OSCAR: Occupancy-based Shape Completion via Acoustic Neural Implicit Representations

Magdalena Wysocki[1,2⋆], Kadir Burak Buldu[1*], Miruna-Alexandra Gafencu[1,2,3], Benjamin D. Killeen[1,2], Mohammad F. Azampour[1,2], and Nassir Navab[1,2]

[1] Chair for Computer Aided Medical Procedures (CAMP)
Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany
{magdalena.wysocki, burak.buldu, miruna.gafencu, bd.killeen, mf.azampour, nassir.navab}@tum.de
[2] Munich Center for Machine Learning (MCML), Munich, Germany
[3] Konrad Zuse School of Excellence in Reliable AI (relAI), Germany

Abstract. Accurate 3D reconstruction of vertebral anatomy from ultrasound is important for guiding minimally invasive spine interventions, but it remains challenging due to acoustic shadowing and view-dependent signal variations. We propose an occupancy-based shape completion method that reconstructs complete 3D anatomical geometry from partial ultrasound observations. Crucially for intra-operative applications, our approach extracts the anatomical surface directly from the image, avoiding the need for anatomical labels during inference. This label-free completion relies on a coupled latent space representing both the image appearance and the underlying anatomical shape. By leveraging a Neural Implicit Representation (NIR) that jointly models both spatial occupancy and acoustic interactions, the method uses acoustic parameters to become implicitly aware of the unseen regions without explicit shadowing labels through tracking acoustic signal transmission. We show that this method outperforms state-of-the-art shape completion for B-mode ultrasound by 80% in HD95 score. We validate our approach both *in silico* and on phantom US images with registered mesh models from CT labels, demonstrating accurate reconstruction of occluded anatomy and robust generalization across diverse imaging conditions. Code and data availabe at https://github.com/magdalena-wysocki/Oscar.

Keywords: Shape Completion · Implicit Representation · Ultrasound

1 Introduction

Ultrasound (US) image-guided spinal surgery offers significant benefits for both patients and clinicians, avoiding excessive ionizing radiation while retaining the benefits of minimally invasive approaches. Unfortunately, severe acoustic shadowing and reverberation significantly degrades image quality, so that only the near surface of hard tissue can be directly observed. To provide intra-operative

⋆ These authors contributed equally to this work.

guidance, then, expert sonographers rely on strong prior knowledge of spine anatomy to acquire multiple viewpoints and mentally infer the complete 3D geometry. Anatomical shape completion mimics this cognitive process, moving beyond visible surface reconstruction [2,16] to create the anatomical digital twin, *e.g.*, for robotic- and computer-assisted approaches transitioning from distinct (visible) to completed (occluded) [3] anatomical reconstruction. In the past, shape completion methods have used statistical shape models to approximate this expert capability [5,4,7], but the reliance on explicitly annotated, discrete point clouds inherently ignores the acoustic image formation process, treating shadows as missing data and ignoring their view-dependent properties. Our goal is to provide robust, physics-informed shape completion by leveraging both anatomical and acoustic priors to reconstruct the complete vertebral geometry from fragmented B-mode images, thereby improving 3D spatial guidance for minimally invasive spine interventions and robotic surgery.

Recently, Neural Implicit Representations (NIRs) have emerged as a powerful alternative to discrete surfaces for complex shape modeling. Foundational works in computer vision, such as DeepSDF [9] and Occupancy Networks [8], demonstrated that NIRs can effectively encode complex 3D shapes. Within medical imaging, approaches like ImplicitAtlas [18], alongside other recent implicit shape models [14,17] have further shown that these networks can learn rich, continuous shape priors, while shared latent spaces encode both anatomical shape and image appearance [12,13]. By training on both images and anatomical segmentations, these methods enable self-supervised test-time optimization (TTO), where the shared latent space is optimized using only a medical image to accurately infer the segmentation. Thus far, however, NIRs with coupled latent spaces have been ill-suited to the task of shape completion in ultrasound due to the severe occlusions of acoustic shadowing and view-dependent characteristics of B-mode images. Without a modality-specific model, incomplete and contradicting information from multiple views corrupts the reconstruction process.

