Using Deep Learning to Estimate Frame-to-Frame Angle Displacements in 2D Ultrasound Image Sequences of an Infant Hip

A. Kaderdina[1], M.J. Bontá Suárez[2], R. Garbi[3], E. Schaeffer[4], K. Mulpuri[4], and A. Hodgson[1]

[1]Department of Mechanical Engineering, University of British Columbia, Vancouver, Canada
[2]School of Biomedical Engineering, University of British Columbia, Vancouver, Canada
[3]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
[4]Orthopaedic Surgery, British Columbia Children's Hospital, Vancouver, Canada

Abstract— **To assess developmental dysplasia of the hip in infants, evaluations are currently conducted based on 2D ultrasound (US) images. Using 3D US has been shown to markedly reduce inter-rater variability, but 3D scanners are not widely available in pediatric practices. Here, we propose using deep learning to estimate the spatial positions of 2D US image sequences; this can then be used to form 3D reconstructions. In this study, we extracted fan-shaped sets of slices from a database of 1403 3D US volumes and trained a previously proposed standard convolutional neural network (CNN) as well as two variations of a deeper CNN (one augmented with optical flow (OF) information) to estimate the angular distances between separated slices. The deeper CNN most accurately predicted the inter-slice angular displacements, with a mean absolute error of 0.02˚, for displacements of up to 3.0˚ (corresponding to a center-frame displacement of 5.3mm). OF did not appear to improve prediction accuracy in angle estimation. The deeper CNN also achieved a mean end-to-end sweep angle error of -0.8% ± 13.2%, compared with an error of 25.3% ± 14.7% for the previously proposed standard CNN. This relatively low error suggests that it may be feasible to accurately reconstruct a 3D representation of an infant hip using a 2D US video stream alone, without requiring additional probe-tracking devices.**[1]

Keywords— **Ultrasound, Developmental dysplasia of the hip, Deep learning, Orthopaedics.**

I. Introduction

Developmental dysplasia of the hip (DDH) is a malformation of the hip joint, affecting 1-3% of infants [1]. Currently, 2D ultrasound (US) probes are used to screen for DDH. Unfortunately, due to the inherent difficulty of characterizing 3D anatomy based on 2D US images, diagnosis of DDH is prone to large variability, with one study showing 29% of cases being under-treated [2]. Quader [3] showed the inter-rater variability in DDH metrics based on 3D US imaging was reduced by approximately 75% when compared to 2D US. However, since 3D US probes are not readily available in neonatal clinics, it would be useful to instead reconstruct 3D US volumes from 2D US images.

Spatial compounding techniques, in which the locations of 2D US slices are measured using optical tracking tools, have been used for many years [4]. In addition, it has been known that the degree of speckle correlation between nearby US image planes decreases with distance. Recently, Prevost et al. [5] introduced the use of machine learning to infer relative positions of sequential 2D US slices and found that adding optical flow (OF) information and data from an inertial measurement unit could improve performance. Since then, several research groups have offered variations on the deep learning approach [6]–[8]. Luo et al [9] proposed adding self-supervised and adversarial learning into the training process, making use of context cues and shape priors in the volume construction; their 3D US DDH dataset consisted of 101 volumes collected from 14 participants. Most of the proposed approaches involve position estimation of sequential slices, however, we eliminate the temporal aspect and only investigate angular displacements between any two neighboring slices. In this study, we assess the performance of the network proposed by Prevost in the specific context of imaging for DDH and compare it against two variations of a deeper convolutional neural network (CNN) architecture (with and without OF information). Our overall goal is to reconstruct 3D US from spatially located US images, and then apply a dysplasia metric extraction algorithm on the reconstruction.

II. Methods

We sample 2D US slices from a database of previously acquired 3D US volumes at varying angular displacements. Similar to Prevost et al. [5], we use CNNs to predict the angle between 2D US images. The training set consists of neighboring slices with varying known displacements. Additionally, we also assess the effect of adding pre-computed OF displacement fields between two neighboring slices as inputs

[1] Note that this paper is a slightly extended version of a CAOS International paper submission.

to the network. After training the network, a test set collected from a separate set of patients is used for evaluation.

A. Dataset

Our dataset consists of 1403 3D US volumes of the hip from 118 newborns, collected using an Ultrasonix 4DL14-5/38 probe at 7.5MHz with an image depth of 4cm. Each volume is composed of 123 B-mode 2D images collected at regular intervals, spanning an overall angle of 30.1° corresponding to a center-frame sector length of 53.1mm (see Figure 1A). This dataset was collected by several graduate students over the course of 5 years and was approved by the UBC Clinical Research Ethics Board, with certificate numbers: H14-01448, H18-00131, and H18-02024. Infants between the age of 0-6 months who were suspected of having or who were diagnosed with DDH were scanned by US radiologists at BC Children Hospital; the scans were included in this dataset if they lacked any other congenital hip abnormalities. Both healthy and dysplastic infant hips were included in this dataset.

