

UBone3D: Physics-Rectified Conditional Flow Matching for Anatomical 3D Shape Completion from Ultrasound

Weiying Chen¹, Yuchong Gao², Siyuan Li¹, Marek Reformat¹, Rui Zheng², and Edmond Lou¹

¹ University of Alberta, Edmonton, AB, Canada
{weiying3,elou}@ualberta.ca

² ShanghaiTech University, Shanghai, China

Abstract. Three-dimensional ultrasound (US) is a safe, radiation-free complementary modality to CT and X-rays for longitudinal monitoring, yet its segmentation-derived partial point clouds are extremely artifact-laden. Consequently, it is challenging to recover a clean and complete anatomical structure from such US point clouds. In this paper, we present UBone3D, a novel framework based on physics-rectified conditional flow matching (CFM) that performs point cloud completion directly from partial US observations. UBone3D models deterministic physics artifacts (e.g., surface thickening, streaking, dropouts) via a simulated physics proxy, and introduces test-time physics rectification to steer the shape completion. At inference, the completion is jointly steered by two decoupled forces: (1) anatomical plausibility enforced by a CT-trained generative shape prior, **BoneFM**, and (2) physics consistency enforced by **USimNet** in the ultrasound formation space. Extensive experiments on simulated and in-vivo data demonstrate significant improvements in reconstruction accuracy and anatomical fidelity over existing baselines. Project page: <https://answertx.github.io/UBone3D-Proj/>.

Keywords: 3D Shape Completion · Physics-Guided Generation · Conditional Flow Matching · Ultrasound Bone Reconstruction

1 Introduction

Reconstructing accurate three-dimensional (3D) bone geometry directly from ultrasound scans remains a fundamental unsolved challenge in medical imaging. A robust solution would enable the next generation of radiation-free, real-time 3D applications, from intraoperative navigation to affordable long-term monitoring of skeletal deformities. This is especially important for adolescent idiopathic scoliosis (AIS), where patients often need repeated CT or X-rays over years, resulting in high cumulative radiation exposure. In current clinical practice, 3D spine models are typically generated by aggregating tracked 2D B-mode ultrasound frames according to their spatial poses and volumetrically interpolating the data using Voxel Nearest Neighbor (VNN) [7] for downstream assessment.

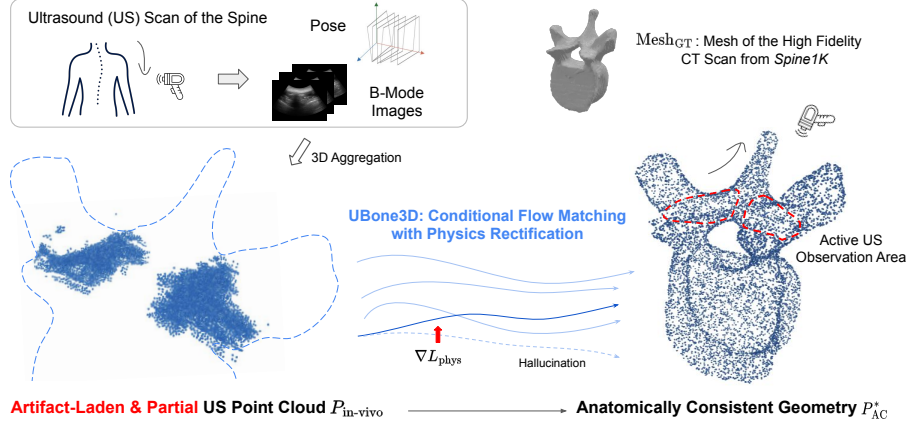


Fig. 1: Physics-rectified shape completion bridges the gap between ultrasound (US) observations and complete anatomy. (Left) Segmentation-derived in-vivo US point clouds suffer from severe acoustic artifacts and only partially cover the underlying bone anatomy. (Middle) UBone3D leverages conditional flow matching (CFM) with a differentiable physics proxy to steer the generative trajectory, constraining completion within the physical bounds of ultrasound formation. (Right) The reconstructed anatomical completion (P_{AC}) achieves global structural integrity while maintaining high local fidelity to the observed ultrasound signals (red regions).

VNN voxel stacking is a lossy discretization that blurs thin cortical surfaces and propagates physics-induced dropouts and reverberation into large empty or spurious regions, further obscuring the true bone geometry.

Ideally, the complete skeletal structure could be recovered from cleanly segmented partial structures; however, achieving perfect segmentation is often impractical in real clinical settings. In this paper, we choose the point-cloud representation to directly model and recover the underlying geometry from **imperfect segmentation-derived ultrasound input**. Recovering an anatomically clean and complete 3D structure from such input is inherently challenging: beyond the missing topology common to any partial observation, the artifact-laden ultrasound point cloud extracted from upstream segmentation exhibits a large gap relative to the desired anatomically complete geometry, as shown in Fig. 1. Importantly, this gap is *not merely one of random noise, but one of deterministic physics*: finite beamwidth, reverberation, and signal dropout caused by the probe’s limited aperture (FOV) together distort the observed signals, leading to incomplete and highly nonuniform point clouds.

Recent 3D reconstruction techniques for ultrasound imaging [9, 16, 37, 38] focused on recovering a clean bone surface of the superficial bone boundaries, assuming access to clean geometric or intensity data, neglecting that in-vivo ultrasound scans do not meet this standard. On the other hand, general 3D completion models were trained on idealized, clean geometric data [36, 42, 45, 48]. The foundational knowledge, or priors, used by these generative methods

often comes from datasets like ShapeNet [6] or natural images, giving them no understanding of the acoustic physics inherent to ultrasound.

In this paper, we introduce **UBone3D**, a novel physics-rectified conditional flow matching framework that casts the recovery of complete anatomical geometry from partial ultrasound as an observation-guided denoising and shape completion problem. The key insight is to decouple two complementary objectives, anatomical plausibility and physical consistency, which are respectively handled by BoneFM and USimNet and synergized during test-time inference. Our main contributions include:

- UBone3D enables the recovery of anatomically consistent complete 3D structures from artifact-laden, partial ultrasound observations by synergizing the generative anatomy prior (BoneFM) with test-time physical rectification, without relying on clean or perfectly segmented inputs.
- We introduce USimNet, a differentiable physics proxy that models acoustic artifacts (e.g., bone shadowing and streaks), supporting an explicit, gradient-based rectification mechanism that aligns the evolving generative trajectory with physical observation constraints during inference.
- Our approach demonstrates strong generalization to unseen viewing angles on simulated data, and remains robust on zero-shot in-vivo data, where pure geometric baselines fail.

