High Volume Rate 3D Ultrasound Reconstruction with Diffusion Models

Tristan S.W. Stevens, *Graduate Student Member, IEEE*, Oisín Nolan, *Graduate Student Member, IEEE*, Oudom Somphone, Jean-Luc Robert and Ruud J.G. van Sloun, *Member, IEEE*

***Abstract*— Three-dimensional ultrasound enables real-time volumetric visualization of anatomical structures. Unlike traditional 2D ultrasound, 3D imaging reduces the reliance on precise probe orientation, potentially making ultrasound more accessible to clinicians with varying levels of experience and improving automated measurements and post-exam analysis. However, achieving both high volume rates and high image quality remains a significant challenge. While 3D diverging waves can provide high volume rates, they suffer from limited tissue harmonic generation and increased multipath effects, which degrade image quality. One compromise is to retain the focusing in elevation while leveraging unfocused diverging waves in the lateral direction to reduce the number of transmissions per elevation plane. Reaching the volume rates achieved by full 3D diverging waves, however, requires dramatically undersampling the number of elevation planes. Subsequently, to render the full volume, simple interpolation techniques are applied. This paper introduces a novel approach to 3D ultrasound reconstruction from a reduced set of elevation planes by employing diffusion models (DMs) to achieve increased spatial and temporal resolution. We compare both traditional and supervised deep learning-based interpolation methods on a 3D cardiac ultrasound dataset. Our results show that DM-based reconstruction consistently outperforms the baselines in image quality and downstream task performance. Additionally, we accelerate inference by leveraging the temporal consistency inherent to ultrasound sequences. Finally, we explore the robustness of the proposed method by exploiting the probabilistic nature of diffusion posterior sampling to quantify reconstruction uncertainty and demonstrate improved recall on out-of-distribution data with synthetic anomalies under strong subsampling.**

***Index Terms*— 3D ultrasound, diffusion models, generative modeling, cardiac ultrasound**

I. Introduction

THREE-dimensional (3D) ultrasound imaging, also known as volumetric ultrasound imaging, is achieved through the use of a 2D matrix probe which enables focusing of ultrasound beams in both azimuth and elevation dimensions through electronic steering [1]. It overcomes several challenges associated with traditional 2D ultrasound imaging, such as the inability to capture out-of-plane motion and the lack of spatial orientation. These limitations make 2D imaging highly operator-dependent, requiring multiple scans to mentally reconstruct anatomy, which is time-consuming, variable across users, and difficult to reproduce. As a result, diagnostic accuracy and interventional guidance can suffer from inconsistency [2], [3].

The transition to 3D ultrasound imaging addresses these limitations by offering volumetric data, enabling applications such as 3D echocardiography [4]–[6] and 3D breast imaging [7]. It also provides improved guidance during image-guided therapy and biopsies [8] and supports greater automation of critical diagnostic measurements, including ejection fraction [9] and strain assessment [10], [11].

However, 3D imaging introduces new challenges, such as increased data acquisition times and higher computational demands [12]. These factors constrain the design of probes and transmit sequences, limiting the accessibility of high-quality 3D imaging in real-time clinical settings. Additionally, the reduced image quality compared to 2D imaging often discourages clinicians from utilizing 3D imaging, as the degradation in resolution and clarity frequently outweighs its potential benefits in practical scenarios.

In addressing the trade-offs inherent to 3D ultrasound, it is beneficial to differentiate between two key phases of the imaging process: *acquisition* and *reconstruction*. Acquisition refers to the process of transmitting and receiving ultrasound waves, which can be optimized to minimize data acquisition times while maintaining sufficient information for effective image reconstruction. The increased field of view in 3D imaging, while enabling more comprehensive anatomical coverage, leads to acquisition times that are prohibitively long under standard transmit schemes. Combined with the large number of elements in 2D matrix arrays, which significantly increases data rates, this has necessitated the development of bandwidth-efficient solutions.

Microbeamforming reduces data transmission by performing partial beamforming within the probe itself, grouping elements into sub-apertures, at the cost of reduced lateral resolution. Slow-time multiplexing limits the number of active elements per transmission, reducing bandwidth but lowering frame rates. Other techniques like compressed sensing (CS) aim to reduce the number of samples needed by exploiting the inherent sparsity of ultrasound signals. In 2D ultrasound, Wagner *et al.* [13] applied CS to radio-frequency (RF) signals using Xampling, simplifying acquisition through structured signal representations. This was extended to 3D ultrasound by Burshtein et al. [14], who demonstrated the approach for volumetric imaging. Learned approaches for designing sparse

Manuscript received May, 2025; This work was performed within the IMPULSE framework of the Eindhoven MedTech Innovation Center (e/MTIC, incorporating Eindhoven University of Technology and Philips Research), including a PPS supplement from the Dutch Ministry of Economic Affairs and Climate Policy.

Tristan S.W. Stevens (e-mail: t.s.w.stevens@tue.nl), Oisín Nolan and Ruud J.G. van Sloun are with the Electrical Engineering Department, Eindhoven University of Technology, the Netherlands. Oudom Somphone and Jean-Luc Robert are with Philips Research, Paris, France and Cambridge MA, USA, respectively.

sensing matrices in ultrasound imaging have shown promising results. Lorintiu *et al.* [15] employed learned overcomplete dictionaries for scanline selection in volumetric ultrasound, while Huijben *et al.* [16] proposed a deep learning-based probabilistic subsampling scheme that generates context-specific sensing matrices for 2D ultrasound.

Reconstruction, on the other hand, involves the process of generating a 3D image or volume from the acquired data. Traditional reconstruction techniques, such as pixel or voxel nearest-neighbor (PNN or VNN) interpolation, simply fill in missing information by relying on nearby sampled points [12], [17]. Function-based interpolation methods are an improved, albeit usually more computationally demanding, alternative, where a particular function, for example, a polynomial, is fitted through the acquired voxels.

Recently, deep learning has emerged as a powerful tool for ultrasound image reconstruction [18]. Specifically, deep generative models (DGMs) have proven to be the missing link in learning expressive image priors [19], [20]. Unlike conventional supervised approaches that learn a direct mapping from measurements to images, generative models learn the underlying data distribution in a self-supervised manner. This enables reconstruction to be framed as *Bayesian* inference, where the *likelihood* captures the forward measurement process and the generative model acts as a learned *prior*. *Posterior* sampling then allows for the recovery of plausible and diverse image reconstructions from sparse or noisy inputs. Consequently, DGMs are task-agnostic and naturally support uncertainty estimation within a probabilistic framework, improving robustness to out-of-distribution (OoD) data.

A specific subset of DGMs, generative adversarial networks (GANs), has been successfully applied in 3D ultrasound imaging to upscale sparsely acquired 2D images to reconstruct a full volume [21], [22]. Also, variational autoencoders (VAEs) have been used to synthesize 3D ultrasound data with the purpose of augmenting existing datasets [23].

More recently, diffusion models have seen a surge of interest due to their powerful generative capabilities and practical training objective. In the field of ultrasound, diffusion models are used for image generation [24]–[26], denoising and dehazing [27]–[29] and image reconstruction [30]–[32]. While diffusion models generally converge more reliably during training compared to GANs and produce higher quality samples than VAEs, they tend to be more computationally demanding during inference due to their iterative nature. Many works have shown, however, that it is feasible to accelerate diffusion models using various techniques, such as leveraging structure in sequential data [33] or model distillation [34].

In this work, we propose a flexible interpolation framework based on diffusion models for 3D ultrasound reconstruction, enabling high-volume rates without compromising image quality. Specifically, we demonstrate our approach on an *in-vivo* cardiac 3D ultrasound dataset, achieving a $4\times$ increase in volume rate while maintaining comparable image quality and performance in downstream tasks.

