Beyond Shadows: Learning Physics-inspired Ultrasound Confidence Maps from Sparse Annotations

Matteo Ronchetti[1,2], Rüdiger Göbl[1], Bugra Yesilkaynak[2], Oliver Zettinig[1], and Nassir Navab[2]

[1] ImFusion GmbH, München, Germany
[2] Computer Aided Medical Procedures (CAMP) Technische Universität München, Germany

Abstract. This paper introduces a novel user-centered approach for generating confidence maps in ultrasound imaging. Existing methods, relying on simplified models, often fail to account for the full range of ultrasound artifacts and are limited by arbitrary boundary conditions, making frame-to-frame comparisons challenging. Our approach integrates sparse binary annotations into a physics-inspired probabilistic graphical model that can estimate the likelihood of confidence maps. We propose to train convolutional neural networks to predict the most likely confidence map. This results in an approach that is fast, capable of dealing with various artifacts, temporally stable, and allows users to directly influence the algorithm's behavior using annotations. We demonstrate our method's ability to cope with a variety of challenging artifacts and evaluate it quantitatively on two downstream tasks, bone shadow segmentation and multi-modal image registration, with superior performance than the state-of-art. We make our training code public.

Keywords: Ultrasound · Confidence Map

1 Introduction

Ultrasound imaging is a widely used diagnostic tool, valued for its non-invasive characteristics, real-time imaging capabilities, and cost-efficiency. Confidence maps aim at offering a quantitative assessment of the reliability of each pixel within the ultrasound image, and have been successfully employed in various applications, including intensity reconstruction [9], volume compounding [2], US-CT registration [19], shadow detection [13] and deep learning segmentation [7]. Of particular interest is the use of confidence maps for probe positioning and contact force optimization in robotic ultrasound [10,1,21,4,18]. Recent methodologies for quantifying uncertainty in B-mode ultrasound images encompass RF-based techniques [13], shadow-specific neural networks [15], Ultra-NeRF based approaches [22,23], and graph-based methodologies [12,11]. Existing methods have different limitations: physics-based approaches overlook artifacts like reverberation [12,11], shadow-based models are restricted by design [15], arbitrary

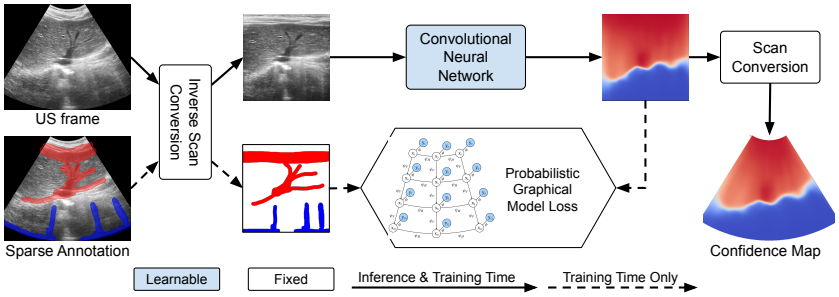


Fig. 1: Overview of our method, showcasing how sparse *Good* (red, high confidence) and *Bad* (blue, low confidence) annotations are utilized to predict confidence maps with a CNN in pre-scan converted space.

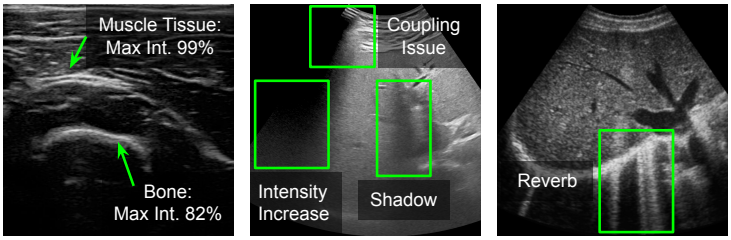
boundary conditions prevent frame-to-frame comparison [12,11], and limited control is offered, as correcting misassigned confidence requires complex modifications to the entire approach. We adopt a user-centered approach, recognizing that ultrasound practitioners can effectively assess confidence. We rely on sparse binary annotations (*Good*/*Bad*), similar to prior segmentation approaches [3,5], and combine them into a physics-inspired probabilistic graphical model (PGM) to estimate the likelihood of confidence maps. We then train a convolutional neural network (CNN) to predict the most likely confidence map. This results in an approach that is fast, capable of dealing with various artifacts, user controllable, and temporally stable.

