

Privileged Anatomical and Protocol Discrimination in Trackerless 3D Ultrasound Reconstruction

Qi Li¹, Ziyi Shen¹, Qian Li^{1,2}, Dean C. Barratt¹, Thomas Dowrick¹, Matthew J. Clarkson¹, Tom Vercauteren³, and Yipeng Hu¹

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.

² State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin, China

³ School of Biomedical Engineering & Imaging Sciences, King's College London, London, U.K.

Abstract. Three-dimensional (3D) freehand ultrasound (US) reconstruction without using any additional external tracking device has seen recent advances with deep neural networks (DNNs). In this paper, we first investigated two identified contributing factors of the learned inter-frame correlation that enable the DNN-based reconstruction: anatomy and protocol. We propose to incorporate the ability to represent these two factors - readily available during training - as the privileged information to improve existing DNN-based methods. This is implemented in a new multi-task method, where the anatomical and protocol discrimination are used as auxiliary tasks. We further develop a differentiable network architecture to optimise the branching location of these auxiliary tasks, which controls the ratio between shared and task-specific network parameters, for maximising the benefits from the two auxiliary tasks. Experimental results, on a dataset with 38 forearms of 19 volunteers acquired with 6 different scanning protocols, show that 1) both anatomical and protocol variances are enabling factors for DNN-based US reconstruction; 2) learning how to discriminate different subjects (anatomical variance) and predefined types of scanning paths (protocol variance) both significantly improve frame prediction accuracy, volume reconstruction overlap, accumulated tracking error and final drift, using the proposed algorithm.

Keywords: Freehand ultrasound · Privileged information · Multi-task learning.

1 Introduction

3D ultrasound (US) reconstruction is a promising technique both for diagnostics and image guidance, such as image fusion with other image modalities [5], registration with preoperative data during surgery [3], volume visualisation and measurement [2]. Currently, most clinically applied 3D reconstruction of 2D freehand

US imaging use spatial tracking devices, such as optical, electromagnetic or mechanical positioning [12]. Research for reducing dependency on external devices in such 3D US reconstruction has been motivated for its portability, accessibility and low-cost. Previous non-learning-based methods utilised the speckle correlation between US frames [1,3]. In recent learning-based approaches, convolutional neural networks and their variants have been proposed to use two adjacent frames as network input [10,15,16], for predicting spatial transformation between them. More generally, sequential modelling methods, e.g. using recurrent neural network [4,7,8,11] and transformer [14], have also been tested with the same goal of localising relative positions between two or more US frames.

With the promising results from deep learning, one might question what was learned in these data-driven approaches for predicting inter-frame transformations. Specifically, in addition to speckle patterns (which holds within a limited spatial scale), what are other factors that generated correlation between US frames, such that their relative locations can be inferred?

We first hypothesized that two factors, common anatomical characteristics between subjects and predefined scanning protocols, are responsible for such predictability in this application. A variance-reduction study is presented in Sec. 3.2 to demonstrate that sufficient anatomical and protocol variance in training data is indeed required for the reconstruction.

In this work, we propose to encode the anatomical and protocol patterns using two classification tasks, discriminating between subjects and between types of scanning paths, respectively. We then investigate methods to train these tasks together with the main reconstruction task to improve the performance of the main task.

In addition to US images as network input, previous studies have also integrated additional information or signals, such as optical flow [21] between US frames and acceleration, orientation, angular rates from inertial measurement units (IMU) [8,9]. Optical flow was derived from image sequence itself, while the IMU-measured signals are required at both training and inference. Different to these applications, the proposed discrimination task labels, subject and protocol indices, are in general available during training, but not required at inference, thus known as privileged information [19], further discussed in Sec. 2.1.

In summary, our contributions include 1) a multi-task learning approach to formulate two factors in freehand US as privileged information, for improving reconstruction accuracy; 2) a mixture model formulation for optimisable branching locations for auxiliary tasks, implemented with a differentiable network and a gradient-based bi-level optimisation; 3) extensive experimental results to quantify the improved performance due to the privileged tasks; and 4) open-source code and data ⁴ for public access and reproducibility.