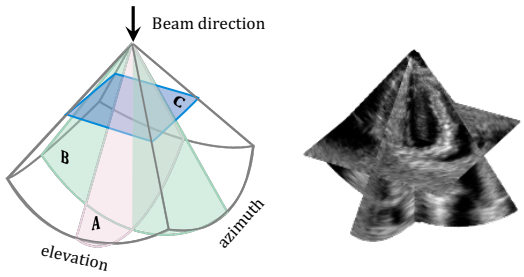

Fig. 1: Geometric visualization of the three common cross-sections in volumetric ultrasound imaging (left) with a B-mode example (right).

Our main contributions can be summarized as follows:

- Development of a deep generative prior for 3D cardiac ultrasound in the form of a diffusion model.
- A practical posterior sampling framework that can interpolate sparsely sampled ultrasound volumes.
- Techniques to mitigate and visualize uncertainty of the generative model by drawing multiple posterior samples.
- An extensive comparison with existing interpolation techniques, showcasing the advantage of using deep generative priors, with evaluation based on both image quality metrics and a clinically relevant downstream task.
- Code and supplementary materials are available.[1]

II. Background

In this section, we will briefly discuss basic concepts in 3D ultrasound imaging and diffusion models and their use in learning image priors, which form the basis of the proposed methodology.

A. 3D Ultrasound Acquisition

Modern ultrasound probes that enable 3D imaging form a matrix of transducer elements, that allows for dynamic beam steering along both azimuth (lateral) and elevation axes. The full 3D ultrasound volume $\mathbf{X} \in \mathbb{R}^{N_{\text{el}}, N_{\text{az}}, N_{\text{ax}}}$, after acquisition and beamforming, can be written as a tensor:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,N_{\text{az}}} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,N_{\text{az}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{N_{\text{el}},1} & \mathbf{x}_{N_{\text{el}},2} & \cdots & \mathbf{x}_{N_{\text{el}},N_{\text{az}}} \end{bmatrix}, \tag{1}$$

with elements $\mathbf{x} \in \mathbb{R}^{N_{\text{ax}}}$. The dimensions N_{el}, N_{az}, and N_{ax} represent the number of samples acquired along the elevation, azimuth, and axial directions, respectively. From $\mathbf{X}$, three orthogonal planes can be identified: the A, B, and C planes. Rows in $\mathbf{X}$ can be considered lateral 2D images at a fixed elevation plane, akin to traditional 2D B-mode imaging (A plane). The columns represent elevation images (B plane) which are perpendicular to the A plane. If A plane images provide a long-axis view, the B plane offers the corresponding short-axis view. The C plane, represents slices parallel to the

[1] https://3d-ultrasound-diffusion.github.io

probe surface, acquired at a fixed axial depth. See Fig. 1 for a visualization of all three planes in relation to each other.

For transthoracic transducers, the design is often asymmetric, featuring a smaller aperture along the elevation axis to facilitate access through narrow imaging windows between the ribs. The smaller aperture broadens the beam and reduces elevational resolution, while a lower element count with wider spacing increases grating lobe artifacts, degrading image quality. Typically, to achieve sufficient temporal resolution, a faster acquisition scheme such as diverging waves is employed along the azimuth dimension, while focused transmits are applied along elevation at a coarse interval. Employing only diverging waves in both dimensions significantly degrades image quality because the unfocused beams produce lower acoustic pressure, resulting in reduced harmonic generation and increased multipath scattering. To compensate for these effects, higher transmit power is required, which raises the mechanical index (MI) and poses challenges for transducer design and patient safety. In contrast, fully focusing across both dimensions improves image quality and resolution, but at the cost of frame rate, since more transmits are needed to cover the entire field of view. As a compromise, combining diverging waves in azimuth with focused transmits along elevation balances image quality and acquisition speed. Limiting the number of focused elevation angles directly impacts the maximum achievable volume rate, but introduces gaps in the volumetric data. To mitigate this, we adopt a sparse interlocking acquisition scheme, as illustrated in Fig. 2, where different elevation planes are acquired at each time step in a staggered pattern. This ensures that across consecutive frames, complementary slices are captured, allowing temporal redundancy to aid in volumetric reconstruction. Still, at lower acquisition rates, high-quality interpolation remains a challenge.

B. Diffusion Models

Generative models seek to learn the data distribution $p(\mathbf{x})$ of some random variable $\mathbf{x}$. Instead of explicitly modeling the distribution, which is often intractable given high-dimensional data, these generative models implicitly model their distribution using a generative process, i.e. producing samples from the underlying distribution $\mathbf{x} \sim p(\mathbf{x})$. Diffusion models (DMs) [35], [36] are a class of generative models that define the generative process as the reversal of a corruption process, which transforms $\mathbf{x}_0 \equiv \mathbf{x} \sim p(\mathbf{x})$ into a Gaussian base distribution $\mathbf{x}_\mathcal{T} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This continuous forward process $\mathbf{x}_0 \to \mathbf{x}_\tau \to \mathbf{x}_\mathcal{T}$, with diffusion time $\tau \in [0, \mathcal{T}]$, can be trivially executed through adding Gaussian noise to initial data samples:

$$\mathbf{x}_\tau = \alpha_\tau \mathbf{x}_0 + \sigma_\tau \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

where α_τ and σ_τ are the signal and noise terms respectively, defined by some pre-defined noise schedule. Naturally, we are interested in the reversal of this corruption process, which is akin to sampling from the target distribution $p(\mathbf{x})$. This can be interpreted as iteratively denoising the noisy estimate $\mathbf{x}_\tau$ [37]. Tweedie's formula relates the minimum mean square error estimator (MMSE) to the score of the distribution:

$$\mathbf{x}_0 \approx \mathbb{E}[\mathbf{x}_0|\mathbf{x}_\tau] = \frac{1}{\alpha_\tau}(\mathbf{x}_\tau + \sigma_\tau^2 \underbrace{\nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)}_{\text{score}}), \tag{3}$$

where $\mathbf{x}_{0|\tau} := \mathbb{E}[\mathbf{x}_0|\mathbf{x}_\tau]$ represents the one-step denoised estimate from diffusion step τ. The score function, defined as the gradient of the log-likelihood of $\mathbf{x}_\tau$, points towards the data distribution. Since Tweedie's formula provides only a local (at time τ) estimate, the final prediction of $\mathbf{x}_0$ is obtained iteratively. At each step, $\mathbf{x}_0$ is estimated using (3) and mapped back to $\mathbf{x}_{\tau-1}$ via forward diffusion in (2), ensuring a smooth sampling trajectory. The score function can be parameterized using a neural network, conditioned on the diffusion timestep τ. For practical reasons, the noise is predicted instead of the score, as it directly relates to the score via $\epsilon_\theta(\mathbf{x}_\tau, \tau) \approx -\sigma_\tau \nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)$ [38]. The network can be simply trained using the denoising score-matching objective as follows:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0 \sim p(\mathbf{x}_0), \epsilon, \tau}\left[\|\epsilon_\theta(\mathbf{x}_\tau, \tau) - \epsilon\|^2\right]. \tag{4}$$

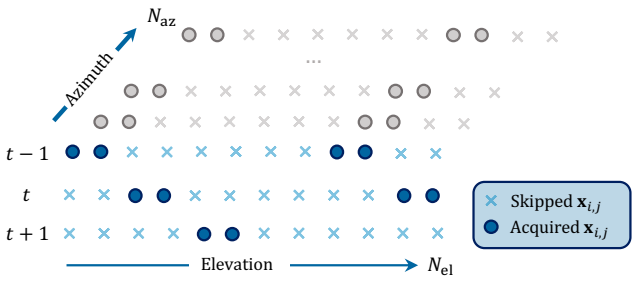

Fig. 2: Schematic representation of the acquisition scheme, with sparse interlocking patterns over time. A full set of azimuthal lines (corresponding to a single elevation plane) is acquired at a time.

III. Methods

We formulate sparse 3D ultrasound reconstruction as an inverse problem, using DMs as image priors. This allows training a self-supervised model that learns the data distribution without paired examples. The learned prior is then applied during the inversion process, offering a task-agnostic framework that can be applied to various inverse problems without retraining. For the problem at hand, this flexibility allows for modifying the sampled scan-lines at test time, effectively changing the measurement matrix (i.e. sampling mask) adaptively. A conceptual overview of the proposed method is shown in Fig. 3.