Here, we present Occupancy-based Shape Completion via Acoustic NIRs (OSCAR), a physics-aware framework for US-based spinal shape completion that integrates view-dependent acoustic properties of tissue into a coupled-latent NIR. By modeling acoustic image formation and signal propagation as a ray-tracing process in this NIR [1,15,10], we enable the latent space to resolve multi-view pixel contradictions and reconstruct unobserved regions with information from disparate viewpoints. Further, this coupling facilitates bidirectional optimization. Given an inferred anatomical shape, our model can reconstruct a plausible acoustic space, allowing for novel view synthesis of B-mode images with known shape. We evaluate our approach both *in silico* and on real images of an anthropomorphic phantom, with ground truth from registered computed tomography (CT) imaging. Compared to the state-of-the-art in ultrasound shape completion SITD [5], OSCAR improves the HD95 score of reconstructed shapes by 80% while maintaining anatomically plausible shapes across its latent space, in addition to enabling realistic image synthesis from known shape.

2 Method

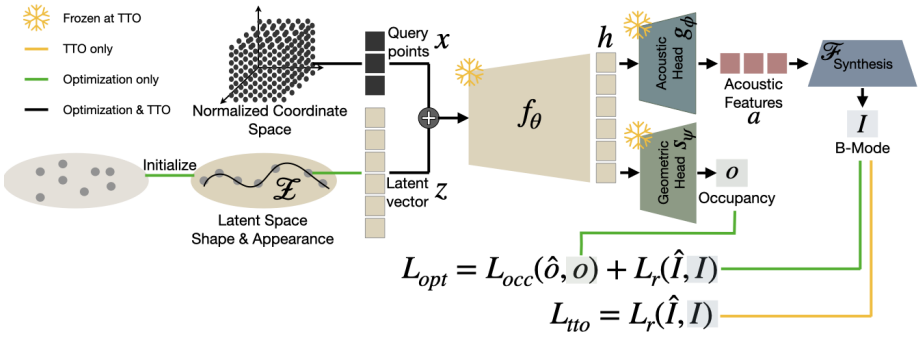

Fig. 1. Proposed framework pipeline. Joint Training: The network couples an acoustic head (g_ϕ) and a geometric head (s_ψ) via a shared latent space ($\mathcal{Z}$) and backbone (f_θ). Supervised by both B-mode synthesis ($\mathcal{L}_{photo}$) and ground-truth shape occupancy ($\mathcal{L}_{occ}$), the ray-based rendering inherently models acoustic shadowing.**Test-Time Optimization:** At inference, network weights are frozen. A novel latent code $\mathbf{z}^*$ is optimized directly from fragmented B-mode observations without shape labels ($\mathcal{L}_{tto}$). The optimized prior inherently extracts the complete 3D geometry $o(\mathbf{x} \mid \mathbf{z}^*)$.

2.1 Problem Formulation and Method Overview

Given a sequence of partial B-mode ultrasound observations $\mathcal{D} = \{(I_k, T_k)\}_{k=1}^{K}$, with images I_k and corresponding probe poses $T_k \in SE(3)$, our goal is to reconstruct the complete 3D vertebral geometry, defined as a continuous occupancy field $O : \mathbb{R}^3 \rightarrow [0, 1]$. We propose a NIR-based framework (Fig. 1) that jointly models this geometry and acoustic properties.

For a 3D coordinate $\mathbf{x}$ and a subject-specific latent code $\mathbf{z}$, a shared backbone network f_θ extracts a joint feature representation. This feature space is then processed by two distinct output heads. The acoustic head g_ϕ maps the joint features to a localized acoustic property vector $\mathbf{a}(\mathbf{x}) = g_\phi(f_\theta(\mathbf{x}, \mathbf{z}))$, where $\mathbf{a}(\mathbf{x}) = [\beta(\mathbf{x}), \sigma(\mathbf{x}), \mu(\mathbf{x})]^T$ represents acoustic reflection, scattering, and attenuation. Simultaneously, the occupancy head s_ψ maps the same features to the continuous volume occupancy probability $o(\mathbf{x}) = s_\psi(f_\theta(\mathbf{x}, \mathbf{z}))$.

