Nonrigid Reconstruction of Freehand Ultrasound without a Tracker

Qi Li[1(✉)], Ziyi Shen[1], Qianye Yang[1], Dean C. Barratt[1], Matthew J. Clarkson[1], Tom Vercauteren[3], and Yipeng Hu[1]

[1] Centre for Medical Image Computing, Wellcome/EPSRC Centre for Interventional and Surgical Sciences, Department of Medical Physics and Biomedical Engineering, University College London, London, U.K.
qi.li.21@ucl.ac.uk

[2] School of Biomedical Engineering & Imaging Sciences, King's College London, London, U.K.

Abstract. Reconstructing 2D freehand Ultrasound (US) frames into 3D space without using a tracker has recently seen advances with deep learning. Predicting good frame-to-frame rigid transformations is often accepted as the learning objective, especially when the ground-truth labels from spatial tracking devices are inherently rigid transformations. Motivated by a) the observed nonrigid deformation due to soft tissue motion during scanning, and b) the highly sensitive prediction of rigid transformation, this study investigates the methods and their benefits in predicting nonrigid transformations for reconstructing 3D US. We propose a novel co-optimisation algorithm for simultaneously estimating rigid transformations among US frames, supervised by ground-truth from a tracker, and a nonrigid deformation, optimised by a regularised registration network. We show that these two objectives can be either optimised using meta-learning or combined by weighting. A fast scattered data interpolation is also developed for enabling frequent reconstruction and registration of non-parallel US frames, during training. With a new data set containing over 357,000 frames in 720 scans, acquired from 60 subjects, the experiments demonstrate that, due to an expanded thus easier-to-optimise solution space, the generalisation is improved with the added deformation estimation, with respect to the rigid ground-truth. The global pixel reconstruction error (assessing accumulative prediction) is lowered from 18.48 to 16.51 mm, compared with baseline rigid-transformation-predicting methods. Using manually identified landmarks, the proposed co-optimisation also shows potentials in compensating nonrigid tissue motion at inference, which is not measurable by tracker-provided ground-truth. The code and data used in this paper are made publicly available at `https://github.com/QiLi111/NR-Rec-FUS`.

Keywords: Freehand US · Reconstruction · Registration · Deformation.

1 Introduction

With a variety of clinical applications including measurement assessment [13], pre-operative registration [11] and surgical guidance [19], trackless freehand

US reconstruction has been proposed using both non-learning [4] and machine learning-based approaches [29]. Recent learning-based methods vary in their network architectures [25,33,28], training strategies [21,6,24], the use of prior knowledge [22,23], input frames [26,33] and sequential modelling techniques [16,27].

To our knowledge, most existing approaches optimise rigid transformations among US frames, which characterises the spatial movement - rotation and translation of the ultrasound probe during scanning. However, there is evidence that probe pressure and patient movement cause soft-tissue undergoing nonrigid deformation [32]. Previous work compensated this nonrigid deformation using a separate registration algorithm, after a rigid reconstruction [29], aligning the same anatomical structures appeared in repeated scans [32].

However, validating such a nonrigid deformation modelling is challenging due to a lack of general means to obtain soft-tissue-tracking, motion-included ground-truth. In fact, the ground-truth data for supervision used in many learning-based approaches are obtained from spatial tracking devices, such as optical and electromagnetic trackers. They are rigid transformations that localise the rigid probe rather than the deformable anatomical structures.

Withstanding the challenges in validating the estimated nonrigid deformation, we would like to explore other benefits for predicting nonrigid transformation in reconstructing US images. Two types of other applications that require predicting rigid transformation are image registration and image data augmentation/perturbation in training neural networks. Both have reported the difficulties in esimating this constrained transformation. For example, predicting rigid transformation is highly sensitive to initialisation and learning rate in weakly-supervised registration [8] and spatial transformer network training [10], mandating careful hyperparameter tuning in these applications. Predicting a higher degree-of-freedom, flexible nonrigid transformation provides an expanded solution space in optimising these transformation estimation methods, with respect to either rigid or nonrigid ground-truth labels. This should improve the resulting model generalisation, given limited data and compute resources in practice.

In this paper, we propose a co-optimisation deep-learning-based approach, together with a "conventional" learnable rigid reconstruction, to estimate an additional nonrigid deformation between US frames as well as within acquired individual frames. Although the latter may not be plausibly compensate physical deformation due to the fast single-frame imaging process, this should allow the flexibility that benefits the numerical training process discussed above.