We begin by formally describing the inverse problem in Section III-A, defining the interpolation task within a probabilistic framework and outlining how posterior sampling with diffusion models (see Section II-B) can be leveraged for reconstruction. Next, we discuss the choice of prior for 3D ultrasound data and detail the training process of the diffusion model in Section III-B. Finally, we explore strategies to incorporate temporal information during inference (Section III-C) and assess the models' uncertainty in the interpolated results (Section III-D).

A. Interpolation of 3D Ultrasound

As discussed in Section II-A, the fully sampled 3D ultrasound volume is given by $\mathbf{X}$, while the partially observed volume, following the Bayesian framework, is denoted by $\mathbf{Y}$. Using their vectorized forms, $\mathbf{x}$ and $\mathbf{y}$, respectively, the inverse problem of inpainting missing data can be formulated as:

$$\mathbf{y} = \boldsymbol{A}\mathbf{x}, \tag{5}$$

where the binary measurement matrix $\boldsymbol{A} \in \{0,1\}^{m\times n}$ with dimensions $d = N_{\text{el}} \times N_{\text{az}} \times N_{\text{ax}}$ and $m < n$, defines the acceleration rate as $r = n/m$. The measurement matrix is related to the element-wise masking operator $\boldsymbol{M}$ as follows:

$$\boldsymbol{M} = \text{diag}(\boldsymbol{A}^\top \boldsymbol{A}) \tag{6}$$

which allows us to express the backprojected (zero-filled) observations as:

$$\mathbf{y}_{\text{zf}} = \boldsymbol{A}^\top \mathbf{y} = \boldsymbol{M} \odot \mathbf{x}, \tag{7}$$

which we will use later mainly for visualization of both measurements and mask in the full image space. Since (5) is underdetermined, inferring $\mathbf{x}$ from $\mathbf{y}$ is inherently challenging. There are many possible solutions that could explain the observed data. To address this, strong prior knowledge about the structure of $\mathbf{x}$ must be incorporated, guiding the reconstruction towards a plausible solution. This motivates the use of deep generative models, specifically *posterior sampling* with diffusion models. Given the linear forward model in (5), the posterior distribution $p(\mathbf{x}|\mathbf{y})$ can be sampled through simple application of Bayes' rule for scores:

$$\underbrace{\nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau|\mathbf{y})}_{\text{posterior}} = \underbrace{\nabla_{\mathbf{x}_\tau} \log p(\mathbf{y}|\mathbf{x}_\tau)}_{\text{likelihood}} + \underbrace{\nabla_{\mathbf{x}_\tau} \log p(\mathbf{x}_\tau)}_{\text{prior}} \tag{8}$$

while the prior is modeled through the score network (see Section II-B), the likelihood term is closed-form. The likelihood term ensures that generated samples are consistent with the measurement and is also referred to as *guidance* as it guides the diffusion process. This expression needs to be evaluated for all diffusion timesteps τ, which is typically intractable. There are several posterior sampling methods for diffusion models which circumvent this issue [39]. In this work, we will use the popular diffusion posterior sampling (DPS) method [40], which interleaves prior updates (i.e. denoising) with guidance steps (gradient step towards the measurement). We now derive the corresponding expression for the time-dependent log-likelihood score used to guide the diffusion process via DPS, based on the formulation in (8):

$$\begin{aligned}
\nabla_{\mathbf{x}_\tau} \log p(\mathbf{y}|\mathbf{x}_\tau) &\approx \nabla_{\mathbf{x}_\tau} \log p\left(\mathbf{y} \mid \mathbf{x}_{0|\tau}\right) && (9)\\
&= -\frac{1}{2\sigma_{\mathbf{n}}^2} \nabla_{\mathbf{x}_\tau} \left\|\mathbf{y} - \boldsymbol{A}\mathbf{x}_{0|\tau}\right\|_2^2 && (10)\\
&= -\frac{1}{2\sigma_{\mathbf{n}}^2} \nabla_{\mathbf{x}_\tau}^\top \mathbf{x}_{0|\tau} \boldsymbol{A}^\top \left(\mathbf{y} - \boldsymbol{A}\mathbf{x}_{0|\tau}\right) && (11)\\
&\approx -\underbrace{\gamma}_{\substack{\text{Guidance}\\ \text{strength}}} \underbrace{\left(\boldsymbol{I} - \sigma_\tau \nabla_{\mathbf{x}_\tau} \epsilon_\theta(\mathbf{x}_\tau, \tau)\right)^\top \boldsymbol{A}^\top}_{\text{Projection}} \underbrace{\left(\mathbf{y} - \boldsymbol{A}\mathbf{x}_{0|\tau}\right)}_{\substack{\text{Measurement}\\ \text{error}}} && (12)
\end{aligned}$$

where (9) uses the Tweedie's estimate from (3) to guarantee tractability [39], [40]. In (10), we relax the delta likelihood (as follows from our noiseless forward model (5)) as a Gaussian $\mathcal{N}(\mathbf{y}; \boldsymbol{A}\mathbf{x}, \sigma_{\mathbf{n}}^2 \boldsymbol{I})$. Additionally, in (12), we use (3) again and subsequently substitute the actual score for the trained score model, introducing the final approximation. Lastly, as is commonly done, we depart from the theoretical guidance strength by reweighing the likelihood with hyperparameter γ in (12). An overview of the full algorithm for interpolating 3D ultrasound data, combining both diffusion (prior) and guidance (likelihood) steps, is shown in Algorithm 1, also incorporating design choices detailed in the following two Sections III-B and III-C

Algorithm 1 3D Ultrasound Interpolation using DMs

Require: subsampling rate $r \in (0,1]$, partial volume $\mathbf{Y} \in \mathbb{R}^{N_{\text{el}}\times N_{\text{az}}\times N_{\text{ax}}}$, measurement matrix $\boldsymbol{A} \in \{0,1\}^{(r\cdot N_{\text{el}})\times N_{\text{el}}}$, score model $\epsilon_\theta(\cdot)$, guidance strength γ, diffusion steps $\mathcal{T}$, noise schedule α_τ, σ_τ for $\tau \in [0, \mathcal{T}]$ (discretized)
Optional: previous reconstruction $\mathbf{X}^{\text{prev}}$, accelerated step τ_{init}
Ensure: reconstructed volume $\mathbf{X} \in \mathbb{R}^{N_{\text{el}}\times N_{\text{az}}\times N_{\text{ax}}}$

Initialize $\mathbf{X} \leftarrow \mathbf{0} \in \mathbb{R}^{N_{\text{el}}\times N_{\text{az}}\times N_{\text{ax}}}$
Define $\{(\mathbf{y}, i)\}$ as B plane slices and their indices from $\mathbf{Y}$
▷ Looping through volume (in parallel)
for each $(\mathbf{y}, i)$ in parallel **do**
  if $\mathbf{X}^{\text{prev}}$ available **then**
    $\mathbf{x}_0 \leftarrow \text{slice}(\mathbf{X}^{\text{prev}}, i)$ ▷ SeqDiff start [33]
    $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $\mathbf{x}_\tau \leftarrow \alpha_{\tau_{\text{init}}} \mathbf{x}_0 + \sigma_{\tau_{\text{init}}} \epsilon$ ▷ Forward diffusion
  else
    $\mathbf{x}_\tau \leftarrow \mathbf{x}_\mathcal{T} \sim \mathcal{N}(\mathbf{0}, \sigma_\mathcal{T}^2 \mathbf{I})$ ▷ Cold start
    $\tau_{\text{init}} \leftarrow \mathcal{T}$
  ▷ Reverse (guided) diffusion loop
  for $\tau = \tau_{\text{init}}$ to 0 **do**
    $\epsilon \leftarrow \epsilon_\theta(\mathbf{x}_\tau, \tau)$ ▷ Predict noise
    $\mathbf{x}_{0|\tau} \leftarrow \frac{1}{\alpha_\tau}(\mathbf{x}_\tau - \sigma_\tau \epsilon)$ ▷ Denoise (prior)
    $\mathcal{M}_\tau \leftarrow \mathbf{y} - \boldsymbol{A}\mathbf{x}_{0|\tau}$ ▷ Measurement error
    $\mathcal{P}_\tau \leftarrow \left(\mathbf{I} - \sigma_\tau \nabla_{\mathbf{x}_\tau} \epsilon_\theta(\mathbf{x}_\tau, \tau)\right)^T \boldsymbol{A}^T$ ▷ Projection
    $\mathbf{x}_{0|\tau} \leftarrow \mathbf{x}_{0|\tau} - \gamma \mathcal{P}_\tau \mathcal{M}_\tau$ ▷ Guidance step
    $\mathbf{x}_\tau \leftarrow \alpha_\tau \mathbf{x}_{0|\tau} + \sigma_\tau \epsilon$ ▷ Forward diffusion
  $\mathbf{X} \leftarrow \text{insert}(\mathbf{X}, \mathbf{x}_0, i)$ ▷ Insert result into i'th B plane
return $\mathbf{X}$

