Deep Learning-based Plane Pose Regression towards Training in Freehand Obstetric Ultrasound

Chiara Di Vece, Brian Dromey, Francisco Vasconcelos, Anna L David, Donald M Peebles and Danail Stoyanov

May 31, 2022

Deep Learning-based Plane Pose Regression towards Training in Freehand Obstetric Ultrasound

Chiara Di Vece[1], Brian Dromey[1,2], Francisco Vasconcelos[1], Anna L. David[2,3], Donald Peebles[2,3], and Danail Stoyanov[1]

[1]*Wellcome/EPSRC Centre for International and Surgical Sciences, University College London*
[2]*Elizabeth Garrett Anderson Institute for Women's Health, University College London*
[3]*NIHR UCL Hospitals Biomedical Research Centre, University College London*
chiara.divece.20@ucl.ac.uk

INTRODUCTION

In obstetrics ultrasound (US) training, the learner's ability to mentally build a three-dimensional (3D) map of the fetus from a two-dimensional (2D) US image represents a major challenge. Pre-defined anatomical planes, known as standard planes (SPs), retain a significant clinical relevance. Their correct acquisition requires extensive experience due to the complexity of SP definition, and their high intra- and low inter-class variation. Previous work proposed automating the extraction of SPs from data acquired with a simplified protocol rather than assisting operators in acquiring typical freehand 2D SPs. These methods, based on traditional machine learning (ML) approaches [1] or convolutional neural network (CNN) [2], are mostly confined to selection or classification of SPs. Automatic approaches for the localisation of planes in 3D volumes are based either on supervised learning (SL) [3] or reinforcement learning (RL) [4]. The first ones, even though effective, can mainly localize one single plane at a time or are tailored to just one organ. Also, they attempt to learn a mapping from high-dimensional volumetric data to low-dimensional abstract features directly, making the training difficult. In the second case, most of the available solutions employing deep reinforcement learning (DRL) are based on robotic navigation and are used at the point of care. To the best of our knowledge, there are no available training systems to guide the sonographers in freehand obstetric US.
Currently, training in obstetric US is focused on SPs recognition. The autonomous probe navigation towards SPs remains a highly challenging task due to the need to interpret variable and complex images and their spatial relationship. Our work aims to develop a real-time training platform to guide inexperienced sonographers in acquiring proper obstetric US images that could be potentially deployed for existing US machines.

MATERIALS AND METHODS

The development of our training system has been divided into five main blocks (Figure 1).

1. Unity simulator for volume reconstruction and synthetic images acquisition: The fetal volume can be sliced with arbitrarily oriented planes. The user can visualize them and annotate the standard ones. Besides, the environment can be used for the automated generation of supervised data using pre-acquired 3D US volumes.

2. Plane pose regression system: We use a regression CNN [5] to estimate the 6D pose of arbitrarily oriented US planes using a continuous rotation representation [6]. Our method is purely image-based and, therefore, does not require tracking sensors. Besides, it differs from traditional slice-to-volume registration methods since it does not require a previously acquired volume of the same subject being scanned. Instead, we predict poses relative to a generalized brain center, where training and test data belong to different subjects. The network was trained on synthetic images acquired from phantom US volumes and fine-tuned on real scans. Training data was generated by slicing US volumes in Unity at random coordinates and more densely around the manually annotated transventricular (TV) SP.

Network architecture. We used ResNet-18 as a backbone for feature extraction. We modified the network by re-initialising the fully connected layer and adding a regression head to directly output the rotation and translation representations. The input is the US image I (128×128) obtained by slicing the volume. The CNN predicts the 6D pose $\theta_{Pred} = (t'_x, t'_y, t'_z, \alpha'_x, \alpha'_y, \alpha'_z)$. Specifically, the network outputs a vector of nine parameters $\theta_{Out} = (t_1, t_2, t_3, r_1, ..., r_6)$; the first three are used for the translation and the last six for the rotation, then used internally by our CNN to reconstruct the rotation matrix $\mathbf{R}'$ in the forward pass.

Loss function. We used the mean squared error (MSE) between predicted ($\mathbf{t}'$,$\mathbf{R}'$) and ground truth ($\mathbf{t}$,$\mathbf{R}$) values: $\mathcal{L}_{Tot} = \frac{1}{N}\sum_{t=1}^{N} \|\mathbf{R}' - \mathbf{R}\|_2 + \lambda \frac{1}{N}\sum_{t=1}^{N} \|\mathbf{t}' - \mathbf{t}\|_2$, where N denotes the total number of images I within one training epoch, and λ the hyperparameter balancing between the rotation and translation losses.

Experiments. Our framework is implemented in *PyTorch* and trained using a Tesla® V100-DGXS-32GB GPU of an NVIDIA DGX station for 50 epochs with batch size of K = 100, Adam optimiser and weighting factor λ = 0.01. We performed two experiments.

