UltrON: Ultrasound Occupancy Networks

Magdalena Wysocki[1,3][0009−0008−8239−1629], Felix Duelmer[1,3][0009−0000−1702−1746], Ananya Bal[2][0000−0002−9592−1085], Nassir Navab[1,3][0000−0002−6032−5611], and Mohammad Farid Azampour[1,3][2222−−3333−4444−5555]

[1] Chair for Computer Aided Medical Procedures (CAMP)
Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany
{magdalena.wysocki, felix.duelmer, nassir.navab, mf.azampour}@tum.de
[2] Robotics Institute
Carnegie Mellon University, 5000 Forbes Avenue Pittsburgh, PA 15213, USA
abal@andrew.cmu.edu
[3] Munich Center for Machine Learning (MCML), Munich, Germany

Abstract. In free-hand ultrasound imaging, sonographers rely on expertise to mentally integrate partial 2D views into 3D anatomical shapes. Shape reconstruction can assist clinicians in this process. Central to this task is the choice of shape representation, as it determines how accurately and efficiently the structure can be visualized, analyzed, and interpreted. Implicit representations, such as SDF and occupancy function, offer a powerful alternative to traditional voxel- or mesh-based methods by modeling continuous, smooth surfaces with compact storage, avoiding explicit discretization. Recent studies demonstrate that SDF can be effectively optimized using annotations derived from segmented B-mode ultrasound images. Yet, these approaches hinge on precise annotations, overlooking the rich acoustic information embedded in B-mode intensity. Moreover, implicit representation approaches struggle with the ultrasound's view-dependent nature and acoustic shadowing artifacts, which impair reconstruction. To address the problems resulting from occlusions and annotation dependency, we propose an occupancy-based representation and introduce Ultrasound Occupancy Network (UltrON) that leverages acoustic features to improve geometric consistency in weakly-supervised optimization regime. We show that these features can be obtained from B-mode images without additional annotation cost. Moreover, we propose a novel loss function that compensates for view-dependency in the B-mode images and facilitates occupancy optimization from multiview ultrasound. By incorporating acoustic properties, UltrON generalizes to shapes of the same anatomy. We show that UltrON mitigates the limitations of occlusions and sparse labeling and paves the way for more accurate 3D reconstruction. Code and dataset is available at https://github.com/magdalena-wysocki/ultron.

Keywords: Ultrasound · Implicit Neural Representation · Surface Reconstruction.

1 Introduction

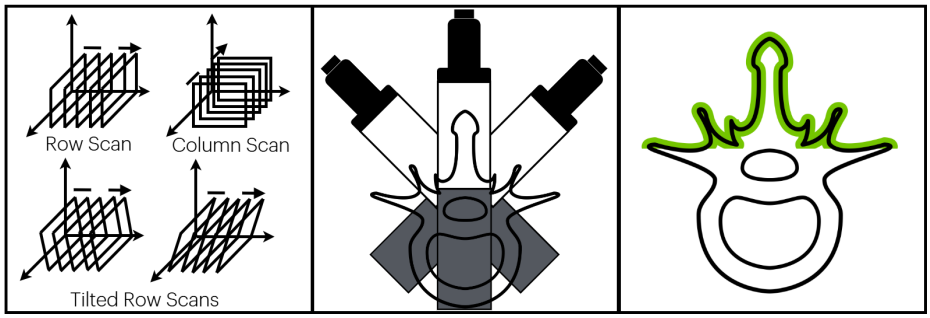

Fig. 1. (a) The method uses multiview ultrasound scans. We use row and column scans as proposed in RoCoSDF [3] and tilted scans as proposed in Ultra-NeRF [16]. (b) Occlusions in ultrasound B-mode imaging create partial observations. The regions in acoustic shadow (gray) are undefined. (c) Because of the occlusions we can define only the partial shape (green).