B. Training and Choice of Prior

Selecting an appropriate representation for learning 3D ultrasound priors is nontrivial, particularly when considering all possible design options for how to present the data to the diffusion model. A seemingly straightforward approach is to employ a fully 3D generative model to learn the 3D prior; however, this introduces several challenges. First, as the dimensionality of both the data and model increases, the amount of training data required to capture the distribution grows exponentially [41], making 3D generative models highly *data hungry*. Second, 3D ultrasound data exhibits strong spatial and temporal correlations, which can be leveraged to improve efficiency when transitioning from 2D to 3D representations.

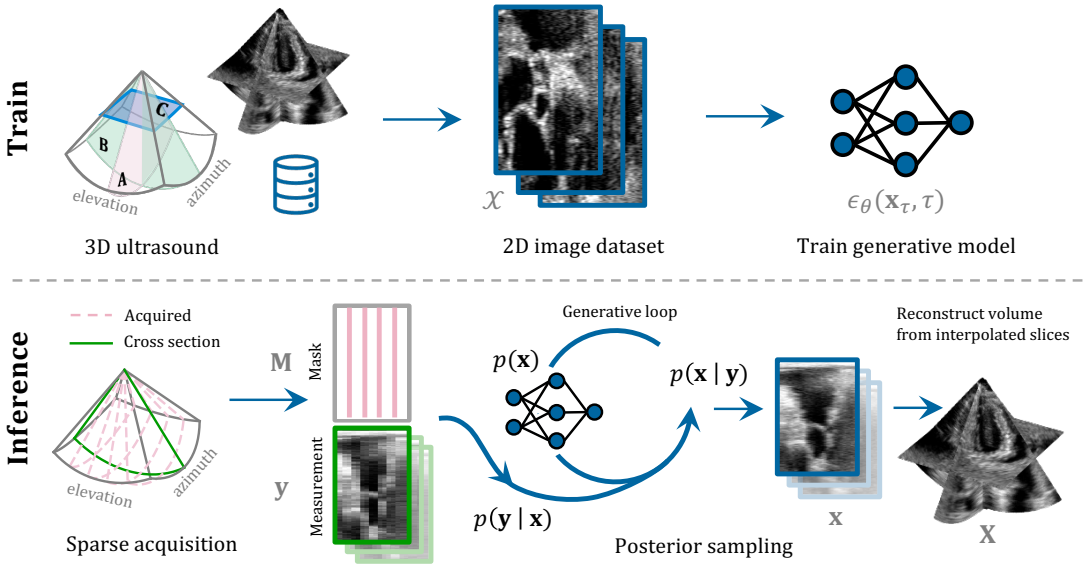

Fig. 3: Overview of the training and inference pipeline for diffusion-based interpolation of 3D ultrasound volumes. A 2D score-based prior is learned from fully sampled elevation (B-plane) slices extracted from the training dataset. At inference time, this generative prior guides the posterior sampling process to reconstruct from subsampled B-planes. See Algorithm 1 for details of the proposed interpolation method using DMs.

Lastly, 3D models impose significant computational demands, requiring substantial GPU memory and extended training times, which limits their feasibility in practical applications.

To mitigate these challenges, we adopt a strategy that utilizes 2D models for learning a prior across cross-sections (B-planes). Prior work has demonstrated the success of such approaches for high-dimensional data, including 3D [42] as well as sequence data [33], [43]. Additionally, 2D models and 2D ultrasound datasets are more readily available, making it easier to leverage pretrained models and integrate them into the proposed framework. Therefore, given the fully sampled 3D ultrasound volume $\mathbf{X}$ as defined in (1), we construct a dataset $\mathcal{X} = \left\{\mathbf{x}_{\text{el}}^{(1)}, \mathbf{x}_{\text{el}}^{(2)}, \ldots, \mathbf{x}_{\text{el}}^{(|\mathcal{X}|)}\right\} \sim p(\mathbf{x}_{\text{el}})$, which is comprised of elevation slices (B plane). In relation to the full volume $\mathbf{X}$, the slice is highlighted in green as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,2} & \cdots & \mathbf{x}_{1,N_{\text{az}}} \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & \cdots & \mathbf{x}_{2,N_{\text{az}}} \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{\mathbf{x}_{N_{\text{el}},1}}_{\mathbf{x}_{\text{el}(1)}} & \mathbf{x}_{N_{\text{el}},2} & \cdots & \mathbf{x}_{N_{\text{el}},N_{\text{az}}} \end{bmatrix}. \tag{13}$$

Moreover, we use log-compressed image data (B-mode) in the native polar coordinate system, as defined in (II-A). For visualization purposes, all images presented in this paper are shown after scan-conversion to a Cartesian grid, allowing the anatomy to be displayed in its true physical coordinate system.

Our dataset consists of 100 *in vivo* volumetric cardiac cine-loops, recorded using an X5-1C matrix phased array transducer connected to a Philips EPIQ scanner. The volumes were collected from various sites and scanning sessions involving 16 patients and contain, on average, 40 frames, spanning at least one full cardiac cycle. Approximately 10% of the dataset is reserved for validation and testing, with one cine-loop designated for validation and hyperparameter tuning, and the remaining seven cine-loops from three patients used for testing.

The diffusion model is trained using the denoising score matching objective from (4), with a time-conditioned U-Net ($\approx$3.9 M parameters) with sinusoidal embeddings [44] as the backbone architecture. We optimize the model using AdamW with a learning rate of $1\text{e}{-4}$ and weight decay of $1\text{e}{-4}$. Finally, we turn on exponential moving average (EMA) of 0.999, minimizing an MSE loss over approximately 25 epochs.

C. Temporal consistency

Ensuring temporal consistency across frames is crucial for high-quality 3D ultrasound reconstruction. By explicitly incorporating temporal information, we can produce more coherent and stable reconstructions over time, reducing flicker and improving clinical interpretability. To achieve this, we build upon SeqDiff [33], which exploits temporal continuity by initializing the current frame's diffusion process from the previous frame's reconstruction. Rather than starting the generative process from pure noise at the beginning of each frame, SeqDiff warm-starts the diffusion at an intermediate step along the reverse trajectory, effectively integrating temporal consistency into the sampling process. Formally, SeqDiff initializes the diffusion process at some intermediate step τ' along the reverse diffusion trajectory, where $0 < \tau' \ll \mathcal{T}$, effectively reducing the number of iterations. Rather than beginning each generation from scratch at $\tau = \mathcal{T}$ using a pure Gaussian sample $\mathbf{x}_\mathcal{T} \sim \mathcal{N}(\mathbf{0}, \sigma_\mathcal{T}^2 \boldsymbol{I})$, SeqDiff initializes the process using a previous solution $\mathbf{x}_0^{t-1}$, forward-diffused to the current τ'.

