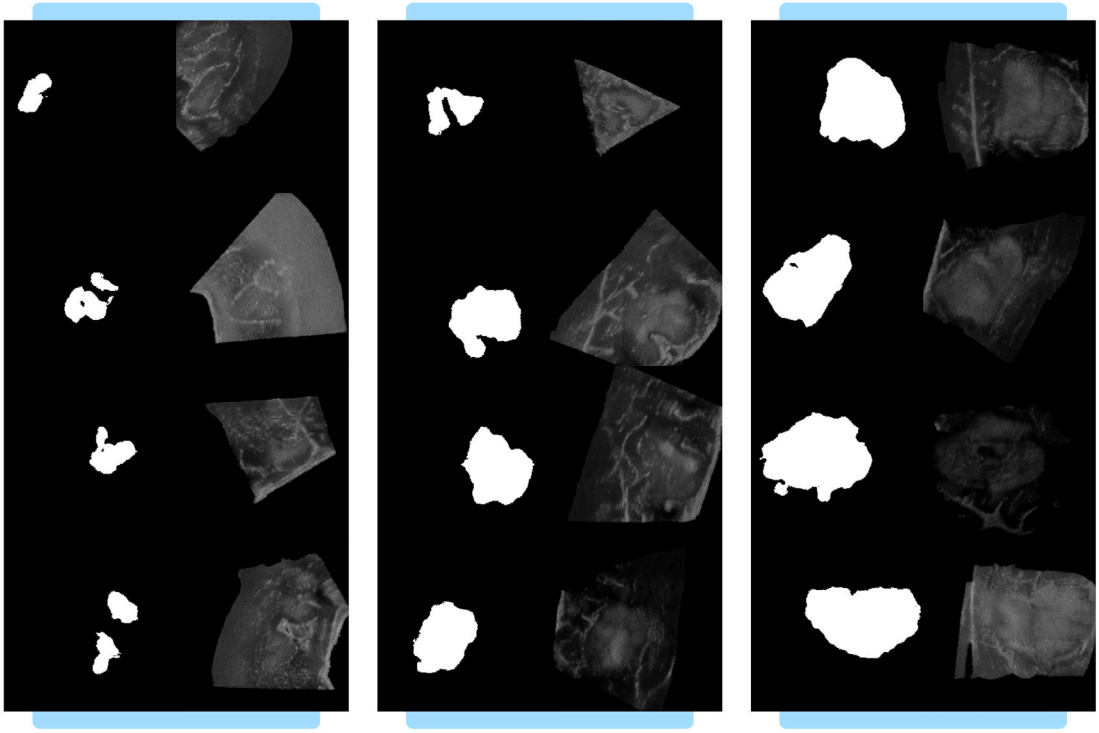


Fig. 10. Representative examples generated by the proposed two-stage framework. Each column shows a generated tumour mask and the corresponding synthetic iUS image, showing morphological variability across different tumour size regimes, including small tumours (left), infiltrative shapes with surrounding functional tissue (centre), and large tumours (right).

of incorporating synthetic data, as HD reflects the spatial accuracy of tumour boundary delineation, a pivotal during intraoperative tumour resection. Improved HD suggests that the model, when trained with incorporation of synthetic data, is better able to predict accurate glioma contours. When the ratio of synthetic to real data exceeded $R = 2$, a slight decline in performance is observed. This finding suggests that an excessive reliance on synthetic data may introduce redundancy into the training set, potentially hindering the model's ability to generalize effectively. Thus, careful calibration of ratio between real and generated data is essential to fully leverage the benefits of generative model as augmentation tool without compromising generalization.

While the proposed framework showed strong performance in generating morphologically realistic tumour representations, some limitations of this study should be listed. The framework follows a two-stage generative design, where tumour morphology is first modelled by the mask generator level and then translated into intraoperative ultrasound images by the iUS generator. The generative capacity of the mask generator remains inherently dependent on the variability represented in the training data. Consequently, although quantitative analyses indicate strong similarity between generated and real tumour geometrical features, the framework does not explicitly enforce coverage of rare or under-represented tumour morphologies. Moreover, the two-stage design implies that the characteristics of the generated masks influence the quality of the final synthetic iUS images, such that deviations from the learned morphological distribution may propagate to the image synthesis stage. Addressing these aspects by systematically characterizing shape variability and monitoring potential distributional biases represents an important direction for future research, particularly when extending the framework to more heterogeneous surgical scenarios. Beyond these methodological aspect, this study also has practical limitations that offer opportunities for future improvement. First, the dataset used in our experiments is based on 3D iUS volumes, which, although valuable for volumetric analysis, differ from native 2D US acquisitions that are more commonly used in clinical practice due to their superior quality. Future work should investigate the performance of generative AI models in the native 2D iUS setting to better align with real-world neurosurgical workflows. Second, the use of binary masks as spatial priors for conditioning the generation process could be seen as a simplification of the complex and often uncertain nature of tumour boundaries in iUS imaging. A promising direction would be to explore more expressive forms of spatial conditioning, such as confidence maps which can better capture the intrinsic ambiguity of tumour margins. Incorporating such uncertainty-aware priors may enhance the anatomical realism of generated images, as shown in related work on US image synthesis tasks [22].