In summary, our contributions include: 1) a novel co-optimisation approach, not only for compensating nonrigid soft tissue motion but also improve the network training for better generalisation based on rigid ground-truth; 2) an open-sourced PyTorch implementation of a practical interpolation method for scatter ultrasound intensity values; 3) a new set of evaluation metrics for reconstruction evaluation, at both global and local levels; and 4) one of the largest *in vivo* US dataset for freehand US reconstruction, with recorded tracker information.

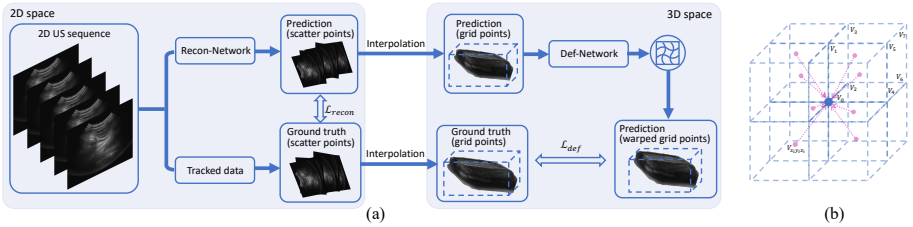

Fig. 1. (a) Overview of the proposed method with rigid transformation and nonrigid deformation prediction. (b) Contributions from support data to query data.

2 Method

An US sequence S consists of a number of US frames $S = \{I_m\}, m = 1, 2, ..., M$, acquired with increasing timestamps. For any pair of US frames I_i and I_j, a spatial rigid transformation parameter vector $\mathbf{t}^{j \leftarrow i}, 1 \leq i < j \leq M$ denotes the relative translation and rotation between the i^{th} and j^{th} frames. An US scan containing several US sequences can be reconstructed in 3D once all the transformations between each US frame and the reference frame can be calculated, where the reference frame can be any frame in the scan. Fig. 1 (a) provides an overview of the proposed method.

2.1 Rigid Reconstruction of Stacked 2D Frames

This section describes a rigid transformation estimation pipeline as proposed in [29], using a deep neural network. As described in the sequence-modelling methods [16], a convolutional neural network f_{recon} with parameters θ_{recon} predicts rigid transformations with respect to the reference frame: $[(\hat{\mathbf{t}}_1^{ref \leftarrow 1})^\top, ..., (\hat{\mathbf{t}}_M^{ref \leftarrow M})^\top] = f_{recon}(S; \theta_{recon})$, where $\{\hat{\mathbf{t}}_m^{ref \leftarrow m}\}_{m=1}^M$, is a set of rigid transformation parameter vectors from all frames to the reference frame. The 3D coordinates of n^{th} pixel in m^{th} frame $\hat{P}_m^n$, in reference frame coordinate system, can be calculated using $\hat{P}_m^n = \mathcal{T}(P_{I_0}^n, \hat{\mathbf{t}}_m^{ref \leftarrow m})$. $P_{I_0}^n$ is the point coordinates in it's tracker tool coordinate system, and can be obtained by using the spatial calibration matrix [7] $\mathbf{t}_{calib}$: $P_{I_0}^n = \mathcal{T}(P_{img}^n, \mathbf{t}_{calib})$, where P_{img}^n denotes the coordinates in the pixel coordinate system [34]. $\mathcal{T}$ is a function applies the $\hat{\mathbf{t}}_m^{ref \leftarrow m}$-converted transformation on the points. While any frame in an US sequence can be regarded as the reference frame, we use the first frame as the reference frame.

2.2 Interpolation for Efficient Resampling Non-parallel Frames

In this section, we first describe an efficient approach for interpolating from scatter data to grid samples. US intensities sampled at regular grid are useful

to efficiently compute similarity measures, deformation regularisation and image warping, as commonly adopted in training registration networks. These are repeatedly applied in the proposed co-optimisation, described in Sec. 2.3. Many interpolation methods for estimating grid data from scatter data [9,1,14,3,31,5] are computationally expensive. We introduce a simple and efficient interpolation, with a time complexity of $O(N)$ where N is number of pixels in a scan.