To bridge these 3D predictions with the 2D observations, a differentiable ray-based forward model synthesizes the expected B-mode intensity $\hat{I}_k(\mathbf{r}) = \mathcal{F}(\mathbf{r}, \mathbf{a})$. By explicitly calculating signal transmission $\mathcal{T}(\mathbf{x})$ along ray $\mathbf{r}$, the model gains an implicit awareness of occlusions, inherently capturing acoustic shadowing without explicit labels and learned latent prior completes the unobserved 3D shape $o(\mathbf{x} \mid \mathbf{z}^*)$.

2.2 Joint Acoustic and Geometric Representation

To capture the codependence between anatomical structure and its acoustic signature, we define a unified latent space $\mathcal{Z} \subset \mathbb{R}^d$. For any 3D coordinate $\mathbf{x} \in \mathbb{R}^3$ and subject latent code $\mathbf{z} \in \mathcal{Z}$, a shared backbone f_θ learns a joint structural prior: $\mathbf{h}(\mathbf{x}) = f_\theta(\mathbf{x}, \mathbf{z})$. This shared feature vector $\mathbf{h}(\mathbf{x})$ branches into two specialized heads. The acoustic head g_ϕ outputs physical tissue properties for wave propagation, $\mathbf{a}(\mathbf{x}) = g_\phi(\mathbf{h}(\mathbf{x})) = [\beta, \sigma, \mu]^T \geq 0$. Concurrently, the occupancy (geometric) head s_ψ yields the continuous surface occupancy, $o(\mathbf{x}) = s_\psi(\mathbf{h}(\mathbf{x})) \in [0, 1]$. By forcing both properties to decode from the shared representation $\mathbf{h}(\mathbf{x})$ conditioned on the same latent space $\mathcal{Z}$, the network intrinsically couples the spaces, projecting the acoustic optimization onto a manifold strictly constrained by the anatomical geometry.

2.3 Physics-Aware Ray-Based Rendering

To optimize the latent code $\mathbf{z}$ directly from B-mode images, we map the 3D acoustic field $\mathbf{a}(\mathbf{x})$ to the 2D image plane using the differentiable physical rendering formulation established in ultrasound NeRFs [15,16]. Unlike standard optical neural rendering that integrates an entire ray into a single pixel, an ultrasound transducer ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ corresponds to a complete 1D scanline. Here, $\mathbf{o}$ is the origin at the transducer (derived from pose T_k), $\mathbf{d}$ is the unit directional vector, and $t \geq 0$ represents the axial depth along the ray.

The remaining acoustic energy, or transmission factor $\mathcal{T}(t)$, reaching depth t is determined by integrating both the predicted attenuation field $\mu(\mathbf{x})$ and the accumulated reflection $\beta(\mathbf{x})$ from the transducer up to that specific depth. This explicitly models acoustic energy conservation:

$$\mathcal{T}(t) = \exp\left(-\int_0^t \big(\mu(\mathbf{r}(s)) + \beta(\mathbf{r}(s))\big)\, ds\right) \tag{1}$$

The expected B-mode intensity $\hat{I}_k(\mathbf{r}(t))$ at depth t is synthesized as the sum of the localized reflection $\beta(\mathbf{x})$ and scattering $\sigma(\mathbf{x})$, modulated by the remaining acoustic energy at that location:

$$\hat{I}_k(\mathbf{r}(t)) = \mathcal{T}(t)\big(\beta(\mathbf{r}(t)) + \sigma(\mathbf{r}(t))\big) \tag{2}$$

Crucially, when $\mathbf{r}(t)$ intersects highly reflective and attenuating structures such as vertebral bone at a specific depth ($t = t_{bone}$), the sum of $\mu(\mathbf{r}(t_{bone}))$ and $\beta(\mathbf{r}(t_{bone}))$ increases. This causes $\mathcal{T}(t)$ to exponentially decay toward zero for all subsequent depths $t > t_{bone}$. Because the synthesized intensity $\hat{I}_k$ at these deeper locations is multiplied by $\mathcal{T}(t) \approx 0$, the gradient flow to the acoustic parameters in the shadowed regions effectively vanishes during backpropagation. Consequently, the network inherently isolates unobserved anatomy without requiring explicit shadow segmentation, forcing the shape head to rely entirely on the learned latent prior to complete the occluded geometry.

2.4 Optimization Strategy: Training and Inference

Our framework operates in two distinct phases: global prior learning across the dataset and label-free test-time optimization (TTO) for unseen subjects.