⁴ <https://github.com/ucl-candi/freehand>

2 Method

2.1 Preliminaries: Privileged Auxiliary Tasks and Shared Parameters

Assume a main task $f_\theta(y|x)$ that predicts y with input data x (here, using a θ -parameterised neural network), which is optimised alongside \mathcal{J} auxiliary tasks $f_{\theta_j^s, \theta_j^a}(y_j^a|x)$, $j = 1, \dots, \mathcal{J}$, where θ_j^s are shared parameters (with the main task), and θ_j^a are task-specific parameters, both for predicting y_j^a . Therefore, $\theta_j^s \subseteq \theta$ and $\theta_j^a \cap \theta = \emptyset$. When the goal is to predict y from x , the supervision and prediction of auxiliary task are only required during training. The main task benefits from this privileged information, which is not required at inference.

Such multi-task learning incorporates the privileged information but may suffer from absolute negative transfer [20] - here, when the additional auxiliary tasks negatively impact the main task performance, and/or relative negative transfer - one auxiliary task reduces the main-task-improving potentials from the other auxiliary task(s). One approach for reducing negative transfer, or optimising the transferability, is adjusting the ratio between the shared θ_j^s and task-specific θ_j^a parameters [13]. As no parameters are shared (i.e. $\theta_j^s = \emptyset$), no negative transfer is possible (although any benefit from this auxiliary task also diminishes).

Assume a binary task descriptor z_j (extended to a one-hot vector in Sec. 2.3) indicating the use of the j^{th} auxiliary task. The main task is thus conditioned on the use of the auxiliary task $f_{\theta, \theta_j^a}(y|x, z_j)$. Importantly, where to place the task descriptor z_j determines which and how many network parameters θ_j^s are shared. As in Fig. 1 (a), the closer the task descriptor is to the input, i.e. early branching, the less shared parameters.

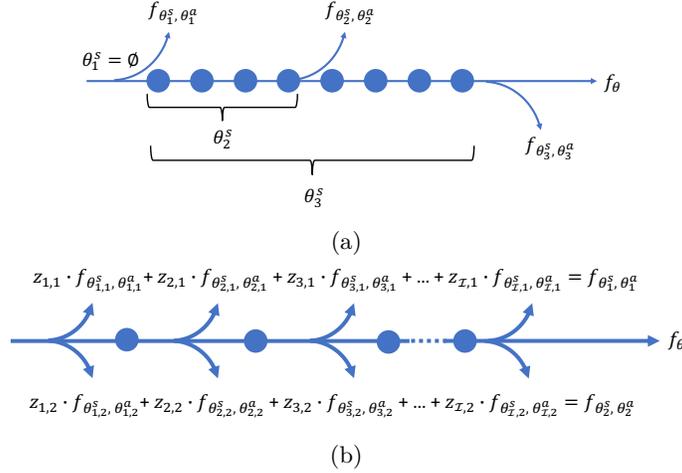


Fig. 1: Example of network architectures with (a) various possible branching locations with different shared parameters and (b) two auxiliary tasks, each modelled as a mixture of \mathcal{I} candidate tasks.

2.2 Example Auxiliary Tasks: Anatomical and Protocol Discrimination

Each auxiliary task is trained with a cross-entropy loss, between predicted class probability vectors y_j^a and ground-truth targets t_j^a in one-hot vectors.

$$\mathcal{L}_j^{CE} = - \sum_{N_j} (t_j^a \cdot \log[f_{\theta_j^s, \theta_j^a}(y_j^a | x)]) \quad (1)$$

The two tasks, minimising $\mathcal{L}_{j=1}^{CE}$ and $\mathcal{L}_{j=2}^{CE}$, classify different training subjects and types of scanning protocols, with $N_{j=1}$ and $N_{j=2}$ number of classes, respectively. An underlying assumption is that the anatomical and protocol variance can impact the 3D US reconstruction performance. This assumption is tested by quantifying the reconstruction accuracy changes, as number of subjects and/or protocol types are reduced. Results are reported in Sec. 3.2.