In conventional free-hand ultrasound imaging, sonographers must rely on their expertise to mentally reconstruct the 3D shape of organs or structures from partial 2D views in order to extract the necessary anatomical information for diagnosis or intervention. Shape reconstruction in medical ultrasound, can help clinicians in this task by enhanced visualization of a target structure. In the field of computer vision, objects can be represented through a diverse array of shape representations. Among them, implicit shape representations provide a method for representing 3D shapes as an implicit function defining the surface of an object. Unlike conventional representations such as meshes, point clouds, or voxel grids, implicit representations map spatial coordinates to a continuous function that describes the shape [11]. Implicit Neural Representation of Shapes (INRS) is a deep-learning-based technique for approximating implicit shape functions. The key concept is that the function is parameterized by a neural network. INRS inherently provides a continuous representation, allowing it to capture fine details at any resolution without increasing memory usage. These representations define inherently smooth surfaces, making them useful for tasks requiring seamless shape transformations [19]. Moreover, they enable learning families of shapes, facilitating the generation of novel shapes [18], [8]. Among INRS methods, Deep Signed Distance Field (SDF) [13] and Occupancy Network (ON) [10] are two widely used approaches. In medical ultrasound, recent methods—UNSR [4] and RoCoSDF [3] —leverage SDF-based neural networks to represent anatomical structures from B-mode images. While UNSR relies on single-view ultrasound data, RoCoSDF demonstrates that incorporating multiple ultrasound views leads to more precise 3D shape reconstruction.

Multiview ultrasound scanning, as shown in Fig 1(a) combines perspectives from various angles, providing a more complete spatial representation and better visualization of complex anatomical structures. However, non-tomographic modalities like ultrasound B-mode imaging are inherently view-dependent and susceptible to acoustic shadowing occlusions, yielding only fragmented observations of the volume [5]. The direction-dependent characteristics of ultrasound imaging create additional challenges for learning a shape representation, as scanning the same target from different directions can result in varying signal intensities in the same spatial location. Moreover, occlusions, visualized in Fig 1(b), make the regions under acoustic shadows undefined, therefore the shape can be only defined in the observed space (Fig 1(c)). Since UNSR and RoCoSDF rely on densely annotated B-mode images for accuracy, they may encounter issues with occlusions, annotation errors, and partial annotations, which are common in B-mode imaging.

In this paper we propose UltrON a novel INRS for ultrasound that integrates acoustic features obtained from B-mode intensities into the representation to address the issues of annotation cost and accuracy. Our contributions can be summarized as followed:

- We demonstrate that by using the information in the B-mode intensities and without additional labels, the proposed method reduces the supervision required for learning INRS from ultrasound scans by 90%.
- We tackle the problem of partial observations due to occlusions, by introducing an attenuation-compensated loss function. This enables optimization directly from multiview annotations.
- By integrating acoustic features into the occupancy function, the proposed method generalizes effectively, learning INRS across different volumes of the same anatomical structure.

2 Method

2.1 Overview

Our objective is to reconstruct the surface of a medical shape from multiview ultrasound scans. As shown in Fig. 2, the proposed method follows a three-step approach. First, we optimize a neural field of tissue-specific acoustic properties (attenuation, reflection, scattering). Second, we optimize an occupancy network that maps these properties to occupancy. Finally, we sample the optimized occupancy network and apply Marching Cubes [9] algorithm to extract the 3D mesh.

2.2 Neural Field of Acoustic Features

Our approach is based on the insight that the same tissue type exhibits consistent acoustic features, therefore if one knows distribution of these features for a specific tissue it can be used to localize this tissue in space. In B-mode images,