During the training phase, the network parameters $\{\theta, \phi, \psi\}$ and the dataset latent codes $\{\mathbf{z}_j\}_{j=1}^{N}$ are jointly optimized. The total training objective $\mathcal{L}_{opt}$ is a weighted sum of four components: a photometric loss $\mathcal{L}_{photo}$, a geometric occupancy loss $\mathcal{L}_{occ}$, an acoustic regularization $\mathcal{L}_{reg\_a}$, and a latent regularization $\mathcal{L}_{reg\_z}$:

$$\mathcal{L}_{opt} = \mathcal{L}_{photo} + \lambda_{occ}\mathcal{L}_{occ} + \lambda_a\mathcal{L}_{reg\_a} + \lambda_z\mathcal{L}_{reg\_z} \tag{3}$$

The photometric loss $\mathcal{L}_{photo}$ measures the reconstruction quality of the synthesized B-mode images. To accurately capture both high-frequency structural details and low-frequency intensity variations, we formulate it as a weighted combination of the Structural Similarity Index Measure (SSIM) and L2 loss:

$$\mathcal{L}_{photo} = \alpha\big(1 - \mathrm{SSIM}(\hat{I}, I)\big) + (1-\alpha)\|\hat{I} - I\|_2^2 \tag{4}$$

where I and $\hat{I}$ are the ground-truth and synthesized B-mode image patches, respectively, and α balances the two terms.

The geometric shape is supervised via the Binary Cross-Entropy (BCE) loss between the predicted occupancy $o(\mathbf{x})$ and the ground-truth shapes $\hat{o}(\mathbf{x})$:

$$\mathcal{L}_{occ} = \sum_{\mathbf{x}\in\Omega} \mathrm{BCE}\big(o(\mathbf{x}), \hat{o}(\mathbf{x})\big) \tag{5}$$

To constrain the physical rendering, $\mathcal{L}_{reg\_a}$ regularizes the predicted acoustic parameter space. Finally, we apply standard L2 regularization to the latent codes, $\mathcal{L}_{reg\_z} = \|\mathbf{z}\|_2^2$, which enforces well-behaved latent manifold.

During inference (TTO) on an unseen subject, the trained network parameters $\{\theta, \phi, \psi\}$ are strictly frozen. We initialize a new latent vector $\mathbf{z}$ and optimize it without any geometric supervision. The TTO objective $\mathcal{L}_{tto}$ relies solely on the self-supervised photometric loss and the regularizers:

$$\mathcal{L}_{tto}(\mathbf{z}) = \mathcal{L}_{photo}(\hat{I}(\mathbf{z}), I) + \lambda_a\mathcal{L}_{reg\_a}(\mathbf{a}(\mathbf{x} \mid \mathbf{z})) + \lambda_z\|\mathbf{z}\|_2^2$$

The optimal latent code $\mathbf{z}^*$ is then obtained by minimizing this objective:

$$\mathbf{z}^* = \arg\min_{\mathbf{z}} \mathcal{L}_{tto}(\mathbf{z})$$

Because the frozen prior tightly couples the acoustic and geometric spaces, optimizing $\mathbf{z}^*$ strictly on the visible, unshadowed anatomy implicitly infers the global structure. The completed 3D vertebral geometry is then directly extracted by querying the occupancy field $o(\mathbf{x} \mid \mathbf{z}^*)$ across the spatial grid.