2.3 Parameterised Task Descriptor Locations

Assume \mathcal{I} locations in the main task network, at which a task descriptor $z_{i,j}$ can be conditioned, where $i = 1, \dots, \mathcal{I}$ for j^{th} auxiliary task. For a single main task in this case, each auxiliary task can branch out from these \mathcal{I} locations, as illustrated in Fig. 1 (b). The task descriptor thus represents the probability of branching location, with additional constraints $z_{i,j} \in [0, 1]$ and $\sum_i z_{i,j} = 1$.

With the task descriptor $z_{i,j}$, the j^{th} auxiliary task is parameterised by a mixture model of \mathcal{I} candidate tasks $f_{\theta_{i,j}^s, \theta_{i,j}^a}(y_{i,j}^a | x)$, each performed by one branch, where the additional subscript i is the candidate task index.

$$f_{\theta_{1,j}^s, \dots, \theta_{\mathcal{I},j}^s}(y_j^a | x, z_j) = \sum_{i=1}^{\mathcal{I}} [z_{i,j} \cdot f_{\theta_{i,j}^s, \theta_{i,j}^a}(y_{i,j}^a | x)] \quad (2)$$

where $z_j = [z_{i,j}]_{i=1, \dots, \mathcal{I}}^\top$ is the task descriptors for all the auxiliary tasks, with shared parameters $\theta_j^s = [\theta_{i,j}^s]_{i=1, \dots, \mathcal{I}}^\top$ and task-specific parameters $\theta_j^a = [\theta_{i,j}^a]_{i=1, \dots, \mathcal{I}}^\top$. For j^{th} auxiliary task, the final prediction is therefore the matrix product of all candidate (branch) predictions $y_j^a = [y_{i,j}^a]_{i=1, \dots, \mathcal{I}}^\top$ (a $N_j \times \mathcal{I}$ matrix) and the location weights z_j (a $\mathcal{I} \times 1$ vector, generated by a softmax function). Multiplicative task conditioning is also known as gating. The loss defined in Eq. 1 remains.

This formulation allows all candidate branches trained together without a predefined architecture decision. We show in Sec. 2.5 that the task descriptor may be considered as a hyperparameter and optimised using a gradient-based bi-level optimisation algorithm, for efficient inference using a single branch.

2.4 The Main Task and Evaluation

We adopt the approach in our previous work [4] for the main task, reconstructing a 3D US scan by sequentially predicting inter-frame transformations with inputs of frame sequences. The same reconstruction loss is used, but is now conditioned

on the two auxiliary tasks, denoted as $\mathcal{L}^{Rec}(t, f_{\theta, \theta_1^a, \theta_2^a}(y|x, z_1, z_2))$ where the network output is the six parameters for rigid spatial transformation. It is important to clarify that this proposed loss function also utilises a multi-task learning for predicting transformations between nearby frames, but different to and independent of that used in this work. Unless specified otherwise, the methodology and the implementation of the main task network, based on EfficientNet (b1) [18], remain the same, with further details in the original publication [4].

To test the generalisation and reconstruction performance of the proposed method, four evaluation metrics are used. For each frame, *frame prediction accuracy* (ϵ_{frame}) is used to evaluate the generalisation of the method, denoting the Euclidean distance between the ground-truth and prediction-transformed four corner points of each frame. The scan reconstruction performance is quantified by an *accumulated tracking error* ($\epsilon_{acc.}$), indicating the averaged point distance on all frame pixels, a *volume reconstruction overlap* (ϵ_{dice}), denoting the overlap of all pixels between the ground-truth and prediction volume, and a *final drift* (ϵ_{drift}), denoting the *frame prediction accuracy* of the last frame in a scan.

2.5 Bi-level Optimisation of Task Descriptor

Given a loss for the main task $\mathcal{L}^{Rec}(t, f_{\theta, \theta_1^a, \theta_2^a}(y|x, z_1, z_2))$ and those for the two auxiliary tasks defined in Eq. 1, the overall loss function thus is:

$$\mathcal{L}(\theta, \theta_1^a, \theta_2^a, z_1, z_2|D) = \mathcal{L}^{Rec}(\theta|x, t) + \sum_{j=1}^2 \mathcal{L}_j^{CE}(\theta, \theta_j^a, z_j|x, t_1^a, t_2^a) \quad (3)$$