2 Ideal Confidence Maps

Ultrasound transducers produce focused sound beams that are used to reconstruct each image scanline [8,6]. The intensity of the echo that arrives back at the transducer $f(d)$ for a scanline from depth d can be modeled as

$$f(d) = \underbrace{p_0}_{\substack{\text{Initial}\\ \text{intensity}}} \underbrace{\cdot \exp\left(-\int_0^d r(t) + s(t)dt\right)}_{\text{Attenuation}} \cdot \underbrace{\left(r_{\uparrow}(d) + s_{\uparrow}(d)\right)}_{\substack{\text{Scatter and reflection}\\ \text{toward transducer}}}, \tag{1}$$

where $r(d)$ and $s(d)$ are the coefficients of the total energy reflected and scattered in all directions, $r_{\uparrow}(d)$ and $s_{\uparrow}(d)$ the reflection and scattering coefficients in the direction of the transducer respectively. This model does not account for multi-path scattering / reverb. We also omit the change in incident pressure over depth, caused by transmit focusing, the weakening of the echo due to the inverse square law, as well as absorption, as we argue these effects are generally accounted for in the signal processing chain, most importantly through the user chosen time-gain-compensation. Overall, this model differs from the models used in [22,23], as it considers off-axis reflection and scattering in the attenuation term. Based on this physical model, an ideal confidence map should exhibit these properties:



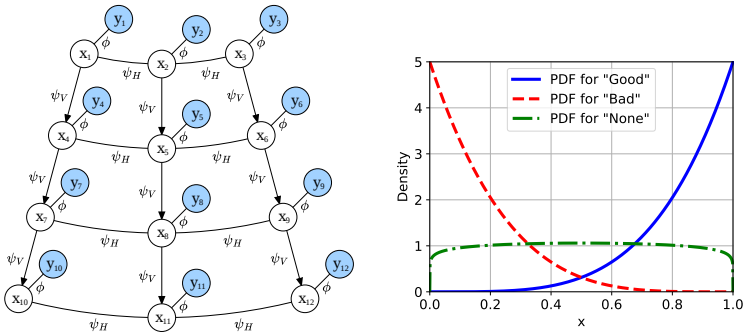
(a) (b) (c)

Fig. 2: Complex relationship between confidence and pixel intensities. (a): tissue that blocks sound (bone) causing a weaker signal than a tissue that doesn't block sound (muscle). (b-c): different common ultrasound artifacts.

1. **Mostly monotonic**. Due to attenuation, the sound beam intensity decreases with depth, and therefore the confidence should also mostly decrease. This is not a hard constraint since a weak sound beam that encounters a strong reflector can still produce a clearly detectable echo.
2. **Loosely related to pixel intensities**. The sound beam attenuation depends on the quantities $r(\cdot)$ and $s(\cdot)$, while the echo depends on $r_{\uparrow}(\cdot)$ and $s_{\uparrow}(\cdot)$. See for example Figure 2a where the muscle tissue, which allows sound transmission, is represented in the B-mode image with a higher brightness than a bone, which blocks all further transmission. As further examples, Figure 2b depicts a shadow appearing without a bright reflection above it and an increase in intensity in a shadowed area due to electronic noise. Therefore, we argue that the relationship between confidence and pixel intensities is complex and cannot be captured by simple models.
3. **Beyond shadows**. Ultrasound images present a variety of artifacts commonly including shadows (Fig. 2b), reverberations (Fig. 2c) and a combination of these when lacking acoustic coupling (Fig. 2b). An ideal confidence map should not only properly handle these but also less common artifacts.
4. **Sound beams-aware**. Confidence map computation should consider the direction of insonication, i.e. the method should compensate accordingly if scanlines are tilted in non-linear fan geometries.
5. **Horizontally smooth**. While the sound beams are narrow, there is still overlap between them due to the width of the point-spread function. Therefore, any horizontal discontinuity in the confidence map would be unrealistic.

3 Approach

Given sparse annotations y, i.e. $y_i \in \{Good, Bad, None\}$, we model the probability of a confidence map x using a Probabilistic Graphical Model (PGM) as shown in Figure 3a. In this model, each confidence map pixel (x_i) is a node in the graph, and its confidence probability distribution depends on its annotation



(a) (b)

Fig. 3: (a) Graphical representation of the PGM used to combine sparse annotations with a physics based prior. Refer to the text for the description of the pairwise potentials ϕ, ψ_V, ψ_H. (b) Plot of the Probability Density Functions (PDF) for the Beta distributions used in the definition of ϕ.