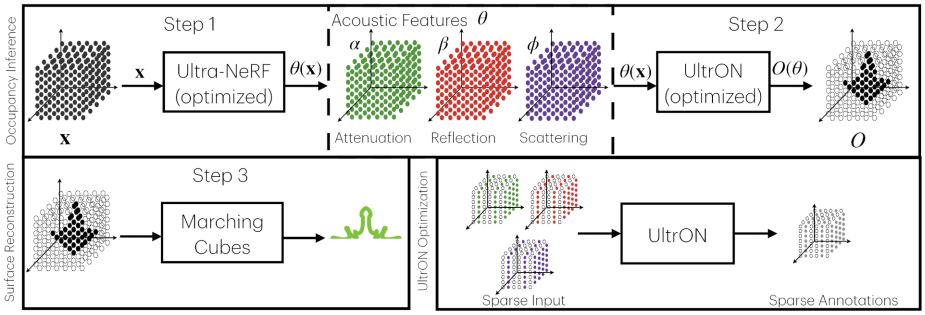

Fig. 2. Overview of the proposed three-step approach for medical shape representation and surface reconstruction from multiview ultrasound scans. First, a neural field of tissue-specific acoustic properties $\boldsymbol{\theta}$—attenuation (α), reflection (β), and scattering (ϕ)—is optimized. Second, an UltrON is optimized to map these properties to occupancy ($o(\boldsymbol{\theta})$). Finally, the optimized occupancy network is sampled, and the Marching Cubes [3] algorithm is applied to extract the 3D mesh. Ultra-NeRF is optimized using the full 3D volume as supervision, whereas UltrON is optimized using sparse 2D ultrasound annotations converted to occupancy.

we observe only pixel intensities as effects of acoustic features, not the features themselves. To learn these features from multiview B-mode scans we employ Ultra-NeRF. Ultra-NeRF is a neural rendering framework that allows synthesis of B-mode images from unobserved viewing points. Additionally it provides a distribution of the three acoustic features—namely, attenuation (α), reflection (β), and scattering (ϕ)-in space. However, the mapping between the distribution and the tissue type is unknown. To find this mapping we propose using an occupancy-based method. Since these features are approximately homogeneous within the same tissue type, we demonstrate that optimizing this mapping requires fewer annotations compared to direct coordinate-to-occupancy mapping.

2.3 Ultrasound Occupancy Network

Occupancy function o of a 3D object is defined as a function that for every 3D point $\boldsymbol{x} \in \mathbb{R}^3$ maps this point to an occupancy value:

$$o : \mathbb{R}^3 \rightarrow \{0, 1\} \tag{1}$$

For a single 3D object, ON [10] is a neural implicit representation used to model 3D object by predicting the occupancy probability of points in continuous space. Given a spatial coordinate $\boldsymbol{x} \in \mathbb{R}^3$, an occupancy network approximates the occupancy function:

$$f_\omega : \mathbb{R}^3 \rightarrow [0, 1] \tag{2}$$

where $f_\omega(\boldsymbol{x})$ represents the probability that the point $\boldsymbol{x}$ is inside the object. In this paper, we introduce UltrON, which extends the standard ON by integrating

acoustic information into the representation. Specifically, instead of relying solely on spatial coordinates, the ultrasound occupancy function o_u is reformulated as

$$o_u : \mathbb{R}^d \to \{0, 1\}, \tag{3}$$

$$o_u(\boldsymbol{\theta}(\boldsymbol{x})) = \begin{cases} 1 & \text{if occupied} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We approximate o_u by the network f_ω and therefore $f_\omega(\boldsymbol{\theta}(\boldsymbol{x})))$ represents the probability that the point $\boldsymbol{x}$ is inside the object given acoustic properties at the point $\boldsymbol{x}$ and $\boldsymbol{\theta}(\boldsymbol{x}) \in \mathbb{R}^d$ is the vector of acoustic properties at point $\boldsymbol{x}$. Since acoustic properties are largely homogeneous within a given tissue type but vary across different tissues, incorporating this information provides a more structured representation of anatomical regions.

2.4 Attenuation-compensated Optimization

Ultrasound is view-dependent, therefore during the training we account for acoustic attenuation when comparing the network output with ground truth labels, ensuring accurate tissue identification across different imaging angles. To this end, we define the loss function based on the binary cross-entropy (BCE) loss that considers effects of attenuation along the propagation path of the ultrasound beam. The rationale behind this is that some regions within the volume of interest may or may not be observed depending on the probe's position and orientation since occlusions can prevent the signal from reaching certain points, leading to incomplete or inaccurate observations. The resulting loss at a given point $\mathbf{x}$ is computed as follows:

$$\mathcal{L}(\boldsymbol{x}) = -\left[y(\boldsymbol{x}) \cdot \log(T(\boldsymbol{x}) \cdot f_\omega(\boldsymbol{\theta}(\boldsymbol{x}))) + (1 - y(\mathbf{x})) \cdot \log(1 - T(\boldsymbol{x}) \cdot f_\omega(\boldsymbol{\theta}(\boldsymbol{x}))\right] \tag{5}$$