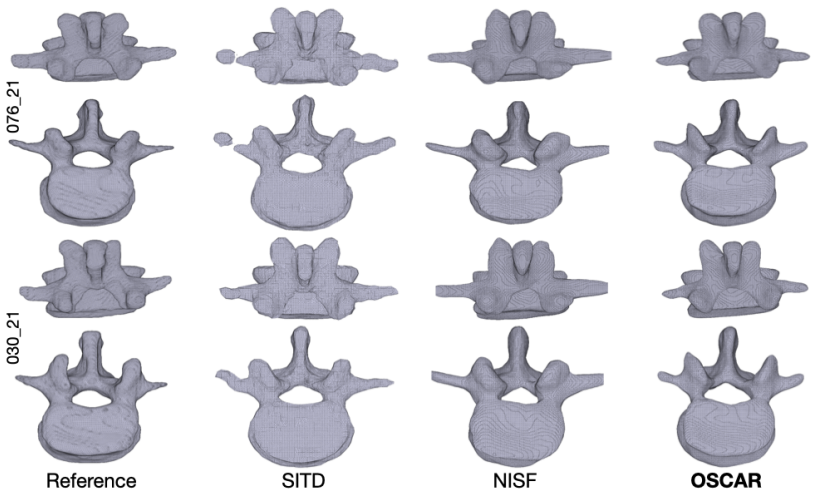

Fig. 2. Visual analysis of reconstructed meshes. OSCAR shows better structural awareness and accurate completion of invisible anatomy compared to NISF (backbone) and SITD (which strictly requires segmentation inputs).

3 Experiments and Results

3.1 Experimental Setup

Data Learning a rich shape prior requires extensive data. We therefore train and benchmark primarily on simulations, while utilizing physical phantoms to evaluate out-of-distribution generalization. **Simulated Data:** Using ImFusion Suite [4] simulation [10], we randomized acoustic properties to generate realistic B-mode images for 53 VerSe [11] vertebrae. We simulated 5 sweeps per subject (perpendicular and $\sim 15^\circ$ tilted), producing 132 training, 10 validation, and 10 test sequences (500 images each with poses and occupancy labels). **Phantom Data:** We scanned three 3D-printed, unused in simulation, VerSe mesh embedded in a tissue-mimicking gelatine-paper pulp phantom [6]. A robot-mounted linear probe acquired the B-mode images, which were then registered to the original 3D model for ground truth comparison.

Network Architecture & Training & TTO Our auto-decoder is an 8-layer MLP backbone (128 hidden units, ReLU) processing the concatenation of a 3D coordinate $\mathbf{x}$ and a 128-dimensional latent vector $\mathbf{z}$. We explicitly omit positional encoding to prevent overfitting to ultrasound speckle noise, yielding smoother surfaces. The backbone branches at the last layer into an acoustic head for physics-aware rendering and an occupancy head. Initial latent codes are sampled from $\mathcal{N}(0, 10^{-3})$ and L_2-regularized. Models are trained for 10 epochs using the

[4] ImFusion Suite, version 3.13.8

Table 1. Quantitative evaluation of reconstructed meshes compared to ground-truth for 10 simulation and 3 phantom test subjects. The methods where trained on simulation data. We compare OSCAR against the baseline backbone (NISF) and SOTA approach (SITD). Performance is reported as Mean ± Std across three metrics: 95% Hausdorff Distance (HD95), Bidirectional Chamfer Distance (CD), and F1-score. In contrast to our method, SITD requires annotations at inference.

	Model	Label	HD95(mm) ↓	CD ↓	F1 ↑
Sim.	SITD	✓	5.93 ± 0.94	119.87 ± 17.04	0.21 ± 0.04
	Nisf	×	3.03 ± 0.92	88.38 ± 14.43	0.33 ± 0.05
	Oscar	×	$\mathbf{1.17 \pm 0.38}$	$\mathbf{56.57 \pm 6.64}$	$\mathbf{0.54 \pm 0.06}$
Phan.	SITD	✓	7.17 ± 1.05	104.66 ± 20.50	0.23 ± 0.07
	Nisf	×	8.61 ± 2.85	146.98 ± 22.57	0.23 ± 0.01
	Oscar	×	7.46 ± 3.41	105.34 ± 20.99	0.27 ± 0.05

Adam optimizer with a learning rate of 10^{-4}. We use validation set to select the best checkpoint and the optimal number of TTO iterations (4k for NISF, 6k for OSCAR). For TTO, the latent code is initialized by the mean of the prior distribution and the optimization updates the latent code to match the B-mode.

3.2 Evaluation

Comparison with State of the Art We benchmark OSCAR against the current SOTA in B-mode ultrasound shape completion, Shape Completion in the Dark [5] (SITD), and our backbone architecture, NISF [12]. Notably, while NISF was originally designed strictly for segmentation, we demonstrate for the first time its inherent capacity for shape completion. Since NISF does not use acoustic features, it effectively provides a geometry-only ablation. As shown in Table 1, our framework outperforms both baselines, yielding overall improvements of 80% and 61% in HD95, respectively. Furthermore, Figure 2 provides a detailed quantitative and visual analysis of two representative samples from the test set. The gain over the baseline NISF proves that aligning acoustic and geometric features and optimizing strictly on visible anatomy better leverages the latent prior to complete occluded regions. Similarly, the improvement over SITD highlights that utilizing a continuous occupancy representation provides the structural awareness necessary for highly detailed shape recovery. In the phantom data, our label free method achieves comparable results to SITD which requires explicit labels.