2 Literature Review

2.1 3D Point Cloud Completion and Generation

Early 3D point cloud completion methods largely followed encoder-decoder designs. Representative works such as PCN [43] and GRNet [40] employed coarse-to-fine generation to progressively recover missing geometry. Transformer-based models later advanced this field by encoding partial inputs into point proxy sequences and leveraging geometry-aware attention to capture long-range dependencies for complete shape decoding [42]. Subsequent methods further refined this paradigm through hierarchical point-wise deconvolution [39], view-structure disentanglement [48], or specialized guidance mechanisms like seed expansion [47] and proxy-based alignment [23]. On the other hand, generative approaches (e.g., ContextualCompletion [10], PCDreamer [36]) formulated completion as conditional sampling from prior distributions. Recent works also extended completion to cross-modal settings [13, 14, 27, 29, 35], which took auxiliary inputs to provide explicit guidance and alleviate the ill-posed nature of shape recovery. Notably, SuperPC [13] unified completion from partial observations with other restoration tasks, including upsampling, denoising, and colorization.

In clinical ultrasound scenarios, however, acquiring perfectly aligned data is highly impractical due to tissue deformation, probe pressure, and respiratory motion during scanning. Driven by similar practical constraints, some works have explored unpaired point cloud completion, e.g., USSPA [28] and UOT-UPC [21]. However, in ultrasound-to-anatomical comprehensive bone completion, the key

challenge lies in the deterministic, physics-driven ultrasound artifacts that systematically distort the observations.

2.2 Physics-Guided Generation and Neural Surrogates

Recent physics-guided frameworks have used learnable models with real-world physics simulations to enhance the generation quality. For example, PIDM [4] imposed governing laws as the training-time regularizer to ensure that predictions remain physically self-consistent. In contrast, PhysDiff [44] incorporated a human motion simulator at inference time as guidance, progressively refining generated samples to enforce physical consistency. Such physics-guided strategies have been extended to different domains, including geometrical design [17, 18, 46], and animation [41]. In our practice, the physics prior, USimNet, serves as a lightweight learned surrogate that captures physics effects while maintaining the gradient flow necessary for test-time rectification.

2.3 Medical Shape Reconstruction for Bone

Traditionally, Statistical Shape Models (SSMs) [3, 11, 32, 34] were used to perform 3D reconstruction by learning a parametric shape representation. More recently, the 3D medical shape reconstruction has shifted toward using Implicit Neural Representations (INRs), e.g., Signed Distance Functions (SDFs) and occupancy fields for the surface reconstruction [2, 8, 9, 38]. Although these works excelled at capturing complex topologies, they typically rely on clean and relatively complete observations. In the context of ultrasound bone imaging, volumetric representations were often reconstructed via Voxel Nearest Neighbor (VNN) [7] interpolation according to spatial poses in clinical practice. Alternatively, point-cloud-based approaches have been proposed to register ultrasound with CT for intraoperative navigation in spine surgery [22]. While these methods do not provide global anatomy, Gafencu *et al.* [15, 16] explored leveraging ultrasound-physics-aware ray-casting to generate synthetic partial vertebrae point clouds for vertebral shape completion. Although accounting for basic ultrasound artifacts, their method was based on a variational autoencoder (VAE) [19], which inherently maps uncured, complex real-world artifacts directly into the latent space. This may lead to shape collapse on in-vivo data, whose distribution differs from that of the simulated data. In contrast, our work formulates US 3D completion as a conditional generative task. By leveraging a large-scale CT dataset (Spine1K [12]) to form the anatomy prior, UBone3D recovers globally consistent geometry while accounting for physics-induced artifacts.

3 Methodology

3.1 Preliminary: Optimal Transport Flow Matching

In this work, we build our generative prior **BoneFM** based on Optimal Transport Flow Matching (OT-FM) [24, 25], which provides a gradient path to transport a simple base distribution to a complex data distribution.

Let $x \in \mathbb{R}^{3N}$ denote the vectorized representation of a point cloud $P = \{p_i\}_{i=1}^N \subset \mathbb{R}^3$, and $t \in [0, 1]$ be the continuous time variable. We define the distributions as:

$$x_0 \sim p_0(x) = \mathcal{N}(0, \mathbf{I}), \quad x_1 \sim p_1(x), \quad \text{where } p_1(x) \approx p_{\text{AC}}(x), \quad (1)$$

where $p_0(x)$ is a standard Gaussian noise distribution and $p_{\text{AC}}(x)$ represents the target distribution of clean and anatomically consistent bone geometries.

To obtain a highly efficient transport path, OT-FM constructs a path under the OT coupling that connects x_0 and x_1 via linear interpolation [1]:

$$x(t) = tx_1 + (1 - t)x_0. \quad (2)$$

Taking the time derivative of Eq. 2 yields the optimal transport velocity:

$$u(x(t), t) = x_1 - x_0 \quad (3)$$

Such straight-trajectory formulations are highly desirable as they accelerate ordinary differential equation (ODE) integration during inference [20]. A neural velocity field $v_\theta(x, t)$ can be trained to approximate $u(t)$ using a simple regression objective:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t \sim \mathcal{U}(0,1), x_0 \sim p_0, x_1 \sim p_1} \left[\|v_\theta(x(t), t) - (x_1 - x_0)\|_2^2 \right]. \quad (4)$$

In our UBone3D, this formulation is later extended to a conditional velocity field, where conditioning variables, i.e., the ultrasound observation, reshape the transport trajectory toward the observation-consistent posterior manifold. Throughout this work, we use x to denote the vectorized representation of the target anatomy point cloud P_{AC} , and y to denote the ultrasound observation point cloud $P_{\text{in-vivo}}$ or $P_{\text{phys-full}}$.

3.2 Framework Overview

UBone3D aims to infer an **anatomically consistent completion** x^* that recovers the full vertebral geometry from the artifact-laden and partial **in-vivo** ultrasound observation y . We formulate this observation-guided shape completion as a probabilistic decomposition:

$$x^* \sim p(x | y) \propto p(y | x)p(x). \quad (5)$$

The final completed point cloud is obtained by interpreting the state x^* as a 3D point cloud $P_{\text{AC}}^* = \{p_i^*\}_{i=1}^N$. As illustrated in Fig. 2, UBone3D integrates two modules during inference: a generative backbone and a physics rectifier:

- **Generative Anatomy Prior (BoneFM):** BoneFM parameterizes a conditional velocity field $v_\theta(x(t), t, y)$ optimized via the OT-FM objective (Eq. 4). Given an observation y , it transports samples from a standard Gaussian base distribution toward the clean anatomical manifold p_{AC} .