where the loss is rearranged as a function of relevant network parameters $\{\theta, \theta_1^a, \theta_2^a\}$ and task descriptors $\{z_1, z_2\}$, with observed data $D = \{x, t, t_1^a, t_2^a\}$. Given a training data set \mathcal{D}_{train} and a validation data set \mathcal{D}_{val} , the empirical losses are $\mathcal{L}_{train}(\cdot) = \mathbf{E}_{D \in \mathcal{D}_{train}}[\mathcal{L}(\cdot|D)]$ and $\mathcal{L}_{val}(\cdot) = \mathbf{E}_{D \in \mathcal{D}_{val}}[\mathcal{L}(\cdot|D)]$, respectively. Optimising the task descriptors subject to optimised network parameters leads to the following meta-learning task with a bi-level optimisation problem:

$$\begin{aligned} \hat{z}_1, \hat{z}_2 = \arg \min_{z_1, z_2} \mathcal{L}_{val}(\hat{\theta}, \hat{\theta}_1^a, \hat{\theta}_2^a; z_1, z_2) \\ \text{s.t. } \hat{\theta}, \hat{\theta}_1^a, \hat{\theta}_2^a = \arg \min_{\theta, \theta_1^a, \theta_2^a} \mathcal{L}_{train}(\theta, \theta_1^a, \theta_2^a; z_1, z_2) \end{aligned} \quad (4)$$

Since the $\frac{\partial \mathcal{L}_{val}}{\partial z_j}$ can be estimated, the task descriptors can be optimised by gradient-based updates alternating between minimising the two empirical loss functions [17]. The numerical algorithm is summarised in Algorithm 1. The three tasks are weighed equally in Eq. 3 as the task descriptor hyperparameters make explicit optimising or predefining these weights redundant.

Algorithm 1: The bi-level optimisation algorithm.

1. Initialise task descriptors $\{z_1, z_2\}$ for two auxiliary tasks.
 2. Optimise the task descriptors using meta-learning algorithm in a differentiable network:
 - while** *not converged* **do**
 - (1) Update network parameters $\{\theta, \theta_1^a, \theta_2^a\}$ for the main task and all auxiliary tasks by descending $\nabla_{\theta, \theta_1^a, \theta_2^a} \mathcal{L}_{train}(\theta, \theta_1^a, \theta_2^a; z_1, z_2)$
 - (2) Update task descriptors $\{z_1, z_2\}$ by descending $\nabla_{z_1, z_2} \mathcal{L}_{val}((\theta, \theta_1^a, \theta_2^a) - \xi \nabla_{\theta, \theta_1^a, \theta_2^a} \mathcal{L}_{train}(\theta, \theta_1^a, \theta_2^a; z_1, z_2); z_1, z_2)$.
 3. Finalise the network architecture using the optimised task descriptors $\{z_1, z_2\}$.
-

In this work, the first-order approximation was used when optimising task descriptors with $\xi = 0$, i.e., $\nabla_{z_1, z_2} \mathcal{L}_{val}(\theta, \theta_1^a, \theta_2^a; z_1, z_2)$ [6].

3 Experiments and Results

3.1 Dataset and Network Settings

The US data used in this study was from our previously published data set in [4], acquired at 20 frame per second (fps) by an Ultrasonix machine (BK, Europe) with a curvilinear probe (4DC7-3/40, 6 MHz, 9 cm depth and a median level of speckle reduction) and an NDI Polaris Vicra (Northern Digital Inc., Canada) tracker. All US images with a size of 480×640 pixels, after spatial and temporal calibration, acquired from 38 forearms of 19 volunteers were used in this study, more than 40,000 US frames in total. For each forearm, three predefined scanning protocols (straight line shape, ‘C’ shape and ‘S’ shape, as in Fig. 2 (a)), in a distal-to-proximal direction, with the US probe perpendicular of and parallel to the forearm, were acquired, resulting in 6 different protocols and 228 scans in total. The US scans have a various number of frames, from 36 to 430 frames, equivalent to a probe travel distance of between 100 and 200 mm. The data was split into train, validation and test sets by a ratio of 3:1:1 on a scan level.