Denote intensity value at any point (x, y, z) as $V_{x,y,z}$, where the coordinate system is defined such that the reconstructed voxel grids are positioned at integer coordinates. Suppose there are N support data points at $\{(x_i, y_i, z_i)\}$, the volume reconstruction process is then given by

$$V_{n_1,n_2,n_3} = \frac{\Sigma_{i=1}^{N} W(x_i - n_1)W(y_i - n_2)W(z_i - n_3)V_{x_i,y_i,z_i}}{\Sigma_{i=1}^{N} W(x_i - n_1)W(y_i - n_2)W(z_i - n_3)} \tag{1}$$

where $(n_1, n_2, n_3) \in \mathbb{N}^3$ is voxel index, and the weight function is given by

$$W(u) = \mathbb{1}_{|u|\leq 1}(u) \cdot (1 - |u|) \tag{2}$$

The indicator function in Eq. 2 suggests that any support data point within one certain cube will only contribute to the eight vertices, or reversely, the value at any query grid point can be calculated based on contributions of support data within adjacent eight cubes. The computation of these contributions thus can be done independently between all these cubes, before gathering all these contributions for estimating the query vertex values, thus $O(N)$[3].

2.3 Deformation Estimation using Co-Optimisation

A transformation-predicting network [2] f_{def} with parameters θ_{def} takes the interpolated rigid-transformed $\hat{V}$ as input and generate a dense displacement field (DDF) ϕ: $\phi = f_{def}(\hat{V}; \theta_{def})$. The DDF then warps the rigid-reconstruction-predicted volume to obtain the final prediction: $\widetilde{V} = \hat{V} \circ \phi$. As illustrated in Fig. 1 (a), the deformation estimation process is co-optimised together with rigid reconstruction in Section 2.1.

The mean squared error, between predicted points coordinates and ground-truth points coordinates, is used to supervise the rigid reconstruction: $\mathcal{L}_{recon} = \frac{1}{M} \times \frac{1}{N} \times \sum_{m=1}^{M} \sum_{n=1}^{N} ||P_m^n - \hat{P}_m^n||_2^2$, where P_m^n and $\hat{P}_m^n$ are the points coordinates of ground-truth and prediction respectively, transformed from ground-truth transformation $\mathbf{t}^{ref \leftarrow m}$ and predicted transformation $\hat{\mathbf{t}}^{ref \leftarrow m}$. $\mathbf{t}^{ref \leftarrow m}$ is calculated using two tool-to-world transformations, $T^{ref \leftarrow m} = (T^{world \leftarrow ref})^{-1} \cdot T^{world \leftarrow m}$, T denoting transformation matrices, converted from parameter vectors. The tool-to-world transformation is obtained from an optical tracker.

The loss function for training the deformation network is a typical registration loss [8], consists of bending energy and intensity similarity between the ground-truth-reconstructed volume V and wrapped predicted volume $\widetilde{V}$, to not

[3] The interpolation process has an average speed of less than 1 ms over the dataset.

only encourage a smooth deformation generated by the network, but also provides an opportunity to rectify any erroneous rigid reconstruction estimation: $\mathcal{L}_{def}(\widetilde{V}, V, \phi) = \mathcal{L}_{sim}(\widetilde{V}, V) + \mathcal{L}_{smooth}(\phi)$

2.4 Training Strategy

In this section, we describe two training strategies that can be used in the proposed pipeline, meta-learning and end-to-end training.

In the meta-learning strategy, network parameters involved in two processes described in Sections 2.1 and 2.3 are separately optimised using a training data set $\mathcal{D}_{train} = \{S, \{\mathbf{t}_m^{ref \leftarrow m}\}_{m=1}^{M}\}$ and a validation data set $\mathcal{D}_{val} = \{\widetilde{V}, V\}$, respectively: $\hat{\theta}_{def} = \arg\min_{\theta_{def}} \mathcal{L}_{def}^{val}(\theta_{def}; \hat{\theta}_{recon}, \mathcal{D}_{val})$, s.t. $\hat{\theta}_{recon} = \arg\min_{\theta_{recon}} \mathcal{L}_{recon}^{train}(\theta_{recon}; \mathcal{D}_{train})$. This bi-level optimisation strategy updates the reconstruction and deformation networks, on separate training and validation sets. This has widely been adopted in previous work, including those for this application [15], to take into account the co-dependency between the optimisation of the two networks and avoid sub-optimum solutions to both optimisations.