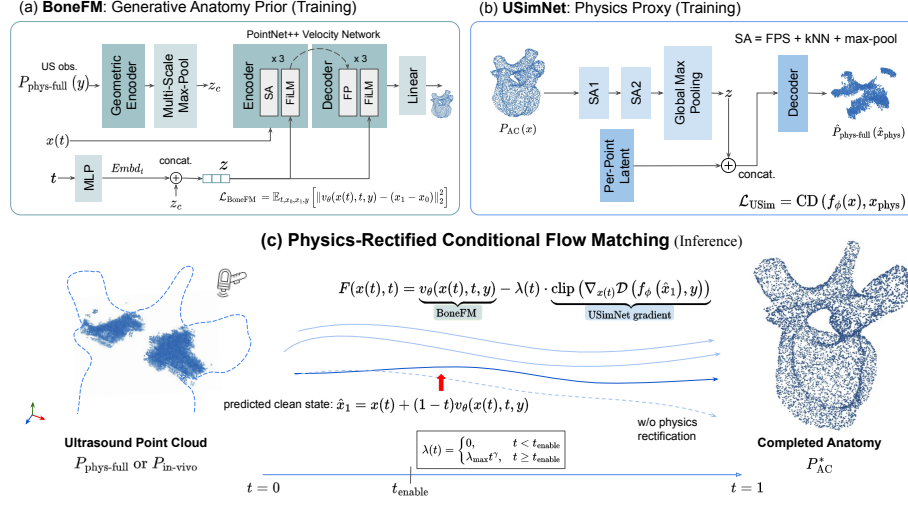


Fig. 2: Overview of the UBone3D framework. (a) **BoneFM Training:** The generative anatomy prior is learned by training a velocity field v_θ to transport a base distribution toward the manifold of clean CT anatomy P_{AC} . (b) **USimNet Training:** Our differentiable physics proxy f_ϕ learns to project a complete anatomical point cloud P_{AC} into its ultrasound-style counterpart $P_{\text{phys-full}}$. (c) **Rectified Inference:** Given the artifact-laden, partial observation, we obtain an anatomically consistent reconstruction P_{AC}^* with the rectified inference. v_θ here denotes the CFG velocity.

- **Physics Proxy (USimNet):** We model $p(y|x)$ with USimNet f_ϕ , a differentiable surrogate network, to *replicate the geometric distribution of in-vivo observations by learning from physics-based simulation*. Given a candidate complete shape x , it acts as a data-driven forward model that maps the clean anatomy into the ultrasound domain: $\hat{x}_{\text{phys}} = f_\phi(x)$.

At test time, BoneFM and USimNet jointly drive a **physics-rectified** conditional flow matching (Sec. 3.5), where the generative anatomy prior provides a shape basis, and USimNet supplies explicit observation-consistency gradients to steer the flow toward the posterior manifold.

3.3 Generative Anatomy Shape Prior: Conditional BoneFM

We train our generative anatomy prior, **BoneFM**, on an anatomically clean dataset, and model it as a *conditional* velocity field, as established in Sec. 3.1.

Conditional Velocity Parameterization. BoneFM employs a PointNet++-based [31] architecture enhanced with Feature-wise Linear Modulation (FiLM) [30]. The partial ultrasound observation y is first processed by a geometric encoder to extract a global context vector. Meanwhile, the continuous integration time t is

mapped to a high-dimensional sinusoidal embedding. These conditional signals are then fused and injected into the intermediate layers of the velocity network via FiLM, which applies affine transformations (scale and shift) to the intermediate point features. This allows the network to dynamically adjust its spatial receptive field based on the ultrasound observation.

Training Objective. Following the optimal transport path defined in Eq. 2, BoneFM minimizes a conditional adaptation of the regression loss established in Eq. 4. The network is optimized to predict the constant target velocity $u(t) = x_1 - x_0$ conditioned on the partial input y :

$$\mathcal{L}_{\text{BoneFM}} = \mathbb{E}_{t, x_0, x_1, y} \left[\|v_\theta(x(t), t, y) - (x_1 - x_0)\|_2^2 \right], \quad (6)$$

where the target state x_1 corresponds to $P_{\text{AC}} = \{p_i\}_{i=1}^N \subset \mathbb{R}^3$, the anatomically consistent ground-truth point cloud, and $y \in \mathbb{R}^{3M}$ denotes the simulated partial ultrasound observation point cloud. By mapping the standard Gaussian base distribution to clean anatomical shapes conditioned on ultrasound observation y , BoneFM establishes a robust and smooth topological foundation for the subsequent test-time shape guidance. To enable classifier-free guidance (CFG) at inference, the conditioning signal y is randomly replaced by a null token during training with probability p_{drop} .

3.4 Differentiable Physics Proxy: USimNet

Physics Simulation. To obtain physically grounded supervision for ultrasound appearance, we employ a customized ray-based simulator based on the open-source *PyMUST* [5], operating directly on anatomically clean meshes. For each mesh, we construct an orthonormal probe basis (u, v, w) and cast a dense set of rays from a start plane along the probe direction $-w$. Specifically, we first enforce a sector-shaped **field-of-view (FOV)** by clipping rays whose angular deviation from the probe apex exceeds a predefined aperture, e.g., 73° . The simulated probe is tilted around the world X -axis, e.g., by $+10^\circ$, to simulate realistic probe handling and incidence variations. We also model key ultrasound artifacts in terms of the following features through analytic filters and augmentations:

- **Angle Gating and Shadowing (Shadow):** Suppressing signal returns behind strong cortical reflections within a set lateral radius and depth margin;
- **Depth Windowing and Attenuation (Atten):** Constraining the effective imaging depth and decaying intensities;
- **Thick-Surface Augmentation (Thick):** Adding samples slightly inside the bone surface along the inward normal to mimic thick reflective surfaces;
- **Axial Streak Artifacts (Streak):** Duplicating hits along the ray direction to emulate streak-like artifacts;
- **Anti-Grid Sampling (Anti-Grid):** Adding stochastic variability (Poisson sampling, jitter, and noise) to reduce grid regularity.

After applying these physically motivated filters, we obtain a simulated ultrasound-style point cloud $P_{\text{phys-full}}$ (as detailed in Sec. 4.1). This simulated point cloud is used as supervision for training our differentiable proxy, **USimNet**.

USimNet Architecture and Training. Although the ray-based simulator described in Sec. 3.4 produces physically grounded supervision, it is non-differentiable and computationally prohibitive for test-time gradient computation. Therefore, we introduce **USimNet**, a lightweight, differentiable surrogate network f_ϕ , to learn a data-driven forward mapping that approximates complex physical artifacts (e.g., acoustic shadows, limited field-of-view) with the simulator’s output. It employs a PointNet++-based [31] encoder-decoder architecture. The encoder captures hierarchical geometric patterns with Set Abstraction (SA) layers. These features are then symmetrically pooled into a global context vector and processed by an MLP-based decoder.