With the EfficientNet (b1) for the main reconstruction task [4], nine locations were used for candidate branches, with each being a single fully-connected classification layer, denoted as Branches 1-9 and Branch 1 being closest to network input. The nine candidate predictions are weighted by the softmax-generated task predictor, as described in Sec. 2.3, to form each final auxiliary task prediction. The two auxiliary tasks resulted in total of 18 branches.

A minibatch of 32, an Adam optimizer were used for model training, with a learning rate of 10^{-4} (tested among $\{10^{-3}, 10^{-4}, 10^{-5}\}$) and the input sequence length being $M = \{100, 140\}$ (tested among $M = \{49, 75, 100, 140\}$), selected based on the validation set performance. Each model was trained for at least 15,000 epochs until convergence, for up to 4 days, on Ubuntu 18.04.6 LTS with a single NVIDIA Quadro P5000 GPU card. The model with the best validation set performance was selected to evaluate test set performance. Other hyperparameters were found relatively insensitive to model performance and configured empirically based on validation performance.

3.2 The Effect of Anatomical and Protocol Variances

The impact on performance of the main task was tested by training models with various anatomical and protocol variance in the training data: 1) all training data (All); 2) straight only (Straight); 3) C-shape and S-shape (C-S); 4) 25% subjects in training data (Sub 25%); 5) 50% training subjects (Sub 50%); 6) 75% training subjects (Sub 75%); 7) 50% of frames in a scan (Frm 50%) and 8) 75% of frames (Frm 75%), and tested on the same test set. Fig. 2 (b) plotted $\epsilon_{acc.}$ with $M = 20$ as an example of the reconstruction performance using different training sets. The other models and metrics yielded a consistent trend, which saw $\epsilon_{acc.}$ increased with both reduced anatomical and protocol variances. It indicates both are factors impacting the main reconstruction task performance.



Fig. 2: Illustration of different US acquisition protocols and their impact to reconstruction accuracy. (a) Protocols of straight line shape (p1, p4), ‘C’ shape (p2, p5) and ‘S’ shape (p3, p6) with US probe perpendicular of and parallel to the forearm, (b) $\epsilon_{acc.}$ changes due to reduced anatomical and protocol variance.

3.3 Ablation Studies and Comparison

In our experiments, two types protocol discrimination tasks are considered, six-class classification (as in Fig. 2) and three-class classification (combining the perpendicular and parallel scans for the same scan shapes). Each is combined with the 38-class classification task (38 training subjects) as the anatomical discrimination task. Results from different $M = 100, 140$ are also included, together with those from the main task network without any branches (no-branch).

The optimised $z_{i,j}$ values versus the training epochs are plotted in Fig. 3 (a) and (b), where the optimum branch was selected by the maximum $z_{i,j}$ value, at the epoch. Table. 1 summarised the reconstruction performance of the proposed method, using the optimised branches (their indices are denoted with asterisks) for the two auxiliary tasks. Comparing with the no-branch models, the improved performances from the proposed methods can be seen, for both $M = 100$ and $M = 140$, regardless of the three- or six-class were used as protocol discrimination tasks. For example, at $M = 100$, ϵ_{drift} was lowered from 14.52 to 6.56 mm, using the proposed methods with optimised Branches 4 and 4, for protocol and anatomical discrimination tasks, with p -value = 0.013 (unpaired t-test, at

a significance level $\alpha = 0.05$). Statistical significance was found in performance improvement using all four evaluation metrics.

Although no previous work utilise these discrimination tasks as privileged information for assisting this application, two baseline models were implemented as alternatives to the main task network. We have re-implemented the proposed method using one of the first proposed approaches for this application [16] as the main task network, with two adjacent frames as input and outputting the transformation between them. In addition, the same network architecture using more (10) input frames were also tested, denoted as ‘[16]-10’. These results are also summarised in Table. 1, with and without optimised branches. However, we would like to emphasize that the reported inferior results from these compared baselines need to be interpreted with caution, as they were neither designed for nor tuned with incorporating these auxiliary tasks used in this study. They are included for completeness and reference values.

Table 1: Mean and standard deviation of four metrics of the proposed method and no-branch method.