However, we have found that such trivial solutions are unlikely in this application, perhaps due to the highly constrained deformation estimation. Therefore, we propose to use a simple weighted single loss function to co-optimise the two networks in an end-to-end training. The loss functions in Sections 2.1 and 2.3 are combined by weighting to train both network parameters $\hat{\theta}_{recon}$ and $\hat{\theta}_{def}$: $\hat{\theta}_{recon}, \hat{\theta}_{def} = \arg\min_{\theta_{recon}, \theta_{def}} \mathcal{L}_{ete}^{train}(\theta_{recon}, \theta_{def}; \mathcal{D}_{train})$, where the end-to-end loss $\mathcal{L}_{ete}^{train}$ used for supervising the co-optimisation process consists of two, with a weight α calculated based on the magnitude of gradient of two parts [18]: $\mathcal{L}_{ete}^{train} = \mathcal{L}_{recon}^{train} + \alpha \times \mathcal{L}_{def}^{train}$.

2.5 Evaluation Metrics

Using rigid transformation, e.g. recorded by a spatial tracker, as ground-truth, the weighting between translation and rotation components may be difficult to interpret. In this paper, we design and propose four streamlined evaluation metrics, on pixel and landmark reconstruction error, at local and global levels.

We first define two types of transformation-representing displacement vectors - global displacement vectors and local displacement vectors, where the former represents the displacement between each frame and the reference frame (i.e., the first frame in this work) and the latter denotes the displacement between each frame and the immediately previous frame.

We then define two types of errors, consisting of 1) pixel reconstruction error, where the reconstruction error is the averaged Euclidean distance between ground-truth- and predicted- reconstructed points locations, averaged over all pixels of all but the reference frame in a scan; and 2) landmark reconstruction error, where the reconstruction error is averaged over landmarks in a scan, with the same measurement as pixel reconstruction error.

Thus, the four evaluation metrics used in this study are based on the two types of displacement vectors and the two types of error measurements: 1) global

pixel reconstruction error (GPE), reconstruction error on all pixels based on global displacement vectors; 2) global landmark reconstruction error (GLE), reconstruction error on landmarks based on global displacement vectors; 3) local pixel reconstruction error (LPE), reconstruction error on all pixels based on local displacement vectors; and 4) local landmark reconstruction error (LLE), reconstruction error on landmarks based on local displacement vectors.

The proposed global and local levels of displacement vectors are capable of reflecting the reconstruction error on both frame-level and accumulated error of the algorithm [16]. In addition to the scenario where the performance on the entire scan or adjacent frames is required, as measured by the above metrics, other clinical applications may reconstruct a sequence of US frames using different application-dependent intervals. Nonetheless, these four metrics should still provide an estimate of performance range, for these applications with varying trade-off between accumulated error and reference updating.

3 Experiments

Data Acquisition: The *in vivo* data [4] used in this paper were acquired from 60 volunteers, using Ultrasonix machine (BK, Europe) with a curvilinear probe (4DC7-3/40), tracked by an NDI Polaris Vicra (Northern Digital Inc., Canada). Other US imaging parameters are empirically configured based on the visual quality of acquired US images. For example, the frequency was set at 6MHz with a depth of 9 cm, and the dynamic range is 83 dB with an overall gain of 48%. The US frame, with an image size of 640×480, was recorded at 20 frames per second (fps), without speckle reduction. The spatial calibration was obtained using a pinhead-based method [7], and the temporal difference between the optical tracker and imaging was calibrated using the Plus Toolkit [12].

Twelve scans were acquired for each subject, from both left and right arms, with the US probe perpendicular of and parallel to the scanning direction, in three different scanning trajectories - straight, c-shape and s-shape, in a distal-to-proximal direction, resulting in 720 scans in total. The average number of frames per scan is 500, equivalent to $200 - 300$ mm. The data set was split into train, validation and test sets by a ratio of 3:1:1 on subject level, where scans from the same subject cannot be in different sets.

Network Development and Implementation: EfficientNet (b1) [30] is adopted as the backbone of the reconstruction network. A fully connected layer is added at the end to output $(M-1) \times 6$ rigid transformation parameters, with US sequence containing M frames as input. After calculating the points locations in real-world space for pixels in the input US sequence and interpolating into an US volume, the rigid-transformed volume is fed into an adapted VoxelMorph [2] network, with an input channel of 1.

[4] This study was performed in accordance with the ethical standards in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Approval was granted by the Ethics Committee of local institution (UCL Department of Medical Physics and Biomedical Engineering) on 20^{th} Jan. 2023 [24055/001].