Given the complete anatomically consistent point cloud $P_{\text{AC}} = \{p_i\}_{i=1}^N \subset \mathbb{R}^3$, represented as $x \in \mathbb{R}^{3N}$, USimNet learns a non-linear spatial transformation to project the input geometry into the simulated ultrasound point cloud $\hat{P}_{\text{phys-full}} \subset \mathbb{R}^3$:

$$\hat{x}_{\text{phys}} = f_\phi(x), \quad (7)$$

where $\hat{x}_{\text{phys}} \in \mathbb{R}^{3N'}$ denotes the simulated ultrasound point cloud $\hat{P}_{\text{phys-full}}$ for geometric evaluation.

Training Objective. USimNet is trained by minimizing the symmetric Chamfer Distance (CD), which can be formulated as:

$$\mathcal{L}_{\text{USim}} = \text{CD}(f_\phi(x), x_{\text{phys}}) \quad (8)$$

where x and x_{phys} denote vectorized representations of P_{AC} and $P_{\text{phys-full, train}}$, respectively.

In summary, USimNet provides a differentiable mapping $P_{\text{AC}} \rightarrow \hat{P}_{\text{phys-full}}$. By acting as a structural *physics proxy*, it can supply explicit observation-consistency gradients during the test-time flow inference.

3.5 Inference: Physics-Rectified Conditional Flow Matching

Given the artifact-laden and partial ultrasound observation y , we reconstruct the complete anatomy P_{AC}^* via physics-rectified conditional flow matching.

Starting from pure noise $x(0) \sim \mathcal{N}(0, \mathbf{I})$, we numerically solve the flow from $t = 0$ to $t = 1$ with Heun’s second-order method for a stable integration of the generative prior and the physics correction. At each step Δt , the state is updated as:

$$\begin{aligned} k_1 &= F(x(t), t), \\ k_2 &= F(x(t) + \Delta t k_1, t + \Delta t), \\ x(t + \Delta t) &= x(t) + \frac{\Delta t}{2}(k_1 + k_2), \end{aligned} \quad (9)$$

where

$$F(x(t), t) = v_{\text{cfg}}(x(t), t, y) - \lambda(t) \text{clip}(\nabla_{x(t)} \mathcal{D}(f_\phi(\hat{x}_1(x(t), t)), y)), \quad (10)$$

where $v_{\text{cfg}}(x, t, y) = (1 + w) v_\theta(x, t, y) - w v_\theta(x, t, \emptyset)$ is the classifier-free guidance velocity with weight w . $\text{clip}(\cdot)$ denotes a per-point gradient normalization that prevents numerical instability caused by outlier gradients. All gradients are computed via automatic differentiation. At intermediate state $x(t)$, we predict the clean target shape before evaluating the physics proxy f_ϕ since $x(t)$ does not lie on the clean data manifold:

$$\hat{x}_1(x(t), t) = x(t) + (1 - t) v_{\text{cfg}}(x(t), t, y). \quad (11)$$

To form an explicit rectification gradient on BoneFM, we apply a geometric discrepancy $\mathcal{D}$ with the directed Chamfer Distance:

$$\mathcal{D}(f_\phi(\hat{x}_1(x(t), t)), y) = \text{CD}_{\text{directed}}(y, f_\phi(\hat{x}_1(x(t), t))). \quad (12)$$

The directed CD is used to avoid penalizing regions that are not covered by y .

Observation Anchoring. During integration, we hold a subset of N_{vis} observation points on their exact OT-FM path $x_{\text{vis}}(t) = t y_{\text{vis}} + (1 - t) z_{\text{vis}}$ (so $x_{\text{vis}}(1) = y_{\text{vis}}$), letting the rectification complete only the unobserved region.

Time-dependent Guidance Schedule. Applying physical gradients in early, high-noise stages might corrupt the global topology. Therefore, we restrict USimNet’s intervention to the late generative phase using a truncated polynomial schedule:

$$\lambda(t) = \begin{cases} 0, & \text{if } t < t_{\text{enable}}, \\ \lambda_{\text{max}} t^\gamma, & \text{if } t \geq t_{\text{enable}}, \end{cases} \quad (13)$$

where t_{enable} represents the activation threshold and γ represents the scaling factor. Detailed parameters are provided in Sec. 4.2.

4 Experimental Setup

4.1 Dataset

Simulation Data. We curated **Spine1K-PC**, a large-scale simulated point cloud dataset derived from Spine1K [12]. To establish consistent geometric ground truths (GTs), we extracted individual vertebrae from the raw volumetric scans and applied the Marching Cubes algorithm to generate the ground-truth meshes ($\mathcal{M}_{\text{GT}}$). Subsequently, we uniformly sampled $N = 8,192$ points from $\mathcal{M}_{\text{GT}}$ to obtain the **Anatomically Consistent Point Cloud** (P_{AC}). This serves as the complete anatomical reference that our model aims to recover. This dataset was split into 5,502 samples for training and 1,373 samples for testing. We generated four distinct variants of partial ultrasound observations, all based on the same field-of-view cut (FOV-Cut) and vertebral body removal (Body-Cut):

- 1) **Pure Geometric Train** ($P_{\text{geo, train}}$): We simulated a restricted observation by taking the top-35% points (Top 35%) at a predefined tilted angle of the probe on P_{AC} . This serves as the most basic, artifact-free partial input.
- 2) **Simple Physics Train** ($P_{\text{phys-simp, train}}$): We applied PyMUST [5] on $\mathcal{M}_{\text{GT}}$ to simulate minimal physics effects, i.e., first-return ray casting (Ray) and spinous-process removal (SP-Cut) under the same probe angle as P_{geo} .
- 3) **Full Physics Train** ($P_{\text{phys-full, train}}$): We used PyMUST to simulate comprehensive acoustic physics, including acoustic shadowing (Shadow), depth windowing and exponential attenuation (Atten), thick-surface responses (Thick), axial streak artifacts (Streak), and anti-grid sampling perturbations (Anti-Grid: Poisson + jitter + subsampling + noise), under the same probe angle as P_{geo} on the training set.
- 4) **Full Physics Test** ($P_{\text{phys-full, test}}$): To evaluate generalization capability, we simulated full physics effects on 1,373 test samples using a *novel, strictly unseen* probe angle.