B. Network

We implemented the CNN network from Prevost et al. [5], but modified the final fully-connected layer to output only one displacement value corresponding to an angular rotation (see Figure 1B). We also implemented a deeper CNN with 7 down-sampling layers of the form Conv-BatchNorm-LeakyReLu. The first 2D convolution used 256 filters and on every down-sample we reduced the number of filters by a factor of 2. The Farneback algorithm [10] was also used to generate an OF image: a pixel-wise displacement field formed from two input images. The deeper CNN was trained and tested with the OF image set as an additional input.

C. Training & Testing

We divided the participants into training, validation, and test sets using a ratio of 70:10:20 (968 volumes from 82 participants for training, 143 volumes from 11 participants for validation, and 292 volumes from 25 participants for testing). Each volume was resampled to obtain 128 isotropic 2D slices

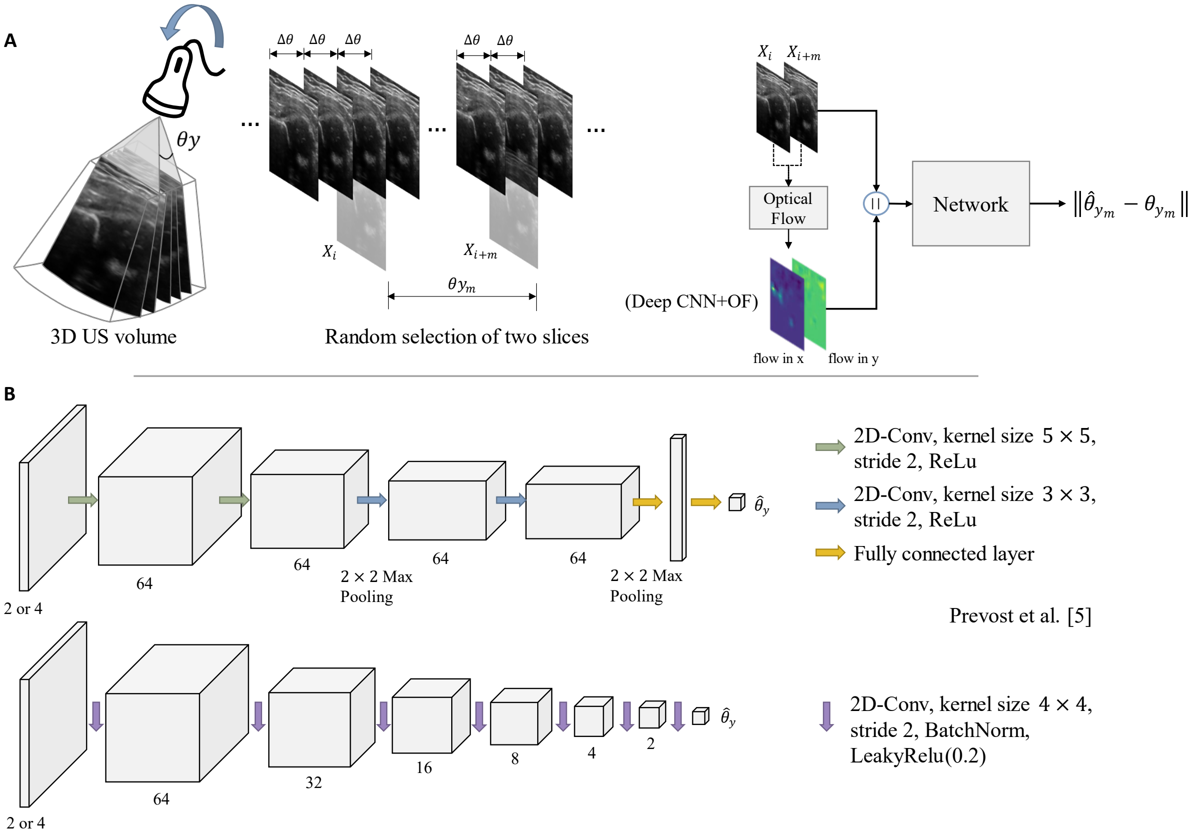
A
θy
$\Delta\theta$ $\Delta\theta$ $\Delta\theta$
$\Delta\theta$ $\Delta\theta$
X_i
X_{i+m}
θy_m
3D US volume
Random selection of two slices
X_i X_{i+m}
Optical Flow
(Deep CNN+OF)
flow in x
flow in y
||
Network
$\|\hat{\theta}_{y_m} - \theta_{y_m}\|$
B
2 or 4
64
64
2 × 2 Max Pooling
64
64
2 × 2 Max Pooling
$\hat{\theta}_y$
2D-Conv, kernel size 5 × 5, stride 2, ReLu
2D-Conv, kernel size 3 × 3, stride 2, ReLu
Fully connected layer
Prevost et al. [5]
2 or 4
64
32
16
8
4
2
$\hat{\theta}_y$
2D-Conv, kernel size 4 × 4, stride 2, BatchNorm, LeakyRelu(0.2)