The input sequence length is set to 100 for reported benchmark performance [17]. The reconstructed volume has a resolution of 1 mm×1 mm×1 mm with various sizes. Other hyper-parameters are less sensitive to reconstruction and selected based on validation performance, including an Adam optimizer with a learning rate of 10^{-4} and $\alpha = 10^3$, on a single NVIDIA Quadro GV100 GPU.

For a fair comparison, we adapted two start-of-the-art (SOTA) methods, [29] and [16], with the same hyper-parameters. We also compared with a model with meta-training strategy, adapted from [20]. The same data split was used with a ratio of 2:2:1, where the train and validation sets for rigid transformation and nonrigid deformation network training, respectively. All models were trained for at least 10,000 epochs until convergence, for up to 5 days.

4 Results

The performance improvement from our proposed method is summarised in Table 1, as well as results from an ablation study with only rigid transformation prediction (baseline) or meta-trained (*Meta*) models. As an surrogate of clinically useful landmarks, four corner points in an image were used. *Recon* and *Def* models come from our proposed approach, using the co-optimised rigid transformation and nonrigid transformation, respectively.

The reconstruction performance of the *Recon* and *Def* models both show performance improvement in global reconstruction metrics (GPE and GLE), compared with the baseline method (p-value = 0.001 and 0.003 for *Def* model, based on paired t-test at a significance level at $\alpha = 0.05$). The performance of *Meta* model is poorer than those from both the end-to-end model and the baseline, for metrics, demonstrating the effectiveness of proposed end-to-end strategy. Most interestingly, although the global reconstruction metrics show

Table 1. Reconstruction performance using proposed four evaluation metrics, among baseline, SOTA and our methods.

Models	GPE (mm)	GLE (mm)	LPE (mm)	LLE(mm)
Baseline	18.48 ± 10.30	19.70 ± 10.42	**0.41±0.17**	**0.44±0.18**
$\text{Recon}_{\text{meta}}$	21.29 ± 11.01	22.82 ± 11.15	3.59 ± 1.44	3.71 ± 1.45
Def_{meta}	20.89 ± 10.52	22.80 ± 11.13	–	–
Recon (ours)	16.69 ± 7.79	18.15 ± 7.91	3.07 ± 0.99	3.38 ± 1.00
Def (ours)	**16.51±7.76**	**17.91±7.85**	–	–
$[29]_{\text{cf}}$	18.33 ± 7.77	20.19 ± 8.04	**0.23±0.07**	**0.25±0.08**
$[16]_{\text{cf}}$	17.32 ± 8.12	18.64 ± 8.53	0.23 ± 0.08	0.25 ± 0.09
$\text{Baseline}_{\text{cf}}$	16.44 ± 7.83	17.75 ± 8.08	0.39 ± 0.16	0.44 ± 0.18
$\text{Recon}_{\text{meta\_cf}}$	18.46 ± 8.77	19.93 ± 8.95	3.29 ± 1.37	3.57 ± 1.48
$\text{Def}_{\text{meta\_cf}}$	18.21 ± 8.37	19.94 ± 8.90	–	–
Recon_{cf} (ours)	15.26 ± 7.17	16.57 ± 7.43	2.71 ± 0.89	3.25 ± 0.97
Def_{cf} (ours)	**15.13±7.12**	**16.37±7.35**	–	–

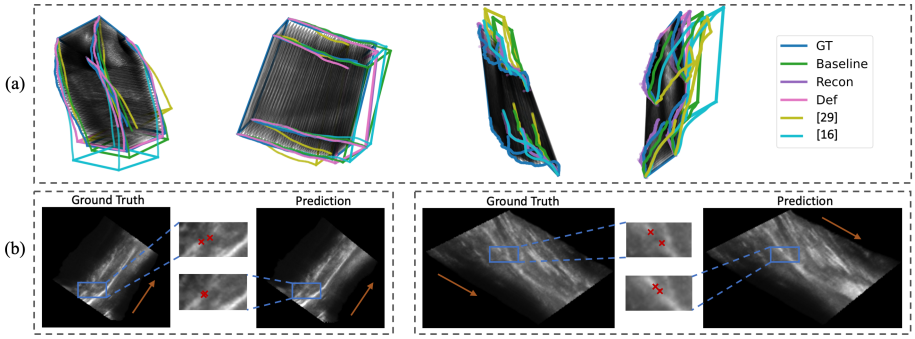

Fig. 2. (a): Reconstructed US scan volumes with ground-truth, proposed and SOTA methods, illustrated with perpendicular c-shape, perpendicular straight, parallel s-shape and parallel c-shape scans, from left to right. (b) Illustration of landmarks in ground-truth- and prediction- reconstructed US scan volumes.

significant improvement, the local metrics has a reduced performance. This may reflect a property of the added regularised deformation estimation, in which the local perturbation (albeit may yield larger variance) is constrained to have a smaller bias, thus reducing the long-term (or long distance in this case) expected error.