In-Vivo Data. With ethics approval and informed consent, we retrospectively collected 5 cases from healthy adult volunteers using a Clarius transducer (Clarius Mobile Health, BC, Canada) following a standard protocol. We then reconstructed a 3D vertebral point cloud of the full spine for each case after segmentation on the B-mode images. Finally, we extracted 24 vertebrae in total from the thoracic (20) and lumbar (4) regions exclusively for the sim-to-real test. The resulting point clouds $P_{\text{in-vivo, test}}$ were normalized into a canonical space.

4.2 Implementation Details

Network Architectures and Training. For the anatomy generative prior, BoneFM, we parameterized the conditional velocity field with PointNet++ [31] backbone modulated by FiLM layers [30]. BoneFM was trained on the 5,502 P_{AC} training samples for 700 epochs, with ≈ 3.6 minutes per epoch, using the AdamW optimizer and an initial learning rate of 10^{-4} . Classifier-free guidance (CFG) was used by randomly dropping the conditioning signal during training with probability $p_{\text{drop}} = 0.2$, with guidance weight $w = 2.0$ at evaluation. For the physics proxy, USimNet, we employed a PointNet++-based architecture. It was trained on the paired simulation dataset using the AdamW optimizer with a learning rate of 5×10^{-5} to minimize the CD loss.

Guided Inference. During the test-time inference, we discretized the continuous time $t \in [0, 1]$ into $K = 35$ integration steps. To enhance the conditional generation quality of BoneFM, we applied classifier-free guidance (CFG) with a scale of $w = 2.0$. For the time-dependent physics guidance schedule $\lambda(t)$, we set the activation threshold to $t_{\text{enable}} = 0.5$, the polynomial scaling factor to $\gamma = 3$, and the maximum guidance weight to $\lambda_{\text{max}} = 0.6$. For observation anchoring, we FPS-subsampled the input to a fixed subset of N_{vis} anchor points and re-projected them onto their OT-FM path at every integration step, while the remaining points were generated freely. To ensure numerical stability during shape deformation, the analytical gradients from the physics proxy were strictly

clipped at a threshold of 1.0. All experiments were conducted on 40GB NVIDIA A100 GPUs. More details are provided in the Supplementary Material.

4.3 Evaluation Metrics and Baselines

Simulation Quality Metrics. To assess simulation fidelity, we evaluated the set-level distributional similarity between the aligned in-vivo test set ($P_{\text{in-vivo, test}}$) and the simulated outputs produced by our method and by Gafencu *et al.*'s method [16], respectively, under an unpaired bootstrap protocol. Since no one-to-one ground-truth correspondence exists, we repeatedly drew random subsets of K simulated point clouds (matching the in-vivo test set size) over 100 trials. All clouds were resampled to 8,192 points for fair comparison. We report the Minimum Matching Distance based on Chamfer Distance (MMD-CD $\downarrow$) and Coverage (COV $\uparrow$) to measure geometric proximity and mode coverage. Additionally, we computed the Density Jensen-Shannon Divergence (JSD $\downarrow$) on voxelized density distributions and the Sliced Wasserstein Distance (SWD $\downarrow$) to rigorously capture occupancy and global distribution mismatches.

Completion Baselines. We compared our method against several point cloud completion baselines, including PoinTr [42] (Transformer-based), SVDFormer [48] (Transformer-based), PCDreamer[†] [36] (Adapted diffusion-based approach used as a multi-modality baseline), Gafencu *et al.*'s method [16] (VAE-based), and one clinical statistical model, SSM-Net* [11]. We implemented SSM-Net*, a parametric baseline inspired by SMPL [26], which models vertebral shapes via PCA-derived blend shapes and linear blend skinning trained on the registered ground-truth meshes ($\mathcal{M}_{\text{GT, train}}$). For a fair evaluation, the resulting mesh outputs were uniformly converted into point clouds.

Completion Metrics (Simulation Data). We evaluated all methods by taking $P_{\text{phys-full}}$ as input and comparing the completed output against the anatomical reference P_{AC} . Following recent works [36, 42], we used $L1$ Chamfer Distance ($L1 \text{ CD} \times 10^3 \downarrow$), Earth Mover's Distance ($\text{EMD} \times 10^3 \downarrow$), and F-score@1% $\uparrow$ to quantitatively evaluate the completion results. All metrics were computed after normalizing both the predicted and reference point clouds to a unit sphere.

Clinical Metrics (In-Vivo Data). We report the laminae distance error (mm) measured by an experienced operator to evaluate the anatomical fidelity of UBone3D-completed vertebrae against both (1) the raw partial observations and (2) the VNN-reconstructed reference volumes. The completion results were compared to SVDFormer[†] (physically-trained) and SSM-Net*.

5 Results and Discussions

5.1 Efficacy of the Physics Proxy

Fig. 3 provides a side-by-side visual comparison between in-vivo ultrasound point clouds $P_{\text{in-vivo}}$, a geometric baseline P_{geo} and simulated point clouds, i.e., $P_{\text{phys-simp}}$ and $P_{\text{phys-full}}$. We also compared to point clouds P_{dark} produced following Gafencu *et al.*'s [16] method. Three distinct patterns are observed: (1)

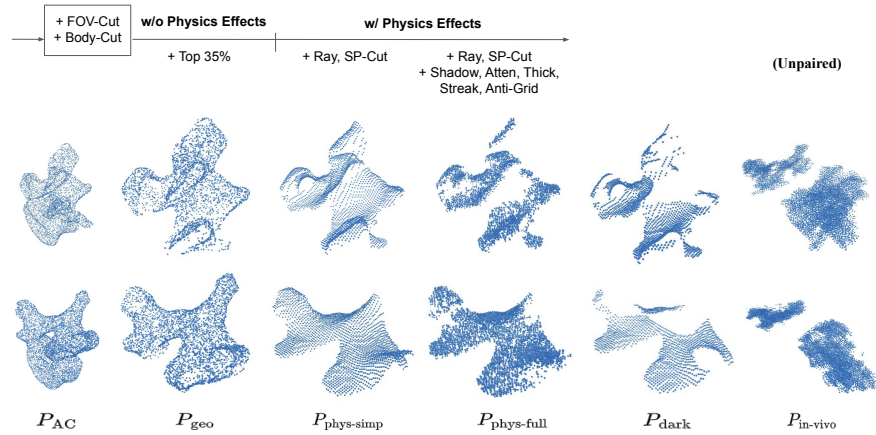


Fig. 3: Comparison of simulation variants derived from the same $\mathcal{M}_{GT}$, all after applying a FOV-Cut and a Body-Cut. P_{geo} is a purely geometric baseline obtained by shallow cropping with top-35% points (Top 35%), $P_{phys-simp}$ simulates minimal physics effects, i.e., first-return ray casting (Ray) and spinous-process removal (SP-Cut). $P_{phys-full}$ further adds acoustic shadowing (Shadow), depth windowing and exponential attenuation (Atten), thick-surface responses (Thick), axial streak artifacts (Streak), and anti-grid sampling perturbations (Anti-Grid: Poisson + jitter + subsampling + noise). The in-vivo examples (right) are shown for reference only and are not paired with the simulated point clouds.