Specifically, we reconstruct from a sequence of time-dependent measurements $\mathbf{y}^t = \boldsymbol{A}^t \mathbf{x}^t$, where $\boldsymbol{A}$ varies accord-

ing to the sparse interlocking acquisition scheme as described in Fig. 2 and $\mathbf{x}^t$ and $\mathbf{x}^{t-1}$ are assumed to exhibit strong structural similarity. By propagating prior reconstructions forward to an intermediate noise level and using them as initialization, we significantly reduce computation and improve temporal consistency in the reconstructions. This approach is well-suited for cardiac ultrasound, where, at high frame rates, the anatomy changes gradually over time. Thereby, the previous frame's reconstruction serves as an effective prior, allowing each newly acquired set of elevation planes to benefit from earlier information.

D. Uncertainty Quantification

Many interpolation methods provide only a single *point-estimate* reconstruction, failing to represent the uncertainty inherent to the inverse problem. Given the probabilistic nature of DPS, however, our method can quantify the uncertainty present in its reconstructions. Quantifying the uncertainty can help mitigate *hallucinations*, in which a model generates one of many plausible anatomies, which may be convincing but not truthful. Or, more critically, overlooks an anatomical feature that is in fact present. In what follows, we describe two methods to quantify this uncertainty and communicate it to the user.

The posterior sampling approach as discussed in Section III-A allows us to draw N samples $\{\mathbf{x}^{(i)}\}_{i=0}^{N} \sim p(\mathbf{x}|\mathbf{y})$ from the posterior, inherently providing an estimate of uncertainty of the interpolated result. Luckily, drawing multiple posterior samples can be executed in parallel on the GPU, minimizing the additional computational expense. We propose two methods for visualizing the model's uncertainty, along with an outlook for future work. An example of both methods for various acceleration rates is shown in Fig. 5.

1) Variance Heat Map: By analyzing the pixel-wise variance across posterior samples, we can assess which regions exhibit higher or lower certainty, where $\bar{\mathbf{x}}$ is the sample mean:

$$\sigma_{\mathbf{x}}^2 = \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])^2\right] \approx \frac{1}{N}\sum_i \left[(\mathbf{x}^{(i)} - \bar{\mathbf{x}})^2\right]. \tag{14}$$

For instance, directly acquired scan lines are well-constrained by the measurement model and therefore exhibit minimal variance. In contrast, non-acquired scan lines rely on the generative model for reconstruction, where higher uncertainty manifests as greater variance in the posterior samples.

2) Composite Image: An alternative approach to uncertainty visualization involves creating a composite image by sampling pixels randomly from each posterior sample.

$$\forall j \in \big[0, \text{len}(\mathbf{x})\big) : \tilde{\mathbf{x}}_j = \mathbf{x}_j^{(i)}, \quad \text{with } i \sim \text{Uniform}(1, N) \tag{15}$$

This introduces perceivable noise in uncertain regions while maintaining a consistent appearance in regions where the samples agree. The rationale for this presentation is that, intuitively, the operator will naturally place less confidence in the regions that appear noisy, as noisier areas indicate greater model uncertainty.

Finally, beyond visualization techniques, it is crucial to recognize the inherent limitations of the model. In particular, it is important to consider that, in unobserved regions, the posterior samples will only vary according to the distribution that has been learned by the model, and potentially fail to include valid reconstructions that fall outside the training distribution. It is therefore essential to understand how much subsampling remains acceptable for a given application. In the following experiments section, we analyze the model's performance across various tasks and acquisition rates to determine the trade-offs between image quality and subsampling levels.

IV. Experiments

We perform both qualitative and quantitative analyses to assess the benefit of advanced interpolation methods for 3D ultrasound reconstruction. The most straightforward comparison investigates image quality directly through the use of common image quality metrics. We leverage both (pixel-wise) distortion and (global) perceptual metrics. Finally, we perform an experiment on speckle tracking, a common clinical downstream task in cardiac imaging, to further validate our approach. All results are generated using parameters $\mathcal{T} = 200, \tau' = 50, \gamma = 35$, which were optimized for the validation split as outlined in Section III-B.

A. Baselines

In every experiment, we consider the fully sampled volume as the ground truth. This allows for both quantitative evaluation of reconstruction quality and comparison of downstream task performance between interpolated and fully acquired volumes. Image quality metrics are computed in the original polar domain, where interpolation is performed, to avoid confounding effects from scan conversion, which itself applies interpolation.

1) Linear and nearest-neighbor interpolation: The most straightforward methods for volumetric interpolation in 3D ultrasound are nearest-neighbor and linear interpolation along the elevation axis. We include these traditional techniques as simple baselines for reference. Nearest-neighbor fills missing slices by duplicating the nearest acquired slice, effectively visualizing the sparse measurements in the full volume. This approach introduces blocky artifacts and discontinuities. Linear interpolation averages neighboring slices to produce smoother transitions but often results in blurring and loss of detail. Both methods are low in complexity and easily implemented in clinical systems. However, they lack anatomical awareness, limiting their effectiveness at higher acceleration rates. While they may suffice at lower rates, doing so constrains the potential temporal resolution gains targeted in this work.

2) Supervised learning: Additionally, we include a baseline that trains a deep learning network in a fully supervised, end-to-end manner for inpainting 3D ultrasound. Specifically, we employ the same U-Net architecture as used by the diffusion model of the proposed approach (see Section III-B), and focus on the inpainting of individual elevation slices, maintaining the same problem scope. Unlike the diffusion model, which leverages an explicit forward model during sampling, this baseline does not explicitly incorporate forward model knowledge during inference. Instead, it learns a data-driven mapping

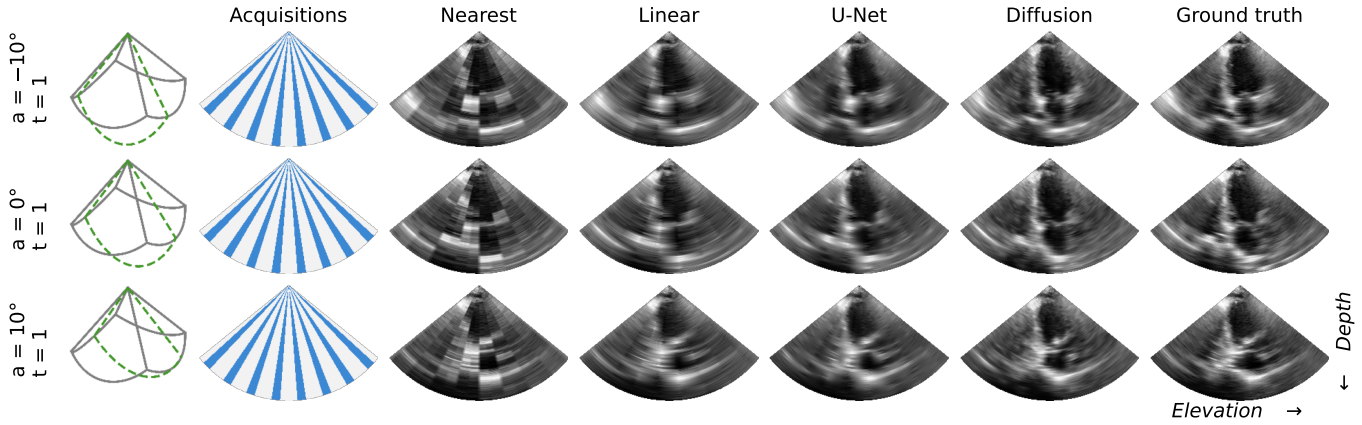

Fig. 4: Qualitative comparison of the proposed Diffusion-based interpolation method with baseline approaches (Nearest, Linear, and U-Net). The figure displays three B-plane slices from a single 3D volume, with their positions within the frustum geometry illustrated on the left. Acceleration rate is set to $r = 3$.