Regarding the use of the proposed framework as a data augmentation tool, limitations related to model generalization must be acknowledged. Although the augmentation experiments showed performance improvements on a test set acquired from the same clinical centre, such findings do not necessarily translate to robust generalization when applied to data collected at different institutions or using different ultrasound devices in prospective studies [49]. Intraoperative ultrasound is characterized by high level of variability arising from differences in glioma morphology, intraoperative acquisition protocols, probe types, and operator-dependent factors. A generative model trained on single-centre data may therefore capture centre-specific imaging characteristics. When used for data augmentation, such a model risks introducing a centre-dependent bias, which may lead to performance degradation rather than improvement under domain shift. These observations highlight the need for caution when interpreting data augmentation gains obtained in single-centre settings. Future work should prioritize multi-centre validation to explicitly assess generalization across devices and institutions [50], and to mitigate centre-dependent biases by incorporating greater real-world variability into the development of generative AI models for iUS.

6. Conclusion

This study introduced two-step LDM framework for generating iUS images of gliomas. By decoupling the unconditional generation of

tumour mask from the conditional generation on iUS image using tumour mask as prior, the proposed framework is able to generate morphologically faithful iUS data while explicitly controlling tumour shape variability. Quantitative and qualitative evaluations show that the generated tumour morphologies are statistically consistent with real glioma characteristics, and that the corresponding synthetic iUS images preserve relevant anatomical structure. Moreover, results from the data augmentation experiments highlight that incorporating synthetic data can significantly enhance the performance of DL models in automatically identifying tumour boundaries, suggesting the potential of LDMs as a valuable tool for addressing the scarcity of annotated iUS data. These findings suggest that diffusion-based generative frameworks can play a meaningful role in addressing data scarcity in intraoperative ultrasound imaging, supporting the development of more robust segmentation models. Future research will focus on extending the framework to more heterogeneous acquisition settings, exploring multi-centre data, and further improving the generalization of tumour shape modelling to better capture the variability observed in real surgical scenarios.

CRediT authorship contribution statement

Angelo Lasala: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Maria Chiara Fiorentino:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Andrea Bandini:** Writing – review & editing, Project administration, Formal analysis. **Sara Moccia:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Stamatia Giannarou:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Funding

This work was supported by the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union – NextGenerationEU – Project Title "THAI-MIA" – CUP B53D23005000006. We would like also to acknowledge the "Italian Fund for Applied Sciences" (FISA) grant no. FISA2022-00696.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledge

The authors acknowledge the CINECA award under the ISCRA project, for the availability of high-performance computing resources.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.bspc.2026.110037.

Data availability

The authors do not have permission to share data.

References

[1] A. Omuro, L.M. DeAngelis, Glioblastoma and other malignant gliomas: A clinical review, JAMA 310 (2013) 1842–1850, http://dx.doi.org/10.1001/jama.2013.280319.

[2] D.N. Louis, A. Perry, P. Wesseling, D.J. Brat, I.A. Cree, D. Figarella-Branger, C. Hawkins, H.K. Ng, S.M. Pfister, G. Reifenberger, R. Soffietti, A.V. Deimling, D.W. Ellison, The 2021 WHO classification of tumors of the central nervous system: A summary, Neuro-Oncol. 23 (2021) 1231–1251, http://dx.doi.org/10.1093/neuonc/noab106.

[3] W. Wick, M. Osswald, A. Wick, F. Winkler, Treatment of glioblastoma in adults, Ther. Adv. Neurol. Disord. 11 (2018) http://dx.doi.org/10.1177/1756286418790452.

[4] M.J. Mair, M. Geurts, M.J. van den Bent, A.S. Berghoff, A basic review on systemic treatment options in WHO grade II-iii gliomas, Cancer Treat. Rev. 92 (2021) http://dx.doi.org/10.1016/j.ctrv.2020.102124.