Table 1: Set-level distributional similarity with respect to in-vivo data $P_{in-vivo}$.

Data	MMD-CD↓	COV↑	JSD↓	SWD↓
P_{AC}	1298.480	0.124	0.664	16.246
P_{geo}	961.705	0.140	0.649	14.083
P_{dark} [16]	970.563	0.179	0.655	14.293
$P_{phys-simp}$	964.187	0.176	0.656	14.309
$P_{phys-full}$	877.394	0.205	0.638	12.917

P_{geo} and P_{dark} mainly preserve a clean bone surface, (2) physics-driven structural missingness emerges in the simulated point clouds due to ultrasound view dependence and acoustic dropouts, and (3) $P_{phys-full}$ additionally exhibits realistic, structured artifacts. In particular, the full simulation captures key ultrasound characteristics which are commonly observed in real scans. The simulated point clouds show strong qualitative agreement with in-vivo data. Table 1 shows the set-level distributional similarity between simulated and in-vivo point clouds $P_{in-vivo}$. The sampled CT point clouds P_{AC} , as the baseline, exhibit the largest distributional distance. Geometrically-cut P_{geo} and simple physics $P_{phys-simp}$ show a closer match to in-vivo data. P_{dark} does not exhibit a clear advantage over our settings and remains relatively far from the in-vivo distribution. Overall, our full physics simulation $P_{phys-full}$ shows the tightest statistical

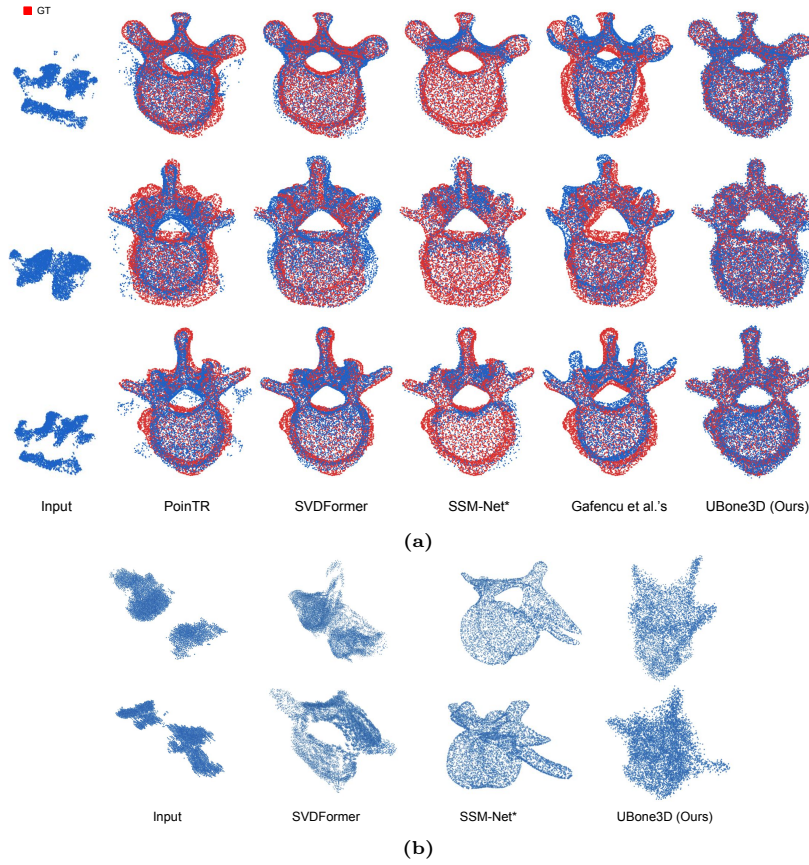


Fig. 4: Qualitative comparisons on simulation data (a) and in-vivo data (b).

alignment with the in-vivo distribution. Taking both results into account, we conclude that $P_{\text{phys-full}}$ is a trustworthy proxy for real-world physics.

5.2 Completion Results of Simulation Data

Fig. 4 (a) presents qualitative visualization comparisons of UBone3D against baselines. Visually, PoinTr [42], SVDFormer [48], and SSM-Net* produced inferior results, exhibiting additional noise, lacking anatomical individuality, or generating the transverse process independently of the input. Although recovering generally good vertebral arch regions, Gafencu et al.'s method [16] produced a small number of points on the vertebral bodies. The adapted baseline, PC-Dreamer [36], unfortunately, failed to generate reasonable outputs. In contrast, our method preserved both global anatomical consistency and local structural sharpness. Table 2 reports the quantitative comparisons. Although Gafencu *et al.*'s method and SVDFormer achieved the lowest CD under the two training set-

Table 2: Quantitative comparisons on the simulation data with geometry-trained and simulation-trained variants ($L1\ CD \times 10^3 \downarrow$, $EMD \times 10^3 \downarrow$, and $F\text{-score}@1\% \uparrow$).

Method	CD↓ (trained with P_{geo})	EMD↓ (trained with P_{geo})	F1↑ (trained with P_{geo})	CD↓ (trained with $P_{\text{phys-full}}$)	EMD↓ (trained with $P_{\text{phys-full}}$)	F1↑ (trained with $P_{\text{phys-full}}$)
PoinTr [42]	70.960	107.701	0.043	32.936	78.404	0.055
SVDFormer [48]	42.106	207.075	0.033	25.463	127.965	0.046
PCDremer† [36]	74.281	194.532	0.032	59.035	92.456	0.025
Gafencu <i>et al.</i> [16]	41.648	250.939	0.056	29.091	256.953	0.063
SSM-Net*	79.907	182.267	0.040	-	-	-
UBone3D	56.294	105.292	0.065	28.161	74.298	0.079

Table 3: Laminae Distance Errors (LDE) compared to raw input¹ and VNN reference².

	SVDFormer†	SSM-Net*	UBone3D
LDE (mm)¹ ↓	0.838	1.253	0.819
LDE (mm)² ↓	3.937	3.382	1.293

Table 4: Ablation on the variants of physics proxy USimNet.

Physics Proxy	CD↓	EMD↓	F1↑
w/o USimNet	49.196	96.673	0.064
Simple	44.287	87.271	0.065
Full (Ours)	28.161	74.298	0.079

Table 5: Ablation on the generative anatomy prior.