Latent Prior To demonstrate that our learned prior captures a continuous anatomical manifold, we perform linear interpolation between the mean latent vector of the training set and a specific target shape. As shown in Figure 3, the interpolation is highly smooth, with the mean shape seamlessly morphing into the target geometry. Every intermediate step yields a valid vertebral structure, confirming that the latent space enforces strict anatomical plausibility.

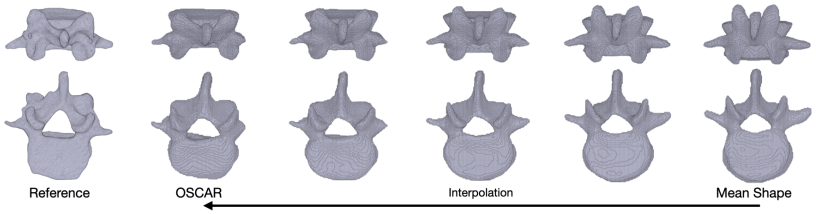

Fig. 3. Latent interpolation We morph the mean shape into a target shape by linearly interpolating their respective latent codes. The anatomically plausible intermediate states show that our framework learns a continuous and valid geometric prior.

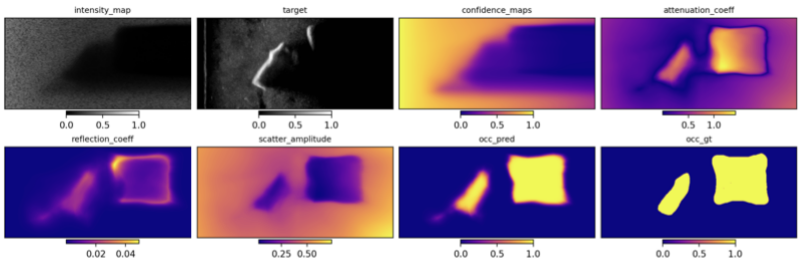

Fig. 4. Bidirectional representation. The acoustic space is accurately predicted by optimizing the latent code $\mathbf{z}$ using only the target shape, demonstrating a strong structural and acoustic coupling.

Acoustic Completion To prove the shared latent space is bidirectional, we invert our optimization to update $\mathbf{z}$ using only the geometric shape. This successfully reconstructs the full acoustic space as visualized in Figure 4 on the example of the phantom image.

Contribution of individual components. OSCAR integrates a forward synthesis model [1,15,16] that incorporates scattering (σ), attenuation (μ), and reflection (β) as an intermediate step to improve shape completion accuracy. Ablation shows that β captures boundary echoes but cannot reconstruct B-mode images independently due to a loss of bidirectional coupling. Furthermore, while β partially compensates for the removal of μ, accurate image synthesis strictly requires μ to be coupled with σ.

Out-of-distribution Shape Completion Because TTO optimizes solely on B-mode intensities, shifting distributions between simulated and real domains present a fundamental challenge. Nevertheless, Figure 5 demonstrates that our framework bridges this sim-to-real gap, accurately completing shapes on physical phantoms using only simulations during training.

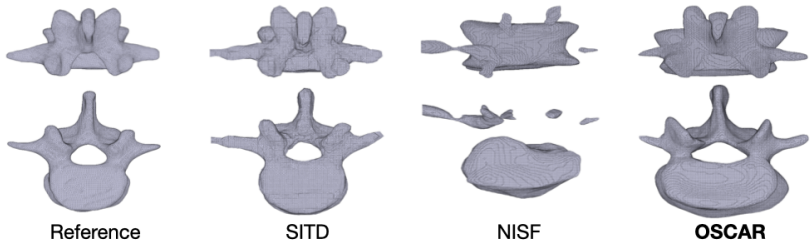

Fig. 5. Evaluation on phantom data. Our framework accurately completes shapes from real B-modes despite the sim-to-real domain gap. In contrast, SITD avoids this OoD challenge only by requiring explicit point cloud labels.