[5] O. Bin-Alamer, H. Abou-Al-Shaar, Z.C. Gersey, S. Huq, J.A. Kallos, D.J. McCarthy, J.R. Head, E. Andrews, X. Zhang, C.G. Hadjipanayis, Intraoperative imaging and optical visualization techniques for brain tumor resection: A narrative review, Cancers 15 (2023) http://dx.doi.org/10.3390/cancers15194890.

[6] P.L. Kubben, K.J. ter Meulen, O.E. Schijns, M.P. ter Laak-Poort, J.J. van Overbeeke, H. van Santbrink, Intraoperative MRI-guided resection of glioblastoma multiforme: a systematic review, Lancet Oncol. 12 (11) (2011) 1062–1070.

[7] L. Dixon, A. Lim, M. Grech-Sollars, D. Nandi, S. Camp, Intraoperative ultrasound in brain tumor surgery: A review and implementation guide, Neurosurg. Rev. 45 (4) (2022) 2503–2515.

[8] A. Šteňo, J. Buvala, V. Babková, A. Kiss, D. Toma, A. Lysak, Current limitations of intraoperative ultrasound in brain tumor surgery, Front. Oncol. 11 (2021) http://dx.doi.org/10.3389/fonc.2021.659048.

[9] A. Weld, L. Dixon, G. Anichini, N. Patel, A. Nimer, M. Dyck, K. O'Neill, A. Lim, S. Giannarou, S. Camp, Challenges with segmenting intraoperative ultrasound for brain tumours, Acta Neurochir. 166 (2024) http://dx.doi.org/10.1007/s00701-024-06179-8.

[10] Y. Xiao, M. Fortin, G. Unsgärd, H. Rivaz, I. Reinertsen, REtroSpective evaluation of cerebral tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries: A, Med. Phys. 44 (2017) 3875–3882, http://dx.doi.org/10.1002/mp.12268.

[11] B. Behboodi, F.X. Carton, M. Chabanas, S. de Ribaupierre, O. Solheim, B.K. Munkvold, H. Rivaz, Y. Xiao, I. Reinertsen, Open access segmentations of intraoperative brain tumor ultrasound images, Med. Phys. (2024) http://dx.doi.org/10.1002/mp.17317.

[12] P. Juvekar, R. Dorent, F. Kögl, E. Torio, C. Barr, L. Rigolo, C. Galvin, N. Jowkar, A. Kazi, N. Haouchine, H. Cheema, N. Navab, S. Pieper, W.M. Wells, W.L. Bi, A. Golby, S. Frisken, T. Kapur, Remind: The brain resection multimodal imaging database, Sci. Data 11 (2024) http://dx.doi.org/10.1038/s41597-024-03295-z.

[13] S. Cepeda, O. Esteban-Sinovas, V. Singh, P. Shetty, A. Moiyadi, L. Dixon, A. Weld, G. Anichini, S. Giannarou, S. Camp, I. Zemmoura, G.R. Giammalva, M.D. Bene, A. Barbotti, F. DiMeco, T.R. West, B.V. Nahed, R. Romero, I. Arrese, R. Hornero, R. Sarabia, Deep learning-based glioma segmentation of 2D intraoperative ultrasound images: A multicenter study using the brain tumor intraoperative ultrasound database (BraTioUS), Cancers 17 (2025) http://dx.doi.org/10.3390/cancers17020315.

[14] L. Canalini, J. Klein, D. Miller, R. Kikinis, Segmentation-based registration of ultrasound volumes for glioma resection in image-guided neurosurgery, Int. J. Comput. Assist. Radiol. Surg. 14 (2019) 1697–1713, http://dx.doi.org/10.1007/s11548-019-02045-6.

[15] R. Dorent, N. Haouchine, F. Kögl, S. Joutard, P. Juvekar, E. Torio, A. Golby, S. Ourselin, S. Frisken, T. Vercauteren, T. Kapur, W.M. Wells, Unified brain MR-ultrasound synthesis using multi-modal hierarchical representations, 2023, http://dx.doi.org/10.1007/978-3-031-43999-5_43, URL http://arxiv.org/abs/2309.08747.

[16] S. Singh, M. Bewoor, A. Ranapurwala, S. Rai, S. Patil, BrainVoxGen: Deep learning framework for synthesis of ultrasound to MRI, 2023, URL http://arxiv.org/abs/2310.08608.