(y_i) and the confidence of its neighboring pixels. According to Property 4 (*see above*), we shape the graph distinguishing between intra- and inter-scanline relationships. The former are shown with a directed arrow since they are causal (Property 1), while the latter are symmetric. Formally, given sparse annotations y, we define the probability of a confidence map x as

$$p(x|y) \propto \underbrace{\prod_i \phi(x_i, y_i)}_{\text{Unary potentials}} \underbrace{\prod_{(i,j)\in V} \psi_V(x_i, x_j) \prod_{(i,j)\in H} \psi_H(x_i, x_j)}_{\text{Pairwise potentials}}, \tag{2}$$

where the pairwise potentials are considered in vertical (V) and horizontal (H) direction, respectively. The reader might find surprising that the ultrasound image intensities are not used by this model. Property 2 suggests that the relation to pixel intensities is too complex to be incorporated in such a simple model. Through the design of pairwise potentials, we thus limit this PGM to only enforce an ultrasound physics prior, and leave the complex pixel intensity relationships to the CNN that is trained on top of the PGM as described in Section 3.3.

3.1 Unary potentials to model sparse annotations

The unary potentials $\phi(x_i, y_i)$ measure the compatibility between the confidence value x_i and the corresponding annotation y_i. We model this using a Beta distribution, whose parameters depend on y_i:

$$\phi(x_i, y_i) = \begin{cases} \text{Beta}(x_i; \alpha = 5, \beta = 1) & \text{if } y_i = \textit{Good} \\ \text{Beta}(1 - x_i; \alpha = 5, \beta = 1) & \text{if } y_i = \textit{Bad} \\ \text{Beta}(x_i; \alpha = 1.1, \beta = 1.1) & \text{if } y_i = \textit{None} \end{cases} \tag{3}$$

As can be seen from Figure 3b, the distributions associated to *Good* and *Bad* annotations are symmetric. We assume that areas with particularly high or low confidence would are likely to be annotated, therefore, for *None* annotations, we assign lower probability to the extremes.

3.2 Ultrasound Physics Prior with Pairwise Potentials

The pairwise potentials ψ_V and ψ_H enforce an ultrasound physics-inspired prior.
Intra-Scanline Potential The asymmetric potential ψ_V acts within scanlines and encourages the confidence along the scanline to be mostly monotonically decreasing according to Property 1. Since confidence is bound below by zero, once a node's confidence is close to zero, subsequent nodes in the same scanline cannot decrease much further. Therefore, a penalty formulated using the confidence values x_i directly would discourage low confidence values from appearing further from the bottom row. To address this issue, we instead employ $\log(x_i)$, which is not bound below. Given a desired decay in confidence between nodes s, we penalize $\log(x_j) + s - \log(x_i)$ taking positive values as follows:

$$\psi_V(x_i, x_j) = \exp\big(-\gamma \max(0, \log(x_j) - \log(x_i) + s)\big)\,, \tag{4}$$

where γ is a parameter that controls the strength of this prior.
Inter-Scanline Potential Given Property 5, ψ_H is used to encourage smooth transitions between scanlines. We achieve this by utilizing a Gaussian function:

$$\psi_H(x_i, x_j) = \exp\big(-\sigma(x_i - x_j)^2\big)\,, \tag{5}$$

where the parameter σ influences the strength of this prior. We empirically determined the parameter values $\gamma = 75$, $s = \frac{3}{2560}$, $\sigma = \frac{1}{4}$.

3.3 Convolutional Neural Network Training

We train a Convolutional Neural Network (CNN), which transforms an image I into a confidence map, by maximizing the likelihood of its prediction according to our probabilistic graphical model (2). Concretely, given a dataset of sparsely annotated ultrasound images $\big\{(I^{(i)}, y^{(i)})\big\}$, we optimize the parameters θ of the CNN $f(I, \theta)$ to minimize the negative log-likelihood of each prediction:

$$\theta^* = \arg\min_\theta \sum_i -\log p\big(f(I^{(i)}, \theta), y^{(i)}\big)\,. \tag{6}$$

To ensure parallel, vertically aligned scanlines even for B-mode images with non-linear fan geometry, e.g. when scanning with convex probes, we apply inverse scan conversion as a preprocessing step (Figure 1). This paper focuses on an alternative method for computing confidence maps rather than CNN architecture optimization. We train a small U-Net [17] with ConvNeXt blocks [14] on a dataset with 291 frames for training and 72 frames for validation. Our model achieved a validation loss of 0.32, closely matching the training loss of 0.25. For details, refer to our publicly available code[3]. On an NVIDIA RTX 4090, our model exceeds 2,300 fps making it suitable for real-time applications.