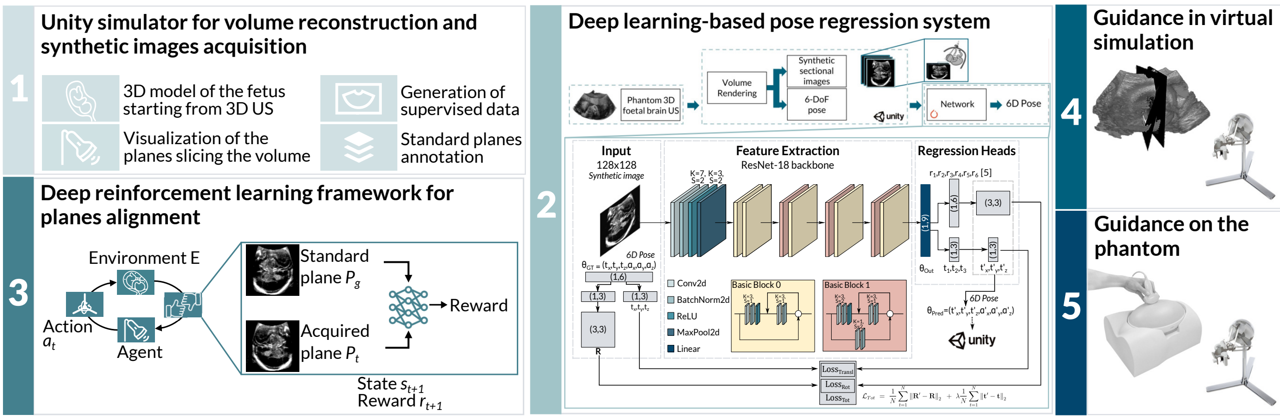

Fig. 1 Overall structure of the training system for obstetric US acquisitions using vision-based DRL

- *Experiment 1*: **(1.1)** Training (4 scans, 75088 images) and testing (2 scans, 37544 images) on phantom data (same phantom, 23 weeks) with weights from ImageNet; **(1.2)** Training (2 scans of one fetus of 23 weeks, 38754 images) and testing (2 scans of another fetus of 24 weeks, 38754 images) on real data with weights from the case 1.1. The test sets are divided into random planes (*Test RP*) and planes around the TV SP (*Test SP*). For translation, we employed the Euclidean distance between the two planes (*mm*). For rotation, we display errors as the geodesic distance to ground truth (°) defined as $Error_{Rot} = \arccos((\mathbf{R}'^{-1}_{00}+\mathbf{R}'^{-1}_{11}+\mathbf{R}'^{-1}_{22}-1)/2)$.
- *Experiment 2*: We fed sectional images of the TV SPs into the network to estimate their pose as a sanity test. We plotted back the two planes within the volume into Unity to visually evaluate the distance between the annotated TV SPs and the predicted ones (**2.1**, **2.2**).

3. DRL framework for planes alignment: The process reminds of the sonographers' behaviour as they continuously manipulate (*action*) the US probe (*agent*) to scan the maternal abdomen (*environment*) while visualising the intermediate planes P_t on the screen (*state*) until the SP P_g is acquired (*reward* and *terminal state*).

4. Guidance in simulation: Once the direction to align the plane P_t to the standard one P_g is found, the users could be guided through a haptic device by making them *feel* a force pointing towards the correct direction.

5. Guidance on the phantom/patient: The system could be developed as an interface and tested in a clinical setting, with validation studies with novice operators.

RESULTS

Considering the pose regression experiments, for phantom data, the median errors are 0.90 mm/1.17° and 0.44 mm/1.21° for random planes and planes close to TV one, respectively. For real data, testing on a fetus of 24 weeks, these errors are 1.89 mm/2.75° and 1.03 mm/0.86°. The average inference time is 2.97 ms per plane. Figures 2 reports the translation and rotation error distributions and the sanity test results for phantom and real data.

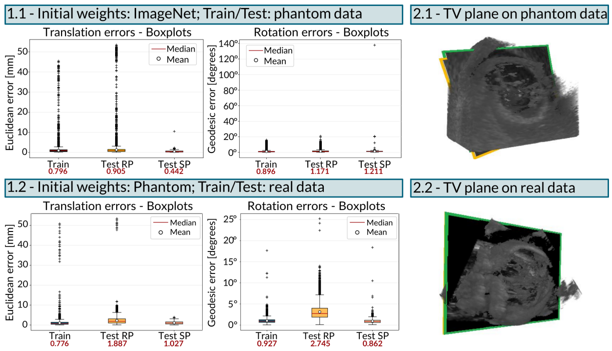

Fig. 2 Left: Translation and rotation error distributions in phantom (1.1) and real (1.2) US data for planes at random coordinates (Test RP) and around the annotated TV SP (Test SP). Right: Visual evaluation of TV SP prediction on phantom (2.1) and real (2.2) US data

DISCUSSION

Our regression CNN can reliably localize US planes within the fetal brain in the phantom, regularly used for clinical training and evaluation of skills; it successfully generalizes pose regression to an unseen fetal brain without the need for real ground truth data in real-time or 3D volume scans of the patient beforehand. Future development will expand the prediction to volumes of the whole fetus and assess its potential for vision-based, freehand US assisted navigation for acquiring SPs.

REFERENCES

[1] D. Ni *et al.*, "Standard Plane Localization in US by Radial Component Model and Selective Search," *Ultrasound Med Biol*, 2014.
[2] C. Baumgartner *et al.*, "SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand US," *IEEE Trans Med Imaging*, 2017.
[3] A. Schmidt-Richberg *et al.*, "Offset regression networks for view plane estimation in 3D fetal US," in *Medical Imaging*, 2019.
[4] K. Li *et al.*, "Autonomous navigation of an us probe towards standard scan planes with deep reinforcement learning," *IEEE ICRA*, 2021.
[5] C. Di Vece *et al.*, "Deep learning-based plane pose regression in obstetric ultrasound," *Int J Comput Assist Radiol Surg*, 2022.
[6] Y. Zhou *et al.*, "On the continuity of rotation representations in neural networks," in *IEEE CVPR*, 2019.