[17] A.G. Eker, M.K. Pehlivanoğlu, N. Duru, T.T. Dündar, BrainPixGAN: Generating intraoperative MRI images with mask-based generative networks, Eng. Sci. Technol. an Int. J. 58 (2024) http://dx.doi.org/10.1016/j.jestch.2024.101827.

[18] M. Rahmani, H. Moghaddasi, A. Pour-Rashidi, A. Ahmadian, E. Najafzadeh, P. Farnia, D2BGAN: Dual discriminator Bayesian generative adversarial network for deformable MR–ultrasound registration applied to brain shift compensation, Diagnostics 14 (2024) http://dx.doi.org/10.3390/diagnostics14131319.

[19] M. Donnez, F.-X. Carton, F. Le Lann, E. De Schlichting, M. Chabanas, Realistic synthesis of brain tumor resection ultrasound images with a generative adversarial network, in: Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 11598, SPIE, 2021, pp. 637–642.

[20] G. Müller-Franzes, J.M. Niehues, F. Khader, S.T. Arasteh, C. Haarburger, C. Kuhl, T. Wang, T. Han, T. Nolte, S. Nebelung, J.N. Kather, D. Truhn, A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis, Sci. Rep. 13 (2023) http://dx.doi.org/10.1038/s41598-023-39278-0.

[21] A. Lasala, M.C. Fiorentino, A. Bandini, S. Moccia, FetalBrainAwareNet: Bridging GANs with anatomical insight for fetal ultrasound brain plane synthesis, Comput. Med. Imaging Graph. 116 (2024) 102405, http://dx.doi.org/10.1016/j.compmedimag.2024.102405.

[22] A. Lasala, M.C. Fiorentino, A. Bandini, S. Moccia, Conditional latent diffusion models for PLAX echocardiographic image synthesis: A geometric-anatomical guided approach, IEEE Trans. Med. Robot. Bionics (2025) http://dx.doi.org/10.1109/TMRB.2025.3617977.

[23] L. Maier-Hein, A. Reinke, P. Godau, M.D. Tizabi, F. Buettner, E. Christodoulou, B. Glocker, F. Isensee, J. Kleesiek, M. Kozubek, et al., Metrics reloaded: recommendations for image analysis validation, Nature Methods 21 (2) (2024) 195–212.

[24] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2017, pp. 5967–5976, http://dx.doi.org/10.1109/CVPR.2017.632.

[25] C. Baldini, K. Kushibar, R. Osuala, S. Balocco, O. Diaz, K. Lekadir, L.S. Mattos, Clinically-guided data synthesis for laryngeal lesion detection, in: J.C. Gee, D.C. Alexander, J. Hong, J.E. Iglesias, C.H. Sudre, A. Venkataraman, P. Golland, J.H. Kim, J. Park (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, Springer Nature Switzerland, Cham, 2026, pp. 54–63.

[26] H. Cai, B. Ji, S. Cai, Y. Liao, J. Chen, W. Huang, GE2hist: Generating histology images from single-cell gene expression via cross-modal generative network, in: J.C. Gee, D.C. Alexander, J. Hong, J.E. Iglesias, C.H. Sudre, A. Venkataraman, P. Golland, J.H. Kim, J. Park (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, Springer Nature Switzerland, Cham, 2026, pp. 240–250.

[27] J. Shentu, M. Watson, N. Al Moubayed, DiDGen: Diffusion-based dual-task synthesis for dermoscopic data generation, in: J.C. Gee, D.C. Alexander, J. Hong, J.E. Iglesias, C.H. Sudre, A. Venkataraman, P. Golland, J.H. Kim, J. Park (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, Springer Nature Switzerland, Cham, 2026, pp. 74–84.

[28] M. Luna, J. Baek, W.H. Kim, W.G. Son, K.M. Lee, H.J. Kim, J. Kim, Improved tumor segmentation using selective synthetic augmentation for enhanced surgical planning in breast MRI, in: J.C. Gee, D.C. Alexander, J. Hong, J.E. Iglesias, C.H. Sudre, A. Venkataraman, P. Golland, J.H. Kim, J. Park (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2025, Springer Nature Switzerland, Cham, 2026, pp. 315–324.

[29] Y. Xie, J. Wang, T. Feng, F. Ma, Y. Li, CCIS-diff: A generative model with stable diffusion prior for controlled colonoscopy image synthesis, in: 2025 IEEE 22nd International Symposium on Biomedical Imaging, ISBI, 2025, pp. 1–5, http://dx.doi.org/10.1109/ISBI60581.2025.10981078.