[3] github.com/ImFusionGmbH/BeyondShadows

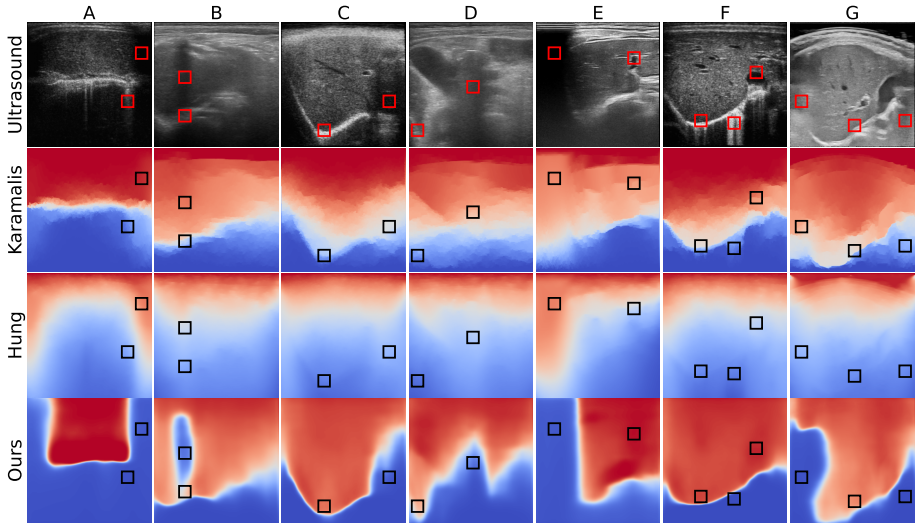


Fig. 4: Ultrasound frames (before scan conv.), with confidence maps generated by three methods. Red ■ and blue ■ represent high and low confidence, respectively. The squares on the confidence maps show regions of interest. *See text for details.*

4 Evaluation

We compare the proposed approach with the methods of Karamalis et al. [12] and Hung et al. [11]. Karamalis' method models pixels as graph nodes with edge weights derived from ultrasound physics, and computes confidence by solving a random walk equilibrium problem with fixed top (1) and bottom (0) boundary conditions. We use the public Python implementation available in MONAI[4], setting $\alpha = 1$. Hung's method reduces speckle in the image using an anisotropic filter and then employs a directed acyclic graph to propagate confidence downward from the top row. We use the official implementation[5] but set the parameters $\alpha = 10^{-2}$ and $\xi = 0.4$ to avoid rapid confidence decay. We exclude from the comparison the approach of [15] since we lack shadow specific annotations.

4.1 Qualitative Evaluation

Figure 4 depicts 7 representative frames (A-G), selected from the validation set. While Frames A-F were acquired with the same setups as our training data, Frame G comes from a completely different setup, demonstrating the generalization capabilities of our approach. Specifically, Frame G was acquired with a different ultrasound machine and through a water bath for optimized acoustic coupling, which causes the unusual artifacts visible on the left of the frame. Our

[4] github.com/Project-MONAI/MONAI
[5] github.com/aL3x-O-o-Hung/ultrasound-confidence-map-with-directed-graphs

Table 1: Random forest shadow segmentation using confidence maps. All rows except the last one are reprinted from [23], *see text for details.*

Method	**Dice** ↑	**Hausdorff** ↓	**Precision** ↑
Karamalis	49.3%	7.9	61.1%
Hung	47.6%	7.2	63.4%
Yesilkaynak	50.4%	**5.6**	71.5%
Ours	**58.9%**	6.2	**86.8%**

method excels in handling shadows, which are poorly managed by Karamalis' and Hung's methods (A-E). Of particular interest is Frame B's partial shadow, which is followed by a strong reflector (diaphragm). While other methods completely miss the shadow, our method detects it correctly and still assigns an intermediate confidence value to the diaphragm. Frame A and E's strong shadows, caused by missing probe contact, are completely mistaken by both competing methods. Hung's method manages to deal with reverberation (A, F) better than Karamalis', still our method provides the cleanest separation between visible structures and artifacts. Visible structures at higher depths are mistakenly assigned low confidence by Karamalis and Hung (C, D, G). Finally, the proposed method correctly recognizes the unusual skin appearance caused by the water bath (G), even though such images were not part of the training data.