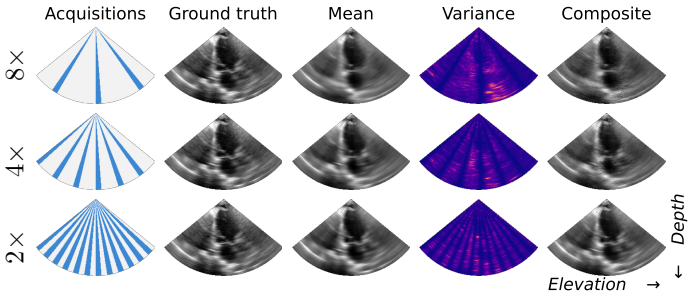

Fig. 5: The posterior mean, variance, and composite images for reconstructions with fractions of the elevation planes acquired $r = (8\times, 4\times, 2\times)$. It is clear that the posterior samples vary more in the unmeasured regions of tissue, and that acquiring more elevation planes reduces the overall uncertainty and boosts the reconstruction quality.

$\mathbf{x} = f_\theta(\mathbf{y}_{\text{zf}})$ from zero-filled input measurements $\mathbf{y}_{\text{zf}}$, see (7), to complete reconstructions. Consequently, the network's performance may degrade on unseen mask patterns, in contrast to the diffusion model's task-agnostic approach.

B. Image quality

Fig. 4 shows a visual comparison of the proposed interpolation method with respect to the baselines. Additionally, we evaluate the interpolated volumes using peak signal-to-noise ratio (PSNR) as a pixel-wise distortion metric and learned perceptual image patch similarity (LPIPS) [45] to measure perceptual quality, assessing both aspects of the perception-distortion tradeoff [46]. Fig. 6 shows the image quality of each interpolation method with respect to the fully sampled and reconstructed 3D volume. The LPIPS is a 2D image metric, which we apply to all the B-planes separately. Unlike the diffusion-based method, all baseline methods come with significant visible interpolation artifacts, which are clearly reflected by the perceptual metric. This trend continues in Fig. 7 and Fig. 8, which show distortion and perceptual image quality performance against various subsampling factors.

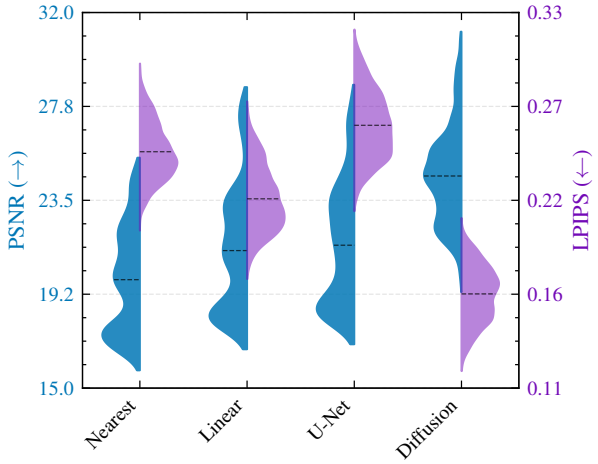

Fig. 6: Image quality for various 3D interpolation methods. Testing for both distortion (PSNR) (↑) (blue) and perceptual (LPIPS) (↓) (purple) metrics. Acceleration rate is set to $r = 6$.

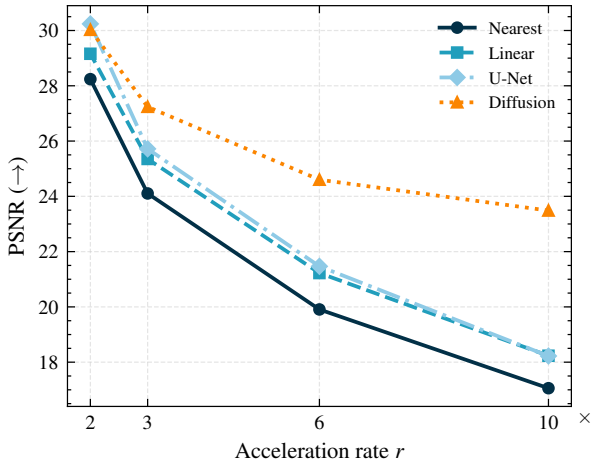

Fig. 7: Reconstruction performance evaluated using a distortion metric (PSNR) across various interpolation methods at increasing acceleration rates $r \in \{2, 3, 6, 10\}$.

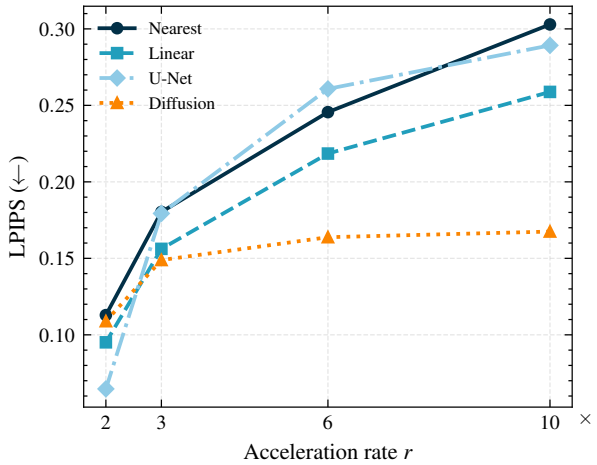

Fig. 8: Reconstruction performance evaluated using a perceptual metric (LPIPS) across various interpolation methods at increasing acceleration rates $r \in \{2, 3, 6, 10\}$.

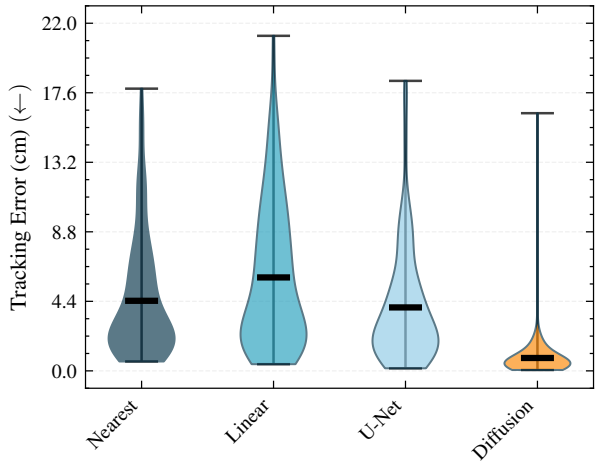

Fig. 9: Speckle tracking performance comparison across various interpolation methods at a acceleration rate of $r = 6$. The proposed method shows significant improvement over baseline approaches.

C. Speckle tracking

Speckle tracking is used to assess tissue motion and deformation by tracking speckle patterns across frames [47], e.g. enabling strain measurements for evaluating myocardial function. While 2D speckle tracking is commonly used, it is limited by out-of-plane motion and foreshortening artifacts. Volumetric ultrasound imaging mitigates these limitations by providing full 3D motion tracking, improving robustness and accuracy. To assess the impact of different reconstruction methods on motion estimation, we evaluate speckle tracking performance on both ground truth and interpolated data. As shown in Fig. 9, the proposed method demonstrates a noticeable reduction in tracking error compared to the other approaches. This can be attributed to the integrated temporal consistency (SeqDiff) of the proposed method as well as improved fidelity of the reconstructions.

D. Out-of-distribution data

Our final experiment evaluates the robustness of the proposed method to *out-of-distribution* (OoD) data. As discussed in Section III-D, a concern with generative models is their tendency to *hallucinate* structures or overlook important anatomical features when faced with ambiguous or unseen data. This experiment specifically investigates such failure modes by introducing synthetic anomalies into the test set.

We augment the test volumes with bright masses, which are simulated as circular inclusions of fixed diameter 5px and uniform brightness (70%), randomly positioned within the volume. These inclusions are clearly OoD, as perfectly circular, high-contrast structures are not present in the training data. Importantly, this setup provides access to ground truth inclusion locations, allowing for quantitative evaluation. After subsampling and reconstruction, we compute the pixel-level recall of the inserted inclusions. A pixel is considered correctly recalled if its intensity lies within 10% of the true inclusion value. The results are averaged over 10 randomly placed inclusions per configuration. See Fig. 10 for the recall performance across acceleration rates, as well as a visual example.