Fig 1 A. Example of 3D US volume dataset. Resampled volumes consist of 128 US slices acquired at regular intervals of $\Delta\theta = 0.235°$. For each training step, slices X_i and X_{i+m} are sampled randomly from every volume and inputted into the network. B. (Top) Block representation of Prevost et al. [5] network with one output channel. (Bottom) Deeper CNN architecture.

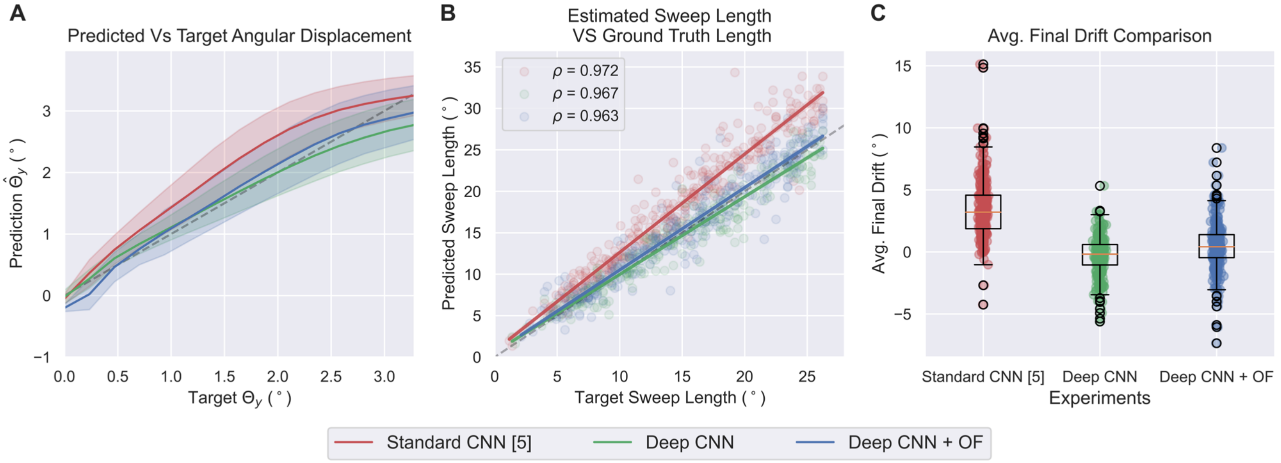

Fig 2 A: Average predicted angle compared to the target angle, where the shaded region represents one standard deviation, and the black dotted line represents $\hat{\theta}_y = \theta_y$. B: Comparison between estimated and ground truth sweep lengths. C: Average final drift error, computed for trajectories ranging between 2˚ and 27˚.

of size 128×128. As an input to the network, two slices separated by an angle in the range of [0.235˚, 3.5˚] were selected at random.

For the Prevost network, we trained with the AdaGrad optimizer, using a learning rate of 0.1 and a batch size of 128. For the deeper CNNs, we implemented the Adam optimizer with a batch size of 32 and a learning rate of 0.03. In all cases, the objective was to minimize the L2 norm between the predicted and known angular steps.

D. Evaluation Metrics

For each volume in the test set, we selected a random slice X_0 and used the trained network to predict the angular separation between X_0 and the 15 subsequent slices. We also evaluated the normalized final drift error (the relative difference between the first and last slices in a sequence normalized by the total trajectory).

III. Results

Figure 2A displays the predicted angular displacement plotted against the actual displacement, for all three network configurations. We calculate the average absolute error as the average absolute difference between the mean prediction and the ground truth angle (dotted black line) across the range of 0-3˚. The deeper CNN with and without OF results in the lowest average absolute error (both at ~0.02˚), followed by the network proposed by Prevost at 0.58˚.

Figure 2B shows predicted angular travel for trajectories ranging between 2˚ and 27˚; all plots show a strong correlation between the predicted and target sweep length. The deeper CNN produced the lowest normalized final drift errors: -0.8% ± 13.2%, compared with 25.3% ± 14.7% for the Prevost network and 4.7 ± 16.5% for the deeper CNN+OF (see Figure 2C). A one-way repeated measures ANOVA showed that there is a statistically significant difference in the average final drift error when comparing the three network configurations (F(48,2) = 22.48, p < .001). Post-hoc pairwise t-tests showed a meaningful decrease in final drift error when implementing the deeper CNN, with and without OF as opposed to the Prevost network (adjusted p = .01 and p = .003 respectively). Including OF when training the deeper CNN network, however, did not have a significant impact on drift error (adjusted p = .63).

