Motion Correction Using Deep Learning Neural Networks – Effects of Data Representation

Rifkat Zaydullin
Department of Bioengineering
Imperial College London
London, United Kingdom
rifkat.zaydullin16@imperial.ac.uk

Anil A. Bharath
Department of Bioengineering
Imperial College London
London, United Kingdom
a.bharath@imperial.ac.uk

Enrico Grisan
School of Engineering
London South Bank University
London, United Kingdom
enrico.grisan@lsbu.ac.uk

Kirsten Christensen-Jeffries
Department of Biomedical Engineering
King's College London
London, United Kingdom
kirsten.christensen-jeffries@kcl.ac.uk

Wenjia Bai
Department of Brain Sciences &
Department of Computing
Imperial College London
London, United Kingdom
w.bai@imperial.ac.uk

Meng-Xing Tang
Department of Bioengineering
Imperial College London
London, United Kingdom
mengxing.tang@imperial.ac.uk

***Abstract*—An in-silico investigation of the effects of ultrasound data representation on the accuracy of the motion prediction made using deep learning neural networks was carried out. The representations studied include: linear ('envelope'), log compressed, linear with phase and log compressed with phase. A UNet model was trained to predict non-rigid deformation field using a fixed and a moving image pair as the input. The results illustrate that the choice of the representation plays an important role on the accuracy of motion estimation. Specifically, representations with phase information outperform the representations without phase. Furthermore, log-compressed data yielded predictions with higher accuracy than the linear data.**

***Index Terms*—deep learning, motion correction, motion estimation, ultrasound, data representation**

I. Introduction

The success of motion correction (MoCo) deep learning (DL) models in the field of computer vision inspired researchers to apply these techniques for MoCo of ultrasound (US) images.

To date, multiple investigation looked at the accuracy of DL MoCo and tracking algorithms. They focused on one key difference between conventional and US images: the presence of speckle pattern and its decorrelation. The studies conclude that the predictions of the DL models are comparable in terms of accuracy to the classical algorithms [1]–[6]. However, there are other notable differences between the conventional and US images. Namely, the availability of the phase information and log-compression of the images.

Therefore, this investigation aims to determine how representation of the ultrasound data affects the accuracy of motion correction of deep learning algorithms. This is the first study carried out with this aim.

II. Method

A. Data representation

Following the acquisition of the signal, IQ demodulation and delay-and-sum beamforming (BF), US data is complex valued. This value at each pixel location can be represented as $re^{i\theta}$ (amplitude with phase). This is referred to as 'envelope with phase' representation, since the real valued r corresponds the 'envelope' of the data. If r is log compressed with a given dynamic range, it is referred to as a 'log' representation. In this investigation a 40 dB dynamic range was used. Analogously to the 'envelope' data, 'log' can be supplemented with the phase information (θ), giving raise to the 'log with phase' representation.

B. Data generation

In this study, in-silico soft tissue phantoms were deformed using non-rigid spatial deformation fields (DFs). Phantoms with realistic structure were obtained by converting publicly available clinical BMode images [7] into 2D probability density functions, which were used to randomly sample scatterer locations. To ensure a fully developed speckle, the phantom was populated with scatterer density of 10 scatterers per resolution cell [8]. The phantom was imaged using single plane wave transmissions, simulated using Field II [9], [10] software. The final images were obtained using delay-and-sum BF. An example of an image produced using this data generation pipeline is illustrated by Fig. 1.

The non-rigid deformation was modelled using polynomials of a random order:

$$di(x,y,z) = a_x P^{(N_x)}(x) + a_y P^{(N_y)}(y) + a_z P^{(N_z)}(z), \quad (1)$$

where $i \in \{x,y,z\}$, di is the displacement along the i-axis, N_i is the order of the polynomial and a_i is a randomised

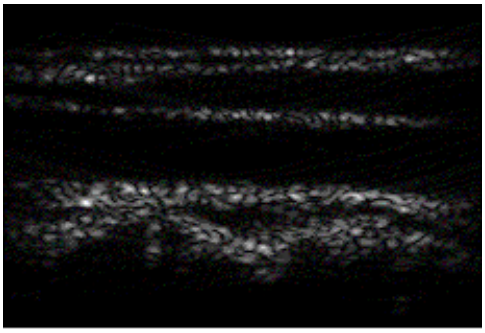

Fig. 1. An example of the simulated BMode image.