Interestingly, our method demonstrates consistently higher recall than baseline approaches, especially under strong subsampling. We attribute this to two key factors. First, the explicit conditioning on measured data during inference ensures consistency with the acquired measurements, improving reliability under distribution shifts. This contrasts with the supervised baselines trained end-to-end without an explicit measurement model. Second, by initialization of each reconstruction with the previous frame via SeqDiff, we do not only accelerate inference but also enhance robustness to anomalies by leveraging temporal continuity in the data.

V. Discussion

This work demonstrates the effectiveness of DM-based interpolation for reconstructing volumetric ultrasound from sparsely acquired elevation planes. This section outlines key implications of our results, including benefits of the probabilistic formulation, temporal consistency, and potential for real-time deployment. We also discuss considerations for integration into clinical workflows and future adaptive imaging systems.

A key strength of our method lies in its probabilistic formulation. The posterior sampling framework not only enhances robustness to OoD data but also yields uncertainty estimates that can inform downstream tasks or guide acquisition. This is especially relevant for integration with active imaging systems [19], [48], [49], where uncertainty can drive intelligent transmit selection to minimize ambiguity in real time. In this work, we focus on reconstruction from a fixed acquisition pattern, but our approach lays important groundwork for closed-loop, adaptive imaging pipelines.

Temporal consistency is another benefit, achieved by incorporating SeqDiff into the sampling process. The integration of prior information stabilizes the generative process over time, enabling more coherent predictions and reducing redundant computations. In practice, we observe it improves both perceptual quality and robustness under aggressive subsampling.

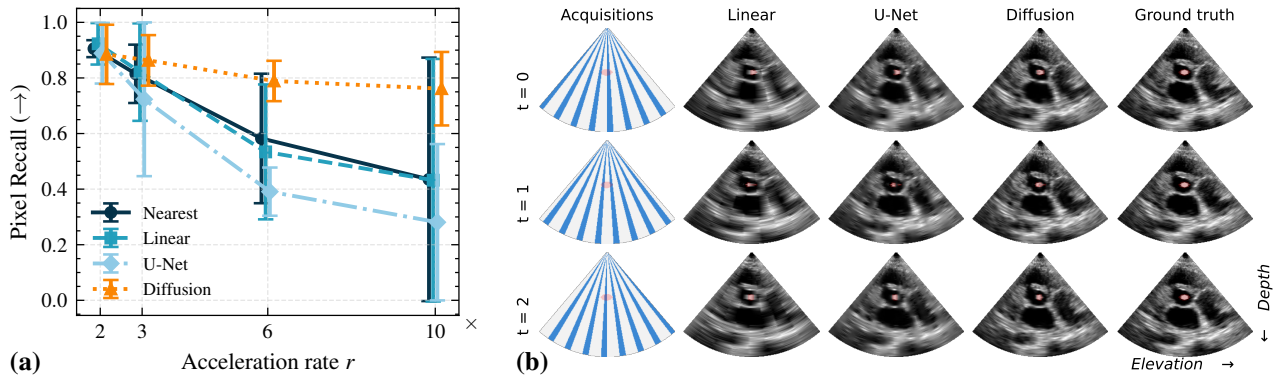

Fig. 10: **(a)** Pixel recall for synthetic out-of-distribution inclusions at varying acceleration rates $r \in \{2, 3, 6, 10\}$, averaged over 10 randomly inserted inclusions with 95% confidence intervals. **(b)** Example of three consecutive frames ($r = 3$), with random inclusions highlighted in red ◉. When the inclusion is not fully covered by an acquisition, baseline methods recover less of the circle, whereas the proposed method achieves better reconstruction due to built-in temporal consistency.

A critical aspect is the feasibility of real-time deployment. While diffusion models are inherently iterative, our current implementation achieves reconstruction times on the order of several ms/step, with convergence typically reached in 20–50 steps. All elevation planes can be reconstructed in parallel on modern GPUs, enabling near real-time throughput for full-volume reconstruction. Further acceleration could be achieved using model distillation [34], latent diffusion models [50], or advanced sampling strategies [51].

Finally, even in the absence of real-time operation, the method is well suited for offline clinical workflows. In many cases, only a 2D view is presented on display even with 3D imaging turned on. Similarly, applications like patch ultrasound [52] do not require any real-time imaging. In such cases, low-resolution previews can guide acquisition, while high-quality reconstructions are generated afterward for diagnostic interpretation. This deferred processing model aligns with existing practices in the clinical workflow, where clinicians often analyze stored image sequences after a scanning session.

VI. Conclusion

3D ultrasound holds significant promise for the future of ultrasound imaging, particularly in cardiac applications. However, its clinical adoption remains limited due to constraints in temporal and spatial resolution, often leading clinicians to fall back to 2D imaging. To push this trade-off, one can employ spatial subsampling to reduce the number of transmit events, thereby enabling higher frame rates or broader coverage. Yet, conventional interpolation methods, as well as supervised learning-based approaches, struggle to fully reconstruct undersampled volumes, especially at high acceleration rates. In this work, we reframe 3D ultrasound reconstruction as a Bayesian inverse problem, leveraging deep generative priors to substantially improve reconstruction quality. Our approach demonstrates clear benefits not only in terms of visual fidelity but also in clinically relevant downstream tasks such as speckle tracking. Moreover, by extending our diffusion-based interpolation framework to exploit the sequential nature of ultrasound data, we achieve both improved temporal consistency and significant acceleration, bringing the method closer to real-time applicability. Finally, the probabilistic nature of our model enables estimation of reconstruction uncertainty, offering a valuable tool for clinical decision-making and a foundation for future work on uncertainty-aware imaging and adaptive acquisition strategies [19].

References

[1] G. D. Stetten, T. Ota, C. J. Ohazama, C. Fleishman, J. Castellucci, J. Oxaal, T. Ryan, J. Kisslo, and O. v. Ramm, "Real-time 3d ultrasound: A new look at the heart," *Journal of Cardiovascular Diagnosis and Procedures*, vol. 15, no. 2, pp. 73–84, 1998.

[2] T. R. Nelson and D. H. Pretorius, "Three-dimensional ultrasound imaging," *Ultrasound in medicine & biology*, vol. 24, no. 9, pp. 1243–1270, 1998.

[3] A. Fenster, G. Parraga, and J. Bax, "Three-dimensional ultrasound scanning," *Interface focus*, vol. 1, no. 4, pp. 503–519, 2011.

[4] J. Pedrosa, D. Barbosa, N. Almeida, O. Bernard, J. Bosch *et al.*, "Cardiac chamber volumetric assessment using 3d ultrasound-a review," *Current pharmaceutical design*, vol. 22, no. 1, pp. 105–121, 2016.

[5] V. C.-C. Wu and M. Takeuchi, "Three-dimensional echocardiography: current status and real-life applications," *Acta Cardiologica Sinica*, vol. 33, no. 2, p. 107, 2017.

[6] X. Papademetris, A. J. Sinusas, D. P. Dione, and J. S. Duncan, "3d cardiac deformation from ultrasound images," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI'99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings 2*. Springer, 1999, pp. 420–429.

[7] O. Catalano, R. Fusco, F. De Muzio, I. Simonetti, P. Palumbo, F. Bruno, A. Borgheresi, A. Agostini, M. Gabelloni, C. Varelli *et al.*, "Recent advances in ultrasound breast imaging: from industry to clinical practice," *Diagnostics*, vol. 13, no. 5, p. 980, 2023.

[8] A. Fenster, K. Surry, W. Smith, J. Gill, and D. B. Downey, "3d ultrasound imaging: applications in image-guided therapy and biopsy," *Computers & Graphics*, vol. 26, no. 4, pp. 557–568, 2002.

[9] E. Spitzer, B. Ren, F. Zijlstra, N. M. Van Mieghem, and M. L. Geleijnse, "The role of automated 3d echocardiography for left ventricular ejection fraction assessment," *Cardiac failure review*, vol. 3, no. 2, p. 97, 2017.