[30] Y.-C. Chou, G.Y. Li, L. Chen, M. Zahiri, N. Balaraju, S. Patil, B. Hicks, N. Schnittke, M. Parker, D.O. Kessler, J. Shupp, C. Baloescu, C. Moore, C. Gregory, K. Gregory, B. Raju, J. Kruecker, A. Chen, Ultrasound image synthesis using generative AI for lung consolidation detection, in: 2025 IEEE 22nd International Symposium on Biomedical Imaging, ISBI, 2025, pp. 1–5, http://dx.doi.org/10.1109/ISBI60581.2025.10980728.

[31] J. Kaleta, D. Dall'Alba, S. Płotka, P. Korzeniowski, Minimal data requirement for realistic endoscopic image generation with stable diffusion, Int. J. Comput. Assist. Radiol. Surg. 19 (3) (2024) 531–539.

[32] H.K. Kim, I.H. Ryu, J.Y. Choi, T.K. Yoo, A feasibility study on the adoption of a generative denoising diffusion model for the synthesis of fundus photographs using a small dataset, Discov. Appl. Sci. 6 (4) (2024) 188.

[33] C. Yu, H. Fang, H. Wang, T. Deng, Q. Du, Y. Xu, W. Yang, Rethinking diffusion-based image generators for fundus fluorescein angiography synthesis on limited data, 2024, arXiv preprint arXiv:2412.12778.

[34] A. Kebaili, J. Lapuyade-Lahorgue, P. Vera, S. Ruan, 3D mri synthesis with slice-based latent diffusion models: Improving tumor segmentation tasks in data-scarce regimes, in: 2024 IEEE International Symposium on Biomedical Imaging, ISBI, IEEE, 2024, pp. 1–5.

[35] Y. Zhou, R. Towning, Z. Awad, S. Giannarou, Image synthesis with class-aware semantic diffusion models for surgical scene segmentation, Heal. Technol. Lett. 12 (1) (2025) e70003.

[36] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, 2022-June, IEEE Computer Society, 2022, pp. 10674–10685, http://dx.doi.org/10.1109/CVPR52688.2022.01042,

[37] C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, et al., Laion-5b: An open large-scale dataset for training next generation image-text models, Adv. Neural Inf. Process. Syst. 35 (2022) 25278–25294.

[38] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., Learning transferable visual models from natural language supervision, in: International Conference on Machine Learning, PmLR, 2021, pp. 8748–8763.

[39] B.K.R. Munkvold, H.K. Bø, A.S. Jakola, I. Reinertsen, E.M. Berntsen, G. Unsgård, S.H. Torp, O. Solheim, Tumor volume assessment in low-grade gliomas: a comparison of preoperative magnetic resonance imaging to coregistered intraoperative 3-dimensional ultrasound recordings, Neurosurgery 83 (2) (2018) 288–296.

[40] L. Zhang, A. Rao, M. Agrawala, Adding conditional control to text-to-image diffusion models, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 3836–3847.

[41] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, Proc. the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2018) 586–595, http://dx.doi.org/10.1109/CVPR.2018.00068.

[42] J. Ho, T. Salimans, Classifier-free diffusion guidance, 2022, arXiv preprint arXiv:2207.12598.

[43] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local Nash equilibrium, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017.

[44] G. Zamzmi, A. Subbaswamy, E. Sizikova, E. Margerrison, J. Delfino, A. Badano, Scorecards for synthetic medical data evaluation and reporting, 2024, arXiv preprint arXiv:2406.11143.

[45] Y. Deo, Y. Jia, T. Lassila, W.A. Smith, T. Lawton, S. Kang, A.F. Frangi, I. Habli, Metrics that matter: Evaluating image quality metrics for medical image generation, 2025, arXiv preprint arXiv:2505.07175.

[46] M. Bińkowski, D.J. Sutherland, M. Arbel, A. Gretton, Demystifying mmd gans, 2021, arXiv preprint arXiv:1801.01401.

[47] J. Ma, Y. He, F. Li, L. Han, C. You, B. Wang, Segment anything in medical images, Nat. Commun. 15 (1) (2024) 654.

[48] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211.

[49] J.S. Yoon, K. Oh, Y. Shin, M.A. Mazurowski, H.-I. Suk, Domain generalization for medical image analysis: A review, Proc. IEEE 112 (10) (2024) 1583–1609, http://dx.doi.org/10.1109/JPROC.2024.3507831.

[50] S. Kumari, P. Singh, Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives, Comput. Biol. Med. 170 (2024) 107912, http://dx.doi.org/10.1016/j.compbiomed.2023.107912.