coefficient of the polynomial $P^{(N_i)}$. The final deformation was obtained by combining the displacements along the three axes:

$$d\boldsymbol{r} = b_x dx\hat{\boldsymbol{x}} + b_y dy\hat{\boldsymbol{y}} + b_z dz\hat{\boldsymbol{z}}, \tag{2}$$

where b_i is a randomised constant and $\hat{\boldsymbol{i}}$ is the unit vector. In this investigation, $a_y = b_y = 0$, which removed y-dependence and the out-of-plane motion. Furthermore, the 0^{th} order term in the polynomial was set to zero to remove global translational movement. This produced a smooth DF, dependent on x and z coordinates of the phantom, example of which is illustrated by Fig. 2.

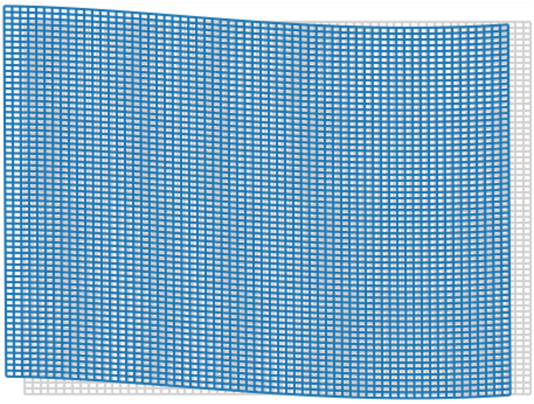

Fig. 2. An example of a position dependent non-rigid deformation field. In gray: regular grid. In cyan: regular grid deformed using the non-rigid deformation field.

C. Architecture and the loss function

The model selected for this investigation was a UNet [11] architecture, illustrated by the Fig. 3. The input of the model was a pair of images (fixed and moving) and the output was the DF. The model was implemented using the PyTorch [12] deep learning framework. In the case of 'envelope' and 'log' representations, the model had two input channels. In the case of 'envelope with phase' and 'log with phase' representations, the model had four input channels. For example, 'envelope with phase' representation had two channels dedicated to r component of the pair images and the other two dedicated to the θ component. Due to the varied number of input channels, the model had 200,000 $\pm 1\%$ trainable parameters.

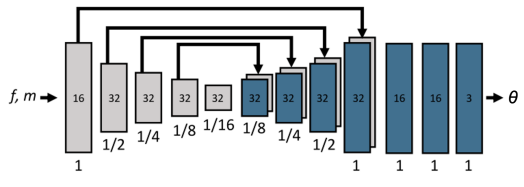

Fig. 3. UNet architecture used. The input to the model is the pair of fixed (f) and moving (m) images. The output is the deformation field (θ). Figure adapted from [13].

The model was trained using the mean square error between the predicted DF of the model and the ground truth DF. The model was trained on the each data representation independently. Each model was trained for 300 epochs and the best performing epoch (in terms of validation loss) was selected for the subsequent evaluation.

III. Results

The different representations were evaluated by calculating the vector magnitude error (VME) between the ground truth ($\vec{\phi}_{\text{GT}}$) and the predicted ($\vec{\phi}_{\text{pred}}$) DFs, given by the equation:

$$\text{VME} = |\vec{\phi}_{\text{GT}} - \vec{\phi}_{\text{pred}}|. \tag{3}$$

VME was calculated for each pixel in the entire testing dataset. The results of studied the data representations are illustrated by Fig. 4 and Fig. 5.

The null hypothesis used in the investigation assumes a zero vector at all locations as the predicted DF, which illustrated the distribution of the vector magnitudes of the ground truth DFs.

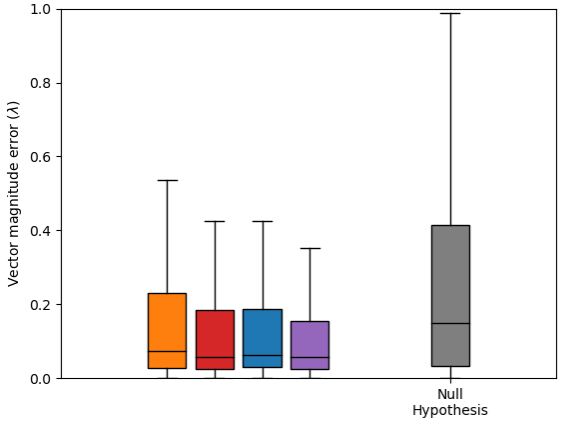

Fig. 4. The effect of data representation on the accuracy on the predicted deformation fields for displacements up to 2.5λ. The data representations are colour coded as: orange – 'envelope', red – 'envelope with phase', blue – 'log', purple – 'log with phase'.