[10] M. De Craene, S. Marchesseau, B. Heyde, H. Gao, M. Alessandrini, O. Bernard, G. Piella, A. R. Porras, L. Tautz, A. Hennemuth *et al.*, "3d strain assessment in ultrasound (straus): A synthetic comparison of five tracking methodologies," *IEEE transactions on medical imaging*, vol. 32, no. 9, pp. 1632–1646, 2013.

[11] H. de Hoop, E. Maas, J.-W. Muller, H.-M. Schwab, and R. Lopata, "3-d motion tracking and vascular strain imaging using bistatic dual aperture ultrasound acquisitions," *Physics in Medicine & Biology*, vol. 70, no. 4, p. 045013, 2025.

[12] Q. Huang and Z. Zeng, "A review on real-time 3d ultrasound imaging technology," *BioMed research international*, vol. 2017, no. 1, p. 6027029, 2017.

[13] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, 2012.

[14] A. Burshtein, M. Birk, T. Chernyakova, A. Eilam, A. Kempinski, and Y. C. Eldar, "Sub-nyquist sampling and fourier domain beamforming in volumetric ultrasound imaging," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 63, no. 5, pp. 703–716, 2016.

[15] O. Lorintiu, H. Liebgott, M. Alessandrini, O. Bernard, and D. Friboulet, "Compressed sensing reconstruction of 3d ultrasound data using dictionary learning and line-wise subsampling," *IEEE transactions on medical imaging*, vol. 34, no. 12, pp. 2467–2477, 2015.

[16] I. A. Huijben, B. S. Veeling, K. Janse, M. Mischi, and R. J. van Sloun, "Learning sub-sampling and signal recovery with applications in ultrasound imaging," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3955–3966, 2020.

[17] S. Ji, D. W. Roberts, A. Hartov, and K. D. Paulsen, "Real-time interpolation for true 3-dimensional ultrasound image volumes," *Journal of Ultrasound in Medicine*, vol. 30, no. 2, pp. 243–252, 2011.

[18] R. J. Van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2019.

[19] R. J. Van Sloun, "Active inference and deep generative modeling for cognitive ultrasound," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.

[20] T. S. W. Stevens, J. Overdevest, O. Nolan, W. L. van Nierop, R. J. G. van Sloun, and Y. C. Eldar, "Deep generative models for bayesian inference on high-rate sensor data: Applications in automotive radar and medical imaging," *Philosophical Transactions A*, 2025.

[21] X. He, Y. Lei, Y. Liu, Z. Tian, T. Wang, W. J. Curran, T. Liu, and X. Yang, "Deep attentional gan-based high-resolution ultrasound imaging," in *Medical Imaging 2020: Ultrasonic Imaging and Tomography*, vol. 11319. SPIE, 2020, pp. 61–66.

[22] X. Dai, Y. Lei, T. Wang, M. Axente, D. Xu, P. Patel, A. B. Jani, W. J. Curran, T. Liu, and X. Yang, "Self-supervised learning for accelerated 3d high-resolution ultrasound imaging," *Medical Physics*, vol. 48, no. 7, pp. 3916–3926, 2021.

[23] D. Wulff, T. Dohnke, N. T. Nguyen, and F. Ernst, "Towards realistic 3d ultrasound synthesis: deformable augmentation using conditional variational autoencoders," in *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2023, pp. 821–826.

[24] D. Stojanovski, U. Hermida, P. Lamata, A. Beqiri, and A. Gomez, "Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation," in *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer, 2023, pp. 34–43.

[25] M. Domínguez, Y. Velikova, N. Navab, and M. F. Azampour, "Diffusion as sound propagation: Physics-inspired model for ultrasound image generation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 613–623.

[26] B. Freiche, A. El-Khoury, A. Nasiri-Sarvi, M. S. Hosseini, D. Garcia, A. Basarab, M. Boily, and H. Rivaz, "Ultrasound image generation using latent diffusion models," *arXiv preprint arXiv:2502.08580*, 2025.

[27] T. S. W. Stevens, H. van Gorp, F. C. Meral, J. Shin, J. Yu, J. luc Robert, and R. V. Sloun, "Removing structured noise using diffusion models," *Transactions on Machine Learning Research*, 2025. [Online]. Available: https://openreview.net/forum?id=BvKYsaOVEn

[28] T. S. W. Stevens, F. C. Meral, J. Yu, I. Z. Apostolakis, J.-L. Robert, and R. J. Van Sloun, "Dehazing ultrasound using diffusion models," *IEEE Transactions on Medical Imaging*, 2024.

[29] H. Asgariandehkordi, S. Goudarzi, M. Sharifzadeh, A. Basarab, and H. Rivaz, "Denoising plane wave ultrasound images using diffusion probabilistic models," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2024.

[30] Y. Zhang, C. Huneau, J. Idier, and D. Mateus, "Ultrasound image reconstruction with denoising diffusion restoration models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 193–203.

[31] A. Durrer, J. Wolleb, F. Bieder, P. Friedrich, L. Melie-Garcia, M. A. Ocampo Pineda, C. I. Bercea, I. E. Hamamci, B. Wiestler, M. Piraud *et al.*, "Denoising diffusion models for 3d healthy brain tissue inpainting," in *MICCAI Workshop on Deep Generative Models*. Springer, 2024, pp. 87–97.

[32] S. W. Penninga, H. van Gorp, and R. J. van Sloun, "Deep sylvester posterior inference for adaptive compressed sensing in ultrasound imaging," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.

[33] T. S. W. Stevens, O. Nolan, J.-L. Robert, and R. J. van Sloun, "Sequential Posterior Sampling with Diffusion Models," in *2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025.

[34] C. Meng, R. Rombach, R. Gao, D. Kingma, S. Ermon, J. Ho, and T. Salimans, "On distillation of guided diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 297–14 306.

[35] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[36] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[37] P. Milanfar and M. Delbracio, "Denoising: A powerful building-block for imaging, inverse problems, and machine learning," *arXiv preprint arXiv:2409.06219*, 2024.

[38] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[39] G. Daras, H. Chung, C.-H. Lai, Y. Mitsufuji, J. C. Ye, P. Milanfar, A. G. Dimakis, and M. Delbracio, "A survey on diffusion models for inverse problems," *arXiv preprint arXiv:2410.00083*, 2024.

[40] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[41] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, vol. 1, 2000, pp. 4–8.

[42] H. Chung, D. Ryu, M. T. McCann, M. L. Klasky, and J. C. Ye, "Solving 3d inverse problems using pre-trained 2d diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 542–22 551.

[43] G. Daras, W. Nie, K. Kreis, A. G. Dimakis, M. Mardani, N. B. Kovachki, and A. Vahdat, "Warped diffusion: Solving video inverse problems with image diffusion models," in *Advances in Neural Information Processing Systems*, A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, Eds., vol. 37. Curran Associates, Inc., 2024, pp. 101 116–101 143.

[44] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 1243–1252. [Online]. Available: https://proceedings.mlr.press/v70/gehring17a.html

[45] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.

[46] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6228–6237.

[47] D. Garcia, P. Lantelme, and E. Saloux, "Introduction to speckle tracking in cardiac ultrasound imaging," *Handbook of speckle filtering and tracking in cardiovascular ultrasound imaging and video. Institution of Engineering and Technology*, pp. 571–598, 2018.

[48] B. Federici, R. J. van Sloun, and M. Mischi, "Active inference for closed-loop transmit beamsteering in fetal doppler ultrasound," *arXiv preprint arXiv:2410.04869*, 2024.

[49] O. Nolan, T. S. W. Stevens, W. L. van Nierop, and R. V. Sloun, "Active diffusion subsampling," *Transactions on Machine Learning Research*, 2025. [Online]. Available: https://openreview.net/forum?id=OGifiton47

[50] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[51] X. Ma, G. Fang, and X. Wang, "Deepcache: Accelerating diffusion models for free," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 15 762–15 772.

[52] H. Hu, H. Huang, M. Li, X. Gao, L. Yin, R. Qi, R. S. Wu, X. Chen, Y. Ma, K. Shi *et al.*, "A wearable cardiac ultrasound imager," *Nature*, vol. 613, no. 7945, pp. 667–675, 2023.