4.2 Quantitative Evaluation: Bone Shadow Segmentation

For a first evaluation of a downstream task, we build upon the work of Yesilkaynak et al. [23], who evaluated their confidence map estimation using a random forest for bone shadow segmentation. While we could not include their Ultra-NeRF-based approach [22] in Sec. 4.1 due to the requirement of perfectly aligned ultrasound and CT volumes for the training phase, we utilized their publicly available code[6] and dataset, which consists of ultrasound frames and bone shadow masks. First, we ran our proposed confidence estimation on all frames, and then evaluated whether a random forest classifier could predict the shadows without modification of the provided code to ensure a fair comparison, i.e. leaving the bone shadow segmentation untouched. We reproduce the author's results in Table 1 verbatim and added a row with our own method's performance, showcasing that without any task-specific objective in training and without any finetuning, our method outperforms the state-of-art in terms of both Dice score and precision.

4.3 Quantitative Evaluation: Registration Weighting

As a second downstream task, we validate the efficacy of our approach by utilizing confidence as voxel weight for multi-modal intensity-based registration,

[6] github.com/MrGranddy/Redefining-Confidence-Maps

Table 2: Impact of using confidence as voxel weight for registration. A case is considered "converged" if the Fiducial Registration Error after registration is below 15 mm. The best results and the ones not significantly different ($p > 10^{-3}$) are highlighted in bold.

	Converged cases w.r.t. initial error		
Weight	**<25mm**	**25-50mm**	**>50mm**
Variance	69.7%	61.5%	47.9%
Karamalis	31.1%	25.5%	20.0%
Karamalis × Variance	70.4%	61.4%	48.4%
Hung	49.4%	34.7%	25.6%
Hung × Variance	70.7%	63.2%	48.4%
Ours	**77.9%**	**68.2%**	**53.5%**
Ours × Variance	**78.7%**	**71.1%**	**57.5%**

thereby replacing the conventional use of patch variance as the weighting factor [20]. To facilitate a direct comparison, we repeat the evaluation approach used in [16], which involves performing ultrasound fusion on a challenging dataset and reporting the number of converged cases across various algorithm variants. Our dataset includes 28 tracked liver clips from two different machines, with positional information obtained through optical tracking. Each clip is paired with a CT or MR volume, with at least 4 corresponding landmark pairs manually annotated by an expert. For each clip, we compute the confidence of all frames individually, and reconstruct a 3D confidence volume. We experiment with using confidence directly and multiplying it by the local patch variance. As can be observed in Table 2, our method was able to support the registration to converge in a significantly higher number of cases than established approaches.

5 Conclusions

We introduced a novel user-centered approach for generating confidence maps in ultrasound imaging, overcoming limitations of traditional methods with simplified assumptions of ultrasound physics. Our CNN training incorporates a probabilistic graphical model, combining sparse annotations with a physics-inspired prior. A qualitative evaluation demonstrated robustness and accuracy in the presence of a broad variety of typical ultrasound artifacts using diverse transducer geometries, different ultrasound systems, and acquisition environments. Quantitative tests, using confidence maps for shadow detection and multi-modal registration, highlighted our method's potential for downstream tasks. Notably, our method improved state-of-the-art bone shadow segmentation and ultrasound fusion algorithms. Its speed and reliability make it suitable for clinical applications where real-time evaluation is needed. Future work includes integration into clinical workflows and extending the approach to 3D ultrasound for volumetric analysis.

Disclosure of Interests. The authors have no competing interests.