where $f_\omega(\boldsymbol{\theta}(\boldsymbol{x}))$ is the occupancy function, $y(\boldsymbol{x})$ is the ground truth label, and $T(\boldsymbol{x})$ is the transmittance function at point $\boldsymbol{x}$. $T(\boldsymbol{x})$ accounts for the transmission of the acoustic signal based on the probe position and orientation. It reflects occlusions and the visibility of the volume from the given probe viewpoint and is defined as:

$$T(\boldsymbol{x}) = T(0) \cdot \exp^{-\int_0^{x-\epsilon} \beta(n)dn} \cdot \exp^{-\int_0^{x-\epsilon} \alpha(n)dn} \tag{6}$$

To compute $T(\boldsymbol{x})$ we use the ray casting and integrate attenuation α and reflection β along propagation path of the ultrasound beam from transducer to the current position $\boldsymbol{x}$ within the volume. Similar integration of reflection and attenuation has been proposed in [20], [16], [14].

3 Experiments and Results

3.1 Data

Four CAD vertebra models of lumbar spine vertebrae (L2, L3x2, L4), created from the VerSe dataset [15], are used as phantoms for 3D printing. These models

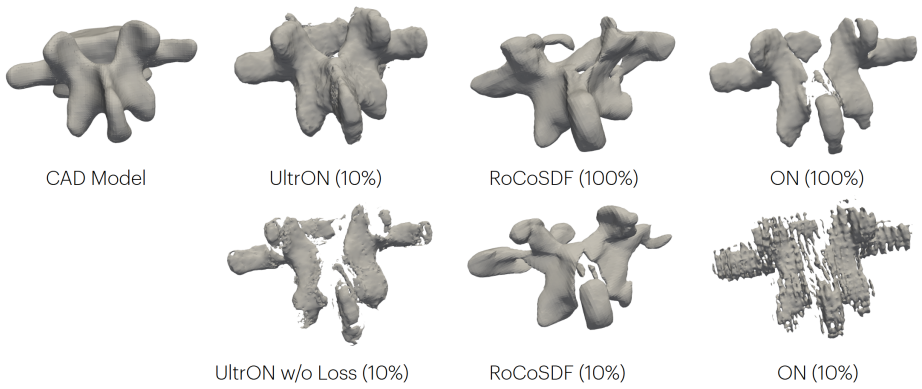

Fig. 3. Visualization of an example if reconstruction of L3 vertebra model from different INRS optimized with dense (100%) and weak supervision (10%). We observe that UltrON optimized on 10% of manual annotations preserves topology more accurately than RoCoSDF and ON optimized on 100% annotations. This shows that UltrON compensates errors in annotations naturally occurring due to tracking errors, occlusions and human errors. Comparison between UltrON with (upper) and UltrON without (bottom) the attenuation-compensated loss shows that the loss compensates acoustic shadows in multiview ultrasound of a highly reflective structure.

are then used to create ballistic gelatin-based phantoms with paper pulp that mimic soft tissue [6], offering better realism compared to models in a water bath. Similar to RoCoSDF, we performed multiview scanning that includes row and column scans, and in addition tilted scans as well. multiview scanning integrates multiple viewpoints as illustrated in Fig 1(a). The tilted scans involve adjusting the probe by -10 and +10 degrees from the standard row scan position. For acquisition, we use a probe mounted on a robotic arm. The robotic tracking system is calibrated to account for the offset caused by the probe's attachment to the arm. After data acquisition, the B-mode images are manually segmented. This segmentation is then converted to occupancy data, where 0 is background label and 1 is bone label. Using the poses corresponding to the B-mode images, we compute the voxel positions in space and normalize them to fit within a unit cube. For training RoCoSDF, we follow the procedure outlined in [3] to generate the point clouds. For Ultra-NeRF, we utilize the poses and B-mode images as described in [16].