IV. Conclusions

We implemented deep learning methods to estimate the relative locations of the different frames in B-mode image sequences of infant hips and found that we could estimate the angular displacements with an average absolute error of 0.02˚. Prevost et al [5] evaluated transformations in 6 degrees of freedom, involving frame-to-frame speed variations below 1mm/frame. Here we show that it is possible to accurately predict center-frame distances of up to 5.3mm. The use of strided convolutions with learnable parameters for downsampling may provide higher flexibility for the local aggregation of information [11]. In our case, OF might not have proven beneficial due to the large separation between slices. Luo et al [9] achieved a final drift error of 5.4% ± 3.0% on their DDH dataset, testing on likely 2 patients; our networks were tested on 25 patients and achieved a drift error of -0.8% ± 13.2% when considering only angular displacement. In the future we expect to extend to 6 degrees of freedom for a more accurate comparison, as well as to experiment with other deep learning algorithms to reduce the standard deviation in the drift error. We intend to also compare the dysplasia metrics evaluated using the reconstructed 3D volumes to those extracted from the reference 3D volumes acquired with a standard 3D US probe to evaluate the variability introduced by the reconstruction process. If the errors are sufficiently low, this technique could enable clinicians to make markedly more repeatable dysplasia metric measurements using widely available 2D US probes.

Acknowledgment

The authors would like to acknowledge the Centre for Aging SMART and ICICS (Institute for Computing Information and Cognitive Systems) for their support during this research.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. M. D. Sewell, K. Rosendahl, and D. M. Eastwood, ‘Developmental dysplasia of the hip’, *BMJ*, vol. 339, Nov. 2009, doi: 10.1136/bmj.b4454.
2. M. Imrie, V. Scott, P. Stearns, T. Bastrom, and S. J. Mubarak, ‘Is ultrasound screening for DDH in babies born breech sufficient?’, J. Child. Orthop., vol. 4, no. 1, pp. 3–8, Feb. 2010, doi: 10.1007/s11832-009-0217-2.
3. N. Quader, ‘Automatic characterization of developmental dysplasia of the hip in infants using ultrasound imaging’, University of British Columbia, 2018. doi: 10.14288/1.0364129.
4. O. V. Solberg, F. Lindseth, H. Torp, R. E. Blake, and T. A. Nagelhus Hernes, ‘Freehand 3D Ultrasound Reconstruction Algorithms—A Review’, *Ultrasound Med. Biol.*, vol. 33, no. 7, pp. 991–1009, Jul. 2007, doi: 10.1016/j.ultrasmedbio.2007.02.015.
5. R. Prevost *et al.*, ‘3D freehand ultrasound without external tracking using deep learning’, *Med. Image Anal.*, vol. 48, pp. 187–202, Aug. 2018, doi: 10.1016/j.media.2018.06.003.
6. M. Luo, X. Yang, H. Wang, L. Du, and D. Ni, ‘Deep Motion Network for Freehand 3D Ultrasound Reconstruction’, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, Sept. 2022, pp. 290–299. doi: 10.1007/978-3-031-16440-8_28.
7. K. Miura, K. Ito, T. Aoki, J. Ohmiya, and S. Kondo, ‘Pose Estimation of 2D Ultrasound Probe from Ultrasound Image Sequences Using CNN and RNN’, in *Simplifying Medical Ultrasound*, 2021, pp. 96–105. doi: 10.1007/978-3-030-87583-1_10.
8. H. Guo, S. Xu, B. Wood, and P. Yan, ‘Sensorless Freehand 3D Ultrasound Reconstruction via Deep Contextual Learning’, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Sept. 2020, pp. 463–472. doi: 10.1007/978-3-030-59716-0_44.
9. M. Luo *et al.*, ‘Self Context and Shape Prior for Sensorless Freehand 3D Ultrasound Reconstruction’, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Sept. 2021, pp. 201–210. doi: 10.1007/978-3-030-87231-1_20.
10. G. Farnebäck, ‘Two-Frame Motion Estimation Based on Polynomial Expansion’, in *Image Analysis*, Berlin, Heidelberg, Jun. 2003, pp. 363–370. doi: 10.1007/3-540-45103-X_50.
11. R. Ayachi, M. Afif, Y. Said, and M. Atri, ‘Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks’, in *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT’18),* Jul. 2019, pp. 234–243. doi: 10.1007/978-3-030-21005-2_23.