Backbone	CD↓ (trained with P_{geo})	EMD↓ (trained with P_{geo})	F1↑ (trained with P_{geo})	CD↓ (trained with $P_{\text{phys-full}}$)	EMD↓ (trained with $P_{\text{phys-full}}$)	F1↑ (trained with $P_{\text{phys-full}}$)
BoneDiff†	59.274	105.920	0.068	42.597	75.987	0.071
BoneFM (Ours)	56.294	105.292	0.065	28.161	74.298	0.079

tings, respectively, they exhibited substantially higher EMD values. Our framework shows consistently stable performance, suggesting that our method is better suited to handle ultrasound-specific artifacts than baselines.

5.3 Completion Results of Zero-Shot In-Vivo Data

The in-vivo qualitative and quantitative comparisons are provided in Fig. 4 (b) and Table 3, respectively. Other baselines often failed to recover anatomical structure and produced distorted output. This might be due to their original design, which makes them struggle to generalize to versatile, artifact-dominated ultrasound inputs. Our method consistently produced anatomically plausible vertebrae with relatively correct bone structure observed from the visualization, and more accurate laminae distance as seen in Table 3. These results suggest that our model has potential for real-world patient data. Nevertheless, compared to the clear visual results on simulated data, the results on real-world (in-vivo) data may require further improvement.

5.4 Ablation Studies

Effectiveness of Physics Rectification. An ablation study on the physics rectification variants is presented in Table 4. The BoneFM alone (w/o USimNet)

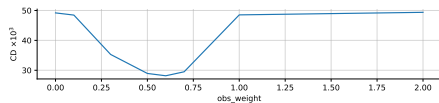


Fig. 5: Effect of observation weight.

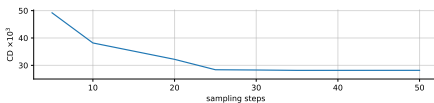


Fig. 6: Ablation on the sampling steps.

yielded competitive completion results. Nevertheless, the introduction of physics rectification further improved completion fidelity. Simplified physics rectification (trained on $P_{\text{phys-simp, train}}$) stabilized the flow trajectory, while the full physics proxy (trained on $P_{\text{phys-full, train}}$) enforced the best physical plausibility.

Effectiveness of the Generative Anatomy Prior (*vs.* DDIM). We ablated the choice of generative model for the anatomy prior by comparing BoneFM to DDIM [33] (BoneDiff_‡), each run at its own optimal number of sampling steps, with BoneDiff_‡ requiring $2\times$ the steps of BoneFM to reach comparable convergence. As seen in Table 5, BoneFM achieved better overall performance while requiring significantly fewer steps compared to BoneDiff_‡.

Ablation on Rectification Strength. Fig. 5 illustrates the effect of varying the physics guidance weight λ on reconstruction quality. The CD exhibited a U-shaped response, degrading under either insufficient or excessive guidance. The rectification strength of $\lambda = 0.6$ achieved the best balance between anatomy prior and physics rectification. t_{enable} is fixed at 0.5.

Ablation on Sampling Steps. Fig. 6 shows the reconstruction quality with respect to the number of ODE integration steps, where λ is fixed at 0.6. The performance increases markedly as the step count rises, with the most significant gains observed between 25 and 35 steps. To achieve a good trade-off between completion fidelity and inference efficiency, we set the number of steps to 35.

6 Conclusion

By leveraging an anatomical generative prior (BoneFM) and a differentiable physics proxy (USimNet), UBone3D completes anatomically consistent vertebral reconstructions from artifact-laden and partial ultrasound observations at inference time. Evaluations across simulation and in-vivo datasets confirm the effectiveness of our proposed framework. While further improvements are still needed, UBone3D provides a new perspective on clinical bone anatomy completion from ultrasound.

Acknowledgements

The authors would like to thank all volunteers who participated in this study. This work was supported in part by the Women and Children’s Health Research Institute (WCHRI), Alberta Innovates, and the Natural Sciences and Engineering Research Council of Canada (NSERC). Ethics approval for this study was granted under protocol Pro00005707.

References

1. Albergo, M.S., Vanden-Eijnden, E.: Building normalizing flows with stochastic interpolants. In: The Eleventh International Conference on Learning Representations (2023)
2. Amiranashvili, T., Lüdke, D., Li, H.B., Zachow, S., Menze, B.H.: Learning continuous shape priors from sparse data with neural implicit functions. *Medical Image Analysis* **94**, 103099 (2024)
3. Aubert, B., Vázquez, C., Cresson, T., Parent, S., de Guise, J.A.: Toward automated 3d spine reconstruction from biplanar radiographs using cnn for statistical spine model fitting. *IEEE Transactions on Medical Imaging* (2019)
4. Bastek, J.H., Sun, W., Kochmann, D.M.: Physics-informed diffusion models. In: The Thirteenth International Conference on Learning Representations (2025)
5. Bernardino, G., Garcia, D.: Pymust: an open-source python library for the simulation and analysis of ultrasound. In: 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS). pp. 1–4. IEEE (2024)
6. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: Shapenet: An information-rich 3d model repository. Zenodo (CERN European Organization for Nuclear Research) (2023)
7. Chen, H.B., Zheng, R., Qian, L.Y., Liu, F.Y., Song, S., Zeng, H.Y.: Improvement of 3-d ultrasound spine imaging technique using fast reconstruction algorithm. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **68**(10), 3104–3113 (2021)
8. Chen, H., Gao, Y., Zhang, S., Wu, J., Ma, Y., Zheng, R.: RoCoSDF: Row-Column Scanned Neural Signed Distance Fields for Freehand 3D Ultrasound Imaging Shape Reconstruction. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. pp. 721–731. Springer Nature Switzerland, Cham (2024)
9. Chen, H., Kumaralingam, L., Zhang, S., Song, S., Zhang, F., Zhang, H., Pham, T.T., Punithakumar, K., Lou, E.H., Zhang, Y., Le, L.H., Zheng, R.: Neural implicit surface reconstruction of freehand 3D ultrasound volume with geometric constraints. *Medical Image Analysis* **98**, 103305 (Dec 2024)
10. Chu, J., Li, W., Wang, X., Ning, K., Lu, Y., Fan, X.: Digging into intrinsic contextual information for high-fidelity 3d point cloud completion. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 2573–2581 (2025)
11. Cootes, T.: Statistical shape models. In: *Medical Image Analysis*, pp. 201–225. The MICCAI Society book Series, Academic Press (2024)
12. Deng, Y., Wang, C., Hui, Y., Li, Q., Li, J., Luo, S., Sun, M., Quan, Q., Yang, S., Hao, Y., Liu, P., Xiao, H., Zhao, C., Wu, X., Zhou, S.K.: Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography. *arXiv (Cornell University)* (2022)
13. Du, Y., Zhao, Z., Su, S., Golluri, S., Zheng, H., Yao, R., Wang, C.: Superpc: a single diffusion model for point cloud completion, upsampling, denoising, and colorization. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 16953–16964 (2025)
14. Du, Z., Dou, J., Liu, Z., Wei, J., Wang, G., Xie, N., Yang, Y.: Cdpnet: Cross-modal dual phases network for point cloud completion. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 38, pp. 1635–1643 (2024)