M	Num. of protocols	Branch index (protocol/anatomy)	ϵ_{frame}	$\epsilon_{acc.}$	ϵ_{dice}	ϵ_{drift}
100	n/a	No-branch	0.18 ± 0.05	7.03 ± 3.97	0.84 ± 0.08	14.52 ± 10.51
100	3	Branches 5*/4*	0.19 ± 0.06	3.92 ± 3.50	0.73 ± 0.21	7.06 ± 7.30
100	6	Branches 4*/4*	0.17 ± 0.08	3.80 ± 3.97	0.76 ± 0.24	6.56 ± 7.53
140	n/a	No-branch	0.14 ± 0.05	3.68 ± 3.10	0.62 ± 0.28	7.30 ± 7.40
140	3	Branches 9*/4*	0.15 ± 0.08	3.36 ± 3.26	0.94 ± 0.00	6.20 ± 6.31
140	6	Branches 5*/7*	0.13 ± 0.05	2.90 ± 2.10	0.89 ± 0.00	6.53 ± 5.98
[16]	n/a	No-branch	0.59 ± 0.28	29.03 ± 9.15	0.43 ± 0.32	35.67 ± 11.20
[16]	3	Branches 1*/9*	0.70 ± 0.50	32.71 ± 18.10	0.60 ± 0.22	59.53 ± 36.87
[16]	6	Branches 4*/9*	0.68 ± 0.46	30.29 ± 17.44	0.67 ± 0.16	55.05 ± 32.69
[16]-10	n/a	No-branch	0.38 ± 0.21	17.60 ± 9.77	0.67 ± 0.22	22.64 ± 12.47
[16]-10	3	Branches 4*/4*	0.42 ± 0.30	19.37 ± 10.63	0.50 ± 0.30	26.32 ± 13.33
[16]-10	6	Branches 9*/4*	0.43 ± 0.38	21.72 ± 12.74	0.50 ± 0.25	29.64 ± 16.29

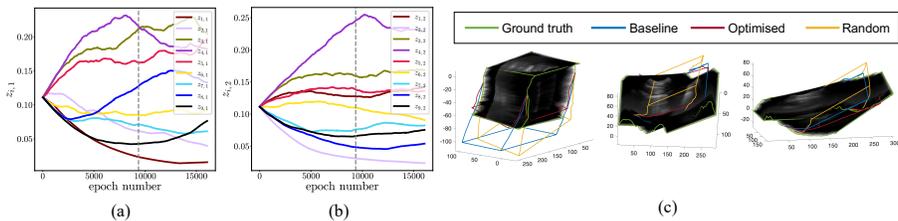


Fig. 3: The trend of task descriptor for anatomy (a) and protocol (b), $M = 100$, and the reconstruction performance (c). The epoch indicating best performance of the model on validation set is denoted by a gray dotted line. Scans with various scanning path are reconstructed using no-branch, optimised branches, and random branches strategies, $M = 100$.

4 Conclusion and Discussion

This work demonstrated the impact of anatomical and protocol variance towards the 3D reconstruction of trackerless freehand US and formulated two respective discrimination tasks for taking advantage these privileged information during training. Using the proposed algorithm, substantially improved reconstruction performance was achieved, which may indicate a promising new direction for improving the potentials of this application for clinical adoption. Future work includes testing clinical applications with specific challenges, such as those without predefined protocol classes (where a clustering task may be used instead), and comparison with approaches such as gradient surgery [22], which may need adaptation for a single main task.

Declarations. This work was supported by the EPSRC [EP/T029404/1], a Royal Academy of Engineering / Medtronic Research Chair [RCSR1819\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. 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 20th Jan. 2023 [24055/001].