4 Conclusion

In this work, we introduced OSCAR, a physics-aware framework for ultrasound-based 3D spinal shape completion. By integrating acoustic image formation and geometric occupancy into a coupled-latent implicit neural representation, OSCAR resolves multi-view inconsistencies and accurately reconstructs occluded anatomy. Crucially, while current SOTA methods rely on explicitly labeled surfaces or segmentation masks, our approach optimizes directly on raw B-mode image intensities. This direct optimization yields an 80% improvement in the HD95 over the SOTA and enforces anatomical plausibility using an implicit prior. Preliminary evaluations on physical phantoms confirm that OSCAR can successfully translate these capabilities to real-world scans. Finally, the bidirectional capacity of our learned joint prior enables the synthesis of plausible acoustic spaces directly from known shapes, with the potential for more scalable registration and data synthesis in US imaging.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Burger, B., Bettinghausen, S., Radle, M., Hesser, J.: Real-time gpu-based ultrasound simulation using deformable mesh models. IEEE transactions on medical imaging **32**(3), 609–618 (2012)
2. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 721–731. Springer (2024)
3. Duque, V.G., Marquardt, A., Velikova, Y., Lacourpaille, L., Nordez, A., Crouzier, M., Lee, H.J., Mateus, D., Navab, N.: Ultrasound segmentation analysis via distinct and completed anatomical borders. International Journal of Computer Assisted Radiology and Surgery **19**(7), 1419–1427 (2024)
4. Gafencu, M.A., Velikova, Y., Navab, N., Azampour, M.F.: Us-x complete: A multimodal approach to anatomical 3d shape recovery. In: International Workshop on Shape in Medical Imaging. pp. 218–231. Springer (2025)
5. 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 (2024)
6. Jiang, Z., Li, Z., Grimm, M., Zhou, M., Esposito, M., Wein, W., Stechele, W., Wendler, T., Navab, N.: Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback. IEEE Transactions on Industrial Electronics **69**(7), 7064–7075 (2021)
7. Massalimova, A., Liebmann, F., Jecklin, S., Carrillo, F., Farshad, M., Fürnstahl, P.: Surgpointtransformer: transformer-based vertebra shape completion using rgbd imaging. Computer Assisted Surgery **30**(1), 2511126 (2025)
8. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
9. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
10. Salehi, M., Ahmadi, S.A., Prevost, R., Navab, N., Wein, W.: Patient-specific 3d ultrasound simulation based on convolutional ray-tracing and appearance optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 510–518. Springer (2015)
11. Sekuboyina, A., Husseini, M.E., Bayat, A., Löffler, M., Liebl, H., Li, H., Tetteh, G., Kukačka, J., Payer, C., Štern, D., et al.: Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images. Medical image analysis **73**, 102166 (2021)
12. Stolt-Ansó, N., McGinnis, J., Pan, J., Hammernik, K., Rueckert, D.: Nisf: Neural implicit segmentation functions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 734–744. Springer (2023)
13. Vyas, K., Veeraraghavan, A., Balakrishnan, G.: Fit pixels, get labels: Meta-learned implicit networks for image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 194–203. Springer (2025)

14. de Wilde, B., Rietberg, M.T., Lajoinie, G., Wolterink, J.M.: Steerable anatomical shape synthesis with implicit neural representations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 630–639. Springer (2025)
15. Wysocki, M., Azampour, M.F., Eilers, C., Busam, B., Salehi, M., Navab, N.: Ultra-nerf: Neural radiance fields for ultrasound imaging. In: Medical Imaging with Deep Learning. pp. 382–401. PMLR (2024)
16. Wysocki, M., Duelmer, F., Bal, A., Navab, N., Azampour, M.F.: Ultron: Ultrasound occupancy networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 606–615. Springer (2025)
17. Yang, J., Sedykh, E., Adhinarta, J.K., Le, H., Fua, P.: Generating anatomically accurate heart structures via neural implicit fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 264–274. Springer (2024)
18. Yang, J., Wickramasinghe, U., Ni, B., Fua, P.: Implicitatlas: learning deformable shape templates in medical imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15861–15871 (2022)