15. Gafencu, M.A., Velikova, Y., Navab, N., Azampour, M.F.: Us-x complete: A multi-modal approach to anatomical 3d shape recovery. In: International Workshop on Shape in Medical Imaging. pp. 218–231. Springer (2025)
16. Gafencu, M.A., Velikova, Y., Saleh, M., Ungi, T., Navab, N., Wendler, T., Azampour, M.F.: Shape completion in the dark: Completing vertebrae morphology from 3D ultrasound. *International Journal of Computer Assisted Radiology and Surgery* **19**(7), 1339–1347 (May 2024)
17. Giannone, G., Srivastava, A., Winther, O., Ahmed, F.: Aligning optimization trajectories with diffusion models for constrained design generation. *Advances in neural information processing systems* **36**, 51830–51861 (2023)
18. Jiang, H., Hsu, H.Y., Zhang, K., Yu, H.N., Wang, S., Li, Y.: Phystwin: Physics-informed reconstruction and simulation of deformable objects from videos. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 7219–7230 (October 2025)
19. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
20. Kornilov, N., Mokrov, P., Gasnikov, A., Korotin, A.: Optimal flow matching: Learning straight trajectories in just one step. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024)
21. Lee, T., Choi, J., Choi, J., Kang, M.: Unpaired point cloud completion via unbalanced optimal transport. In: Singh, A., Fazel, M., Hsu, D., Lacoste-Julien, S., Berkenkamp, F., Maharaj, T., Wagstaff, K., Zhu, J. (eds.) Proceedings of the 42nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 267, pp. 32965–32986. PMLR (13–19 Jul 2025)
22. Li, R., Cai, Y., Davoodi, A., Borghesan, G., Vander Poorten, E.: 3d ultrasound shape completion and anatomical feature detection for minimally invasive spine surgery. *Medical & Biological Engineering & Computing* **63**(7), 1987–2000 (Apr 2025). <https://doi.org/10.1007/s11517-025-03359-1>, <https://doi.org/10.1007/s11517-025-03359-1>
23. Li, S., Gao, P., Tan, X., Wei, M.: Proxyformer: Proxy alignment assisted point cloud completion with missing part sensitive transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9466–9475 (2023)
24. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations (2023)
25. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations
26. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* **34**(6), 248:1–248:16 (Oct 2015)
27. Luo, W., Wu, D., Na, H., Zhu, Y., Hu, M., Quan, G.: Rethinking multimodal point cloud completion: A completion-by-correction perspective. *arXiv preprint arXiv:2511.12170* (2025)
28. Ma, C., Chen, Y., Guo, P., Guo, J., Wang, C., Guo, Y.: Symmetric shape-preserving autoencoder for unsupervised real scene point cloud completion. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13560–13569 (2023)

29. Mao, A., Tang, Y., Huang, J., He, Y.: Dmf-net: Image-guided point cloud completion with dual-channel modality fusion and shape-aware upsampling transformer. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 6063–6071 (2025)
30. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 32 (2018)
31. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems* **30** (2017)
32. Rusli, W.M., Kedgley, A.E.: Statistical shape modelling of the first carpometacarpal joint reveals high variation in morphology. *Biomechanics and Modeling in Mechanobiology* **19**(4), 1203–1210 (2020)
33. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020)
34. Wai, G., Rusli, W., Ghouse, S., Kieser, D.C., Kedgley, A., Newell, N.: Statistical shape modelling of the thoracic spine for the development of pedicle screw insertion guides. *Biomechanics and modeling in mechanobiology* **22**(1), 123–132 (2023)
35. Wang, C., He, S., Fang, X., Han, J., Liu, Z., Ning, X., Li, W., Tiwari, P.: Point clouds meets physics: Dynamic acoustic field fitting network for point cloud understanding. In: *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*. pp. 22182–22192 (June 2025)
36. Wei, G., Feng, Y., Ma, L., Wang, C., Zhou, Y., Li, C.: Pcdreamer: Point cloud completion through multi-view diffusion priors. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 27243–27253 (2025)
37. Wysocki, M., Azampour, M.F., Eilers, C., Busam, B., Salehi, M., Navab, N.: Ultranerf: Neural radiance fields for ultrasound imaging. In: *Medical Imaging with Deep Learning* (Apr 2023)
38. Wysocki, M., Duelmer, F., Bal, A., Navab, N., Azampour, M.F.: Ultron: Ultrasound occupancy networks. In: Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*. pp. 606–615. Springer Nature Switzerland, Cham (2026)
39. Xiang, P., Wen, X., Liu, Y.S., Cao, Y.P., Wan, P., Zheng, W., Han, Z.: Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 5499–5509 (2021)
40. Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., Sun, W.: Grnet: Gridding residual network for dense point cloud completion. In: *European conference on computer vision*. pp. 365–381. Springer (2020)
41. Xie, T., Zhao, Y., Jiang, Y., Jiang, C.: Physanimator: Physics-guided generative cartoon animation. In: *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*. pp. 10793–10804 (June 2025)
42. Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., Zhou, J.: Pointr: Diverse point cloud completion with geometry-aware transformers. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 12498–12507 (2021)
43. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: *2018 international conference on 3D vision (3DV)*. pp. 728–737. IEEE (2018)

44. Yuan, Y., Song, J., Iqbal, U., Vahdat, A., Kautz, J.: Physdiff: Physics-guided human motion diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
45. Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K.: Lion: Latent point diffusion models for 3d shape generation. *Advances in Neural Information Processing Systems* **35**, 10021–10039 (2022)
46. Zheng, H., Zheng, Y., Huang, H., Sun, C., Liao, E., Liu, L., Han, Y., Zhou, H., Liang, S.: Co2-net: A physics-informed spatio-temporal model for global surface co2 reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6220–6230 (October 2025)
47. Zhou, H., Cao, Y., Chu, W., Zhu, J., Lu, T., Tai, Y., Wang, C.: Seedformer: Patch seeds based point cloud completion with upsample transformer. In: European conference on computer vision. pp. 416–432. Springer (2022)
48. Zhu, Z., Chen, H., He, X., Wang, W., Qin, J., Wei, M.: Svdformer: Complementing point cloud via self-view augmentation and self-structure dual-generator. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14508–14518 (2023)