References

1. Chang, R.F., Wu, W.J., Chen, D.R., Chen, W.M., Shu, W., Lee, J.H., Jeng, L.B.: 3-d us frame positioning using speckle decorrelation and image registration. *Ultrasound in medicine & biology* **29**(6), 801–812 (2003)
2. Guo, H., Chao, H., Xu, S., Wood, B.J., Wang, J., Yan, P.: Ultrasound volume reconstruction from freehand scans without tracking. *IEEE Transactions on Biomedical Engineering* **70**(3), 970–979 (2022)
3. Lang, A., Mousavi, P., Gill, S., Fichtinger, G., Abolmaesumi, P.: Multi-modal registration of speckle-tracked freehand 3d ultrasound to ct in the lumbar spine. *Medical image analysis* **16**(3), 675–686 (2012)
4. Li, Q., Shen, Z., Li, Q., Barratt, D.C., Dowrick, T., Clarkson, M.J., Vercauteren, T., Hu, Y.: Trackerless freehand ultrasound with sequence modelling and auxiliary transformation over past and future frames. *arXiv preprint arXiv:2211.04867* (2022)

5. Lindseth, F., Kaspersen, J.H., Ommedal, S., Langø, T., Bang, J., Hokland, J., Unsgaard, G., Nagelhus Hemes, T.A.: 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)
6. Liu, H., Simonyan, K., Yang, Y.: Darts: Differentiable architecture search. arXiv preprint arXiv:1806.09055 (2018)
7. Luo, M., Yang, X., Huang, X., Huang, Y., Zou, Y., Hu, X., Ravikumar, N., Frangi, A.F., Ni, D.: Self context and shape prior for sensorless freehand 3d ultrasound reconstruction. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI* 24. pp. 201–210. Springer (2021)
8. Luo, M., Yang, X., Wang, H., Du, L., Ni, D.: Deep motion network for freehand 3d ultrasound reconstruction. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*. pp. 290–299. Springer (2022)
9. Mikaeili, M., Bilge, H.Ş.: Trajectory estimation of ultrasound images based on convolutional neural network. *Biomedical Signal Processing and Control* **78**, 103965 (2022)
10. Miura, K., Ito, K., Aoki, T., Ohmiya, J., Kondo, S.: Localizing 2d ultrasound probe from ultrasound image sequences using deep learning for volume reconstruction. In: *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis: First International Workshop, ASMUS 2020, and 5th International Workshop, PIPPI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 1*. pp. 97–105. Springer (2020)
11. Miura, K., Ito, K., Aoki, T., Ohmiya, J., Kondo, S.: Pose estimation of 2d ultrasound probe from ultrasound image sequences using cnn and rnn. In: *Simplifying Medical Ultrasound: Second International Workshop, ASMUS 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 2*. pp. 96–105. Springer (2021)
12. Mozaffari, M.H., Lee, W.S.: Freehand 3-d ultrasound imaging: a systematic review. *Ultrasound in medicine & biology* **43**(10), 2099–2124 (2017)
13. Newell, A., Jiang, L., Wang, C., Li, L.J., Deng, J.: Feature partitioning for efficient multi-task architectures. arXiv preprint arXiv:1908.04339 (2019)
14. Ning, G., Liang, H., Zhou, L., Zhang, X., Liao, H.: Spatial position estimation method for 3d ultrasound reconstruction based on hybrid transformers. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. pp. 1–5. IEEE (2022)
15. Prevost, R., Salehi, M., Jagoda, S., Kumar, N., Sprung, J., Ladikos, A., Bauer, R., Zettinig, O., Wein, W.: 3d freehand ultrasound without external tracking using deep learning. *Medical image analysis* **48**, 187–202 (2018)
16. Prevost, R., Salehi, M., Sprung, J., Ladikos, A., Bauer, R., Wein, W.: Deep learning for sensorless 3d freehand ultrasound imaging. In: *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part II*. pp. 628–636. Springer (2017)
17. Rajeswaran, A., Finn, C., Kakade, S.M., Levine, S.: Meta-learning with implicit gradients. *Advances in neural information processing systems* **32** (2019)
18. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning*. pp. 6105–6114. PMLR (2019)

19. Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. *Neural networks* **22**(5-6), 544–557 (2009)
20. Wu, S., Zhang, H.R., Ré, C.: Understanding and improving information transfer in multi-task learning. *arXiv preprint arXiv:2005.00944* (2020)
21. Xie, Y., Liao, H., Zhang, D., Zhou, L., Chen, F.: Image-based 3d ultrasound reconstruction with optical flow via pyramid warping network. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 3539–3542. IEEE (2021)
22. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. *Advances in Neural Information Processing Systems* **33**, 5824–5836 (2020)