3.2 Implementation Details

f_ω, the occupancy function in UltrON, and Ultra-NeRF, use the same architecture which consists of 8 fully-connected layers with 128 hidden channels and with the skip connection at the fourth layer. We use ReLU activation layers except for the last layer. We follow 2-stage optimization. In the first stage, we optimize Ultra-NeRF for 75k iterations, then we optimize UltrON for 50k iterations. We

Table 1. Comparison of the shape reconstruction based on different representations optimized with dense (100%) and weak supervision (10%) as measured by Chamfer Distance (CD), Hausdorff Distance (HD), Mean Absolute Deviation (MAD), and Root Mean Square Error (RMSE). We observe that UltrON with 10% annotations outperforms RoCoSDF trained on the whole available data by a margin of 26% as measured by CD.

Method (supervision)	CD (mm)↓	HD (mm)↓	MAD (mm)↓	RMSE (mm)↓
RoCoSDF [3](100%)	2.98 ± 0.03	9.79 ± 0.02	2.53 ± 0.02	3.67 ± 0.03
ON (100%)	2.91 ± 0.03	9.23 ± 0.02	2.60 ± 0.03	3.70 ± 0.03
UltrON w/o Loss (10%)	3.25 ± 0.12	9.30 ± 0.06	2.82 ± 0.12	3.98 ± 0.15
UltrON (10%)	**2.22 ± 0.02**	**7.98 ± 0.04**	**1.67 ± 0.02**	**2.69 ± 0.03**
UltrON (5%)	2.36 ± 0.03	8.04 ± 0.05	1.85 ± 0.03	2.88 ± 0.03

use the Adam optimizer with 0.0001 learning rate and exponential decay. For training Ultra-NeRF we use regularized version of the method [20] and default settings. We apply positional encoding method presented in NeRF [12] to the input of the network. Our network is implemented using Pytorch and trained on NVIDIA RTX 3090 GPU with 24 GB memory. Smoothing is applied to the occupancy, and after smoothing the threshold for Marching Cubes is set to 0 to extract the zero-level-set of the surface. For the smoothing and Marching Cubes we use PyMCubes[4] implementation. For ON, we use the same architectur as f_ω but we change the network input to coordinates.

3.3 Evaluation Method

We compare the method to RoCoSDF, the state-of-the-art method in INRS for ultrasound imaging, and coordinate-based ON. Four evaluation metrics are used to assess reconstruction quality: Chamfer Distance (CD), Hausdorff Distance (HD), Mean Absolute Distance (MAD), and Root Mean Square Error (RMSE). These metrics are computed by calculating the distances between points randomly sampled from the reconstructed mesh and the corresponding CAD models. To further asses the reconstructed volumes we compare the reconstructed surfaces visually with respect to CAD models.

3.4 Qualitative and Quantitative Results

As shown in Table 1, UltrON improves surface reconstruction by 26% (CD) over RoCoSDF and coordinate-based ON and requires 90% less annotations to achieve this performance. Quantitative results in Fig. 3 show that UltrON better preserves topology even in sparse data regimes since it is using dense information about distribution of the acoustic features within the volume. We

[4] https://github.com/pmneila/PyMCubes

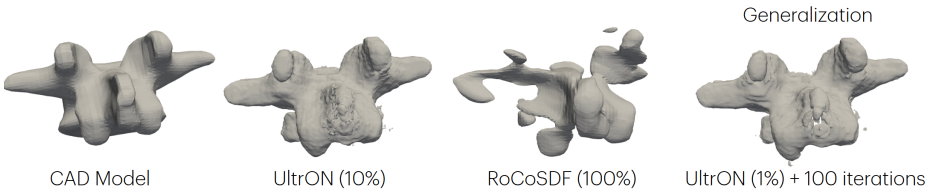

Fig. 4. Visualization of the reconstructed vertebra shapes with UltrON and RoCoSDF in the presence of larger annotation errors and visualization of generalization to the new shape (trained on L3 and fine-tuned on L2).

observe that UltrON is moreover more robust to errors in input data resulting from annotation errors and in the misalignment of the input data. We observe that with these errors RoCoSDF fails to preserve surface topology whereas using acoustic features directly helps UltrON to correct these errors. We show an example qualitatively in Fig 4.

Loss Ablation To visualize the importance of the attenuation compensation we perform ablation on the loss. In Table 1 and Fig 3 we show that using a standard BCE loss decreases performance of the method since the optimization does not compensate the occlusions resulting from attenuation. This, as we can see in the Fig 3 results in errors in the topology due to occlusions.