The number of frames in an input US sequence is set at 100 in [16], with transformation from 20^{th} to 21^{th} frames as an example. As the method in [16] cannot predict the probe trajectory for all frames in a scan. For fair comparison, we subsample the frames and use the same reference frame for the other methods in Table 1, so that all methods predict the transformation for the same subset of frames, i.e. the "common frames", denoted as $_{\text{cf}}$. When comparing on common frames between proposed and baseline models, the improvement was also observed with p-value $= 0.004$ and 0.003 in GPE and GLE, using *Def* model, also illustrated in Fig. 2 (a).

The proposed co-optimisation approach has potentials to rectify and improve rigid reconstruction with ground-truth, by compensating nonrigid deformation. Fig. 2 (b) shows two example slices on the same location, from ground-truth-reconstructed volumes and DDF-predicted volumes, arrow indicating the scanning direction. The landmarks (red crosses) in each slice represent the same anatomical structure, and thus should be at the same 3D location when reconstructed. It can be seen that the anatomical structure is broken in the ground-truth-reconstructed volume, perhaps due to movement or tissue motion. The distance between these landmarks decreased in the predicted volume, from an average distance of 4.65 mm to 1.48 mm.

5 Conclusion and Discussion

This work introduced deformation estimation into rigid reconstruction of freehand US. The experimental results, evaluated on a large data set, shows the efficacy of the proposed algorithm. Examples of compensating nonrigid deformation are also discussed, to open up new avenue for improving this longstanding challenge in ultrasound image computing.

Acknowledgments. This work was supported by the EPSRC [EP/T029404/1], a Royal Academy of Engineering / Medtronic Research Chair [RCSRF1819\7\734] (TV), Wellcome/EPSRC Centre for Interventional and Surgical Sciences [203145Z/16/Z], and the International Alliance for Cancer Early Detection, an alliance between Cancer Research UK [C28070/A30912; C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. TV is co-founder and shareholder of Hypervision Surgical. Qi Li was supported by the University College London Overseas and Graduate Research Scholarships. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

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

References

1. Amidror, I.: Scattered data interpolation methods for electronic imaging systems: a survey. Journal of electronic imaging **11**(2), 157–176 (2002)
2. Balakrishnan, G., Zhao, A., et al.: Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging **38**(8), 1788–1800 (2019)
3. Bartier, P.M., Keller, C.P.: Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (idw). Computers & Geosciences **22**(7), 795–799 (1996)
4. Chen, J.F., Fowlkes, J.B., et al.: Determination of scan-plane motion using speckle decorrelation: Theoretical considerations and initial test. International Journal of Imaging Systems and Technology **8**(1), 38–44 (1997)
5. Ebner, M., Chouhan, M., et al.: Point-spread-function-aware slice-to-volume registration: application to upper abdominal mri super-resolution. In: Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016. pp. 3–13. Springer (2017)
6. Guo, H., Chao, H., et al.: Ultrasound volume reconstruction from freehand scans without tracking. IEEE Transactions on Biomedical Engineering **70**(3), 970–979 (2022)
7. Hu, Y., Gibson, E., et al.: Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks. In: Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment, pp. 105–115. Springer (2017)
8. Hu, Y., Modat, M., et al.: Weakly-supervised convolutional neural networks for multimodal image registration. Medical image analysis **49**, 1–13 (2018)