References

1. Akbari, M., Carriere, J., Sloboda, R., Meyer, T., Usmani, N., Husain, S., Tavakoli, M.: Robot-assisted breast ultrasound scanning using geometrical analysis of the seroma and image segmentation. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3784–3791 (2021). https://doi.org/10.1109/IROS51168.2021.9636401
2. Schulte zu Berge, C., Kapoor, A., Navab, N.: Orientation-driven ultrasound compounding using uncertainty information. In: Stoyanov, D., Collins, D., Sakuma, I., Abolmaesumi, P., Jannin, P. (eds.) IPCAI 2014, LNCS, vol. 8498, pp. 236–245. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07521-1_25
3. Cai, H., Qi, L., Yu, Q., Shi, Y., Gao, Y.: 3d medical image segmentation with sparse annotation via cross-teaching between 3d and 2d networks. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023: 26th International Conference, Vancouver, BC, Canada, October 8–12, 2023, Proceedings, Part III. p. 614–624. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-43898-1_59, https://doi.org/10.1007/978-3-031-43898-1_59
4. Chatelain, P., Krupa, A., Navab, N.: Confidence-driven control of an ultrasound probe. IEEE Trans. Rob. **33**(6), 1410–1424 (2017)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. pp. 424–432. Springer International Publishing, Cham (2016)
6. Cobbold, R.: Foundations of Biomedical Ultrasound. Oxford University Press (2007)
7. Duque, V., Zirus, L., Velikova, Y., Navab, N., Mateus, D.: Can ultrasound confidence maps predict sonographers' labeling variability? In: Kainz, B., Noble, A., Schnabel, J., Khanal, B., Müller, J., Day, T. (eds.) ASMUS 2023, LNCS, vol. 14337, pp. 175–184. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-44521-7_17
8. Hedrick, W., Hykes, D., Starchman, D., Wilson, T.: Ultrasound Physics and Instrumentation. Mosby, St. Louis, MI (1995)
9. Hellier, P., Coupé, P., Meyer, P., Morandi, X., Collins, D.: Acoustic shadows detection, application to accurate reconstruction of 3d intraoperative ultrasound. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. pp. 1569–1572. IEEE (2008)
10. Hu, Y., Tavakoli, M.: Autonomous ultrasound scanning towards standard plane using interval interaction probabilistic movement primitives. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3719–3727 (2023). https://doi.org/10.1109/IROS55552.2023.10341685
11. Hung, A., Chen, W., Galeotti, J.: Ultrasound confidence maps of intensity and structure based on directed acyclic graphs and artifact models. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 697–701. IEEE (2021)
12. Karamalis, A., Wein, W., Klein, T., Navab, N.: Ultrasound confidence maps using random walks. Med. Image Anal. **16**(6), 1101–1112 (2012)
13. Klein, T., Wells, W.: Rf ultrasound distribution-based confidence maps. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) MICCAI 2015, Part II 18, LNCS, vol. 9350, pp. 595–602. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24571-3_71

14. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
15. Meng, Q., et al.: Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. IEEE Trans. Med. Imaging (2019)
16. Ronchetti, M., Wein, W., Navab, N., Zettinig, O., Prevost, R.: Disa: Differentiable similarity approximation for universal multimodal registration. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 761–770. Springer Nature Switzerland, Cham (2023)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
18. Virga, S., et al.: Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 508–513. IEEE (2016)
19. Wein, W., Karamalis, A., Baumgartner, A., Navab, N.: Automatic bone detection and soft tissue aware ultrasound-ct registration for computer-aided orthopedic surgery. Int. J. Comput. Assist. Radiol. Surg. **10**, 971–979 (2015)
20. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical image analysis **12**(5), 577–585 (2008)
21. Welleweerd, M.K., de Groot, A.G., De Looijer, S., Siepel, F.J., Stramigioli, S.: Automated robotic breast ultrasound acquisition using ultrasound feedback. In: 2020 IEEE international conference on robotics and automation (ICRA). pp. 9946–9952. IEEE (2020)
22. Wysocki, M., Azampour, M., Eilers, C., Busam, B., Salehi, M., Navab, N.: Ultranerf: neural radiance fields for ultrasound imaging. In: Medical Imaging with Deep Learning, pp. 382–401. PMLR (2024)
23. Yesilkaynak, V.B., Duque, V.G., Wysocki, M., Velikova, Y., Mateus, D., Navab, N.: Ultrasound confidence maps with neural implicit representation. In: Medical Image Understanding and Analysis: 28th Annual Conference, MIUA 2024, Manchester, UK, July 24–26, 2024, Proceedings, Part II. p. 89–100. Springer-Verlag, Berlin, Heidelberg (2024). https://doi.org/10.1007/978-3-031-66958-3_7