Table 2. Performance on generalization to a new shape of the same anatomy based on Chamfer Distance (CD), Hausdorff Distance (HD), Mean Absolute Deviation (MAD), and Root Mean Square Error (RMSE).

Supervison % + # iter.	CD (mm)↓	HD (mm)↓	MAD (mm)↓	RMSE (mm)↓
w/o fine-tunning	3.44 ± 0.02	13.23 ± 0.07	1.79 ± 0.02	2.67 ± 0.03
1% + 100 iter	2.44 ± 0.02	7.91 ± 0.04	2.00 ± 0.02	2.94 ± 0.03

Shape Generalization To test the generalization, we need to account for real-world variations, as well as variations resulting from the ill-posed nature of optimizing a neural field. To this end, fine-tuning the network to a new shape of the same anatomical structure is necessary. In this process, the last two layers of the network are frozen, and only 1% of the labels are used. The fine-tuning procedure takes approximately 5 seconds to complete (100 iterations). In Table 2 we show that the reconstruction quality as measured by the four reconstruction metrics is comparable to the full training. Fig 4 presents this observation quantitatively.

4 Conclusion

We present UltrON, a novel approach that integrates acoustic features from B-mode intensities into a representation of occupancy. We show that with 90% fewer annotations, UltrON provides a representation that can enhance reconstruction accuracy by 26%. Moreover, we introduce an attenuation compensated loss function that facilitates optimization directly from multiview annotations and tackles the problem of partial observations due to occlusions in multiview ultrasound. Finally, we demonstrate that incorporating acoustic features into the occupancy function enables generalization to the same anatomy across different volumes with 1% of the supervision and only requires 100 iterations of fine-tuning. To further enhance the generalization of UltrON, one could consider incorporating shape priors into the representation [1], [2]. We also believe that explicitly defining unmeasured regions and incorporating uncertainty quantification can enhance reconstruction quality, as demonstrated through visibility analysis [17]. With these improvements, UltrON has the potential to facilitate the creation of realistic, patient-specific 3D anatomical models that could be utilized by both clinical practitioners and automated systems such as robotic ultrasound [7].

Acknowledgments. This work was supported by the HINAV project funded by the Bavarian State Ministry for Economic Affairs, Regional Development and Energy.

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

References

1. 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)
2. Bastian, L., Baumann, A., Hoppe, E., Bürgin, V., Kim, H.Y., Saleh, M., Busam, B., Navab, N.: S3m: scalable statistical shape modeling through unsupervised correspondences. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 459–469. Springer (2023)
3. 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)
4. Chen, H., Kumaralingam, L., Zhang, S., Song, S., Zhang, F., Zhang, H., Pham, T.T., Punithakumar, K., Lou, E.H., Zhang, Y., et al.: Neural implicit surface reconstruction of freehand 3d ultrasound volume with geometric constraints. Medical Image Analysis **98**, 103305 (2024)
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. Jiang, Z., Salcudean, S.E., Navab, N.: Robotic ultrasound imaging: State-of-the-art and future perspectives. Medical image analysis **89**, 102878 (2023)
8. Kong, F., Stocker, S., Choi, P.S., Ma, M., Ennis, D.B., Marsden, A.L.: Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. Medical Image Analysis **97**, 103293 (2024)
9. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353 (1998)
10. 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)
11. Michalkiewicz, M., Pontes, J.K., Jack, D., Baktashmotlagh, M., Eriksson, A.: Implicit surface representations as layers in neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
12. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
13. 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)
14. 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: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18. pp. 510–518. Springer (2015)
15. 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)
16. 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)
17. Wysocki, O., Xia, Y., Wysocki, M., Grilli, E., Hoegner, L., Cremers, D., Stilla, U.: Scan2lod3: Reconstructing semantic 3d building models at lod3 using ray casting and bayesian networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6548–6558 (2023)
18. 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)
19. 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)

20. Yesilkaynak, V.B., Duque, V.G., Wysocki, M., Velikova, Y., Mateus, D., Navab, N.: Ultrasound confidence maps with neural implicit representation. In: Annual Conference on Medical Image Understanding and Analysis. pp. 89–100. Springer (2024)