9. Huang, H., Cui, C., et al.: Grid interpolation algorithm based on nearest neighbor fast search. Earth Science Informatics **5**, 181–187 (2012)
10. Jaderberg, M., Simonyan, K., Zisserman, A., kavukcuoglu, k.: Spatial transformer networks. In: Advances in Neural Information Processing Systems. vol. 28. Curran Associates, Inc. (2015)
11. Lang, A., Mousavi, P., et al.: Multi-modal registration of speckle-tracked freehand 3d ultrasound to ct in the lumbar spine. Medical image analysis **16**(3), 675–686 (2012)
12. Lasso, A., Heffter, T., et al.: Plus: open-source toolkit for ultrasound-guided intervention systems. IEEE transactions on biomedical engineering **61**(10), 2527–2537 (2014)
13. Leblanc, T., Lalys, F., et al.: Stretched reconstruction based on 2d freehand ultrasound for peripheral artery imaging. International Journal of Computer Assisted Radiology and Surgery **17**(7), 1281–1288 (2022)
14. Lee, S., Wolberg, G., et al.: Scattered data interpolation with multilevel b-splines. IEEE transactions on visualization and computer graphics **3**(3), 228–244 (1997)
15. Li, Q., Shen, Z., et al.: Privileged anatomical and protocol discrimination in trackerless 3d ultrasound reconstruction. In: International Workshop on Advances in Simplifying Medical Ultrasound. pp. 142–151. Springer (2023)
16. Li, Q., Shen, Z., et al.: Trackerless freehand ultrasound with sequence modelling and auxiliary transformation over past and future frames. In: International Symposium on Biomedical Imaging. pp. 1–5. IEEE (2023)
17. Li, Q., Shen, Z., et al.: Long-term dependency for 3d reconstruction of freehand ultrasound without external tracker. IEEE Transactions on Biomedical Engineering **71**(3), 1033–1042 (2024)
18. Lin, B., Feiyang, Y., Zhang, Y.: A closer look at loss weighting in multi-task learning (2021)
19. Lindseth, F., Kaspersen, J.H., et al.: Multimodal image fusion in ultrasound-based neuronavigation: improving overview and interpretation by integrating preoperative mri with intraoperative 3d ultrasound. Computer Aided Surgery **8**(2), 49–69 (2003)
20. Liu, H., Simonyan, K., et al.: Darts: Differentiable architecture search. In: International Conference on Learning Representations (2018)
21. Luo, M., Yang, X., et al.: Self context and shape prior for sensorless freehand 3d ultrasound reconstruction. In: Medical Image Computing and Computer Assisted Intervention. pp. 201–210. Springer (2021)
22. Luo, M., Yang, X., et al.: Deep motion network for freehand 3d ultrasound reconstruction. In: Medical Image Computing and Computer-Assisted Intervention. pp. 290–299. Springer (2022)
23. Luo, M., Yang, X., et al.: Multi-imu with online self-consistency for freehand 3d ultrasound reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 342–351. Springer (2023)
24. Luo, M., Yang, X., et al.: Recon: Online learning for sensorless freehand 3d ultrasound reconstruction. Medical Image Analysis **87**, 102810 (2023)
25. Mikaeili, M., Bilge, H.Ş.: Trajectory estimation of ultrasound images based on convolutional neural network. Biomedical Signal Processing and Control **78**, 103965 (2022)
26. Miura, K., Ito, K., et al.: Localizing 2d ultrasound probe from ultrasound image sequences using deep learning for volume reconstruction. In: International Workshop on Advances in Simplifying Medical Ultrasound. pp. 97–105. Springer (2020)

27. Miura, K., Ito, K., et al.: Probe localization from ultrasound image sequences using deep learning for volume reconstruction. In: International Forum on Medical Imaging in Asia. vol. 11792, pp. 133–138. SPIE (2021)
28. Ning, G., Liang, H., et al.: Spatial position estimation method for 3d ultrasound reconstruction based on hybrid transfomers. In: International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
29. Prevost, R., Salehi, M., et al.: 3d freehand ultrasound without external tracking using deep learning. Medical image analysis **48**, 187–202 (2018)
30. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
31. Vercauteren, T., Perchant, A., et al.: Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. Medical image analysis **10**(5), 673–692 (2006)
32. Wein, W., Lupetti, M., et al.: Three-dimensional thyroid assessment from untracked 2d ultrasound clips. In: Medical Image Computing and Computer-Assisted Intervention. pp. 514–523. Springer (2020)
33. Xie, Y., Liao, H., et al.: Image-based 3d ultrasound reconstruction with optical flow via pyramid warping network. In: IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3539–3542. IEEE (2021)
34. Xingfang, Y., Yumei, H., Feng, G.: A simple camera calibration method based on sub-pixel corner extraction of the chessboard image. In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems. vol. 3, pp. 688–692. IEEE (2010)