IV. Discussion

The results demonstrate that data representation has an effect on the accuracy of the predicted DFs with the best performing representation being the 'log with phase'. When comparing 'envelope' & 'envelope with phase' and 'log' &

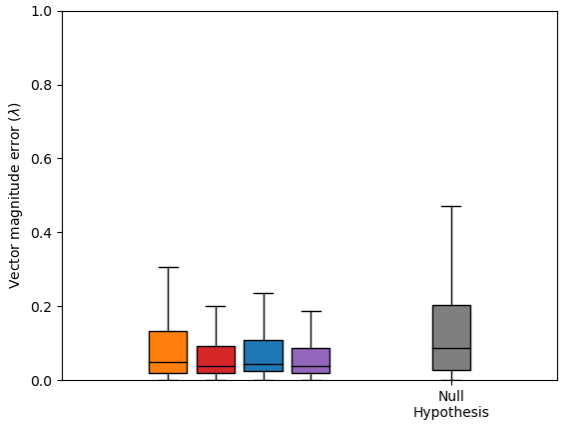

Fig. 5. The effect of data representation on the accuracy on the predicted deformation fields for displacements up to λ. The data representations are colour coded as: orange – 'envelope', red – 'envelope with phase', blue – 'log', purple – 'log with phase'.

'log with phase', the results illustrate that the addition of the phase information improves the accuracy of the prediction. Additionally, they illustrate that log compression of the data can be advantageous. However, the impact of log compression is minimal if displacement does not exceed one wavelength, provided phase information is present (Fig. 5 comparing 'envelope with phase' & 'log with phase').

V. Conclusion

The investigation, carried out in-silico, studied the relationship between ultrasound data representation and the accuracy of the motion prediction using DL MoCo algorithms. This study was carried out in the context of non-rigid deformations and a UNet architecture. The study finds that the choice of data representations plays an important role in the accuracy of the prediction. Specifically, the inclusion of the phase information improves the accuracy of the prediction. Additionally, log-compression of the linear envelope data can be beneficial.

Acknowledgment

The authors of the paper thank Medical Research Council, Engineering and Physical Sciences Research Council and Smart Medical Imaging CDT for their support of the work.

References

[1] E. Evain *et al.*, "A pilot study on convolutional neural networks for motion estimation from ultrasound images," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 67, no. 12, pp. 2565–2573, 2020.

[2] F. Liu *et al.*, "Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences," *Medical Image Analysis*, vol. 65, p. 101793, 2020.

[3] P. Huang *et al.*, "Attention-aware fully convolutional neural network with convolutional long short-term memory network for ultrasound-based motion tracking," *Medical physics (Lancaster)*, vol. 46, no. 5, pp. 2275–2285, 2019.

[4] S. Bharadwaj and M. Almekkawy, "Deep learning based motion tracking of ultrasound image sequences," in *2020 IEEE International Ultrasonics Symposium (IUS)*, 2020, pp. 1–4.

[5] K. Ta *et al.*, "A semi-supervised joint learning approach to left ventricular segmentation and motion tracking in echocardiography," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1734–1737.

[6] C. Xiao *et al.*, "A new deep learning method for displacement tracking from ultrasound rf signals of vascular walls," *Computerized Medical Imaging and Graphics*, vol. 87, p. 101819, 2021.

[7] M. Zukal, R. Beneš, P. Číka, and K. Říha, "Ultrasound image database," Department of Telecommunications, Brno University of Technology, 2011. [Online]. Available: http://splab.cz/en/download/databaze/ultrasound

[8] J. M. Thijssen, "Ultrasonic speckle formation, analysis and processing applied to tissue characterization," *Pattern Recognition Letters*, vol. 24, no. 4, pp. 659–675, 2003.

[9] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordic-Baltic Conf. Biomed. Imag.*, vol. 34, 1996, pp. 351–353.

[10] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 39, no. 2, pp. 262–267, 1992.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

[12] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.

[13] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, p. 1788–1800, 08 2019.