Improving 3D Ultrasound Scan Adequacy Classification Using a Three-Slice Convolutional Neural Network Architecture

Olivia Paserin[1*], Kishore Mulpuri[2], Anthony Cooper[2], Rafeef Abugharbieh[1] and Antony J Hodgson[3]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

[2] Department of Orthopaedic Surgery, British Columbia Children's Hospital, Vancouver, BC, Canada

[3] Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada
opaserin@ece.ubc.ca

Abstract

Accurately detecting and diagnosing developmental dysplasia of the hip (DDH), a common hip instability condition among infants, requires ultrasound (US) image data that captures the relevant anatomical structures. Our group has recently introduced a technique for automatically processing 3D US scans of the neonatal hip that significantly reduces dysplasia metric measurement variability, but it can be challenging for a clinician to know at the time of acquisition if they have acquired a volume suitable for analysis. We have previously introduced a single-slice-based approach for assessing adequacy of acquired volumes that worked reasonably well, but we believe that more explicitly three-dimensional approaches would be more robust and reliable. Here, we propose a new technique based on a convolutional neural network (CNN) architecture that incorporates inter-slice information and transfer learning. Our classifier labels volumes as adequate or inadequate for subsequent interpretation based on detecting the presence of key hip anatomical structures needed for DDH diagnosis. We validate our approach on 40 datasets from 15 pediatric patients and demonstrate a slice classification rate of 93% (improving on our previous implementation by 3%) with average processing time of 2 seconds per US volume. We expect automatic US scan adequacy assessment to have significant clinical impact with the potential to help in imaging standardization, improving efficiency of measuring DDH metrics, and improving accuracy of clinical decision making.

1   Introduction

Accurately detecting and diagnosing developmental dysplasia of the hip (DDH), a common hip instability condition among infants, requires ultrasound (US) image data that captures the relevant anatomical structures. Obtaining adequate quality acquisitions and interpreting them correctly is a challenging task and it is difficult even for experienced clinicians to be sure they have acquired appropriate images. For example, when the quality of hip sonograms across 8 German states were tested in 2011, up to 43% of hip sonographers had their licenses revoked because they could not demonstrate sufficient adherence to the imaging guidelines (Tschauner 2012).

Typically, US images are acquired manually by an experienced clinician as they search for key anatomical structures required to make diagnostic measurements. Once judged to be adequate, the US data are saved during the patient's physical exam for later detailed analysis in which DDH metrics are extracted.

Our group has recently shown that using 3D US scans can markedly improve the reliability of dysplasia metric measurements compared to 2D US as volumetric scans capture the entire hip joint and are less prone to probe orientation errors compared to 2D scans (Jaremko 2014). However, the analysis process takes close to three minutes, which slows clinical workflow. To improve clinical usability, we would like to provide rapid assurance at the time of acquisition that the image acquired is suitable for analysis.

Previous work (Quader 2017) on automatic 2D US scan adequacy for DDH required about 1 second per image, which is suitable for assessing 2D acquisitions, but the ~200 slices in a 3D volume would require about 3 minutes to process, so this approach would not be clinically useful. Our group recently developed a fast approach for automatic 3D US scan adequacy (Paserin 2017), which worked reasonably well with a slice classification accuracy of 90%, but did not make use of inter-slice information within a volume. In this study, our goal is to improve on our previous implementation by incorporating inter-slice information and optimise the training of a deep neural network despite a relatively small dataset through the use of transfer learning.

2   Methods

We propose an automatic, near real-time technique for assessing 3D US scan adequacy for DDH in clinical settings using a convolutional neural network (CNN) structured to process inter-slice information and trained using a transfer learning approach to classify US volumes as adequate or inadequate for diagnostic interpretation.

2.1   Materials and Experimental Setup

The data used in this study was comprised of 40 3D B-mode US volumes acquired by two attending pediatric orthopaedic surgeons from 15 pediatric patients (obtained as part of routine clinical care under appropriate institutional review board approval using a SonixTouch Q+ scanner with a 4DL14-5/38 transducer set at 7MHz). The lead author labelled over 3,000 slices from the data and maintained a balanced training set between adequate and inadequate samples.

2.2   CNN Architecture

We implemented a CNN architecture called Squeezenet (Iandola, 2016) to classify 2D slices as adequate or inadequate based on the presence of anatomical features necessary to extract DDH

metrics, including the ilium, acetabulum, labrum, ischium and entire femoral head. The network receives a three channel 227x227 input image with the target slice in the second channel and the two adjacent slices in the first and last channels; it assigns a binary label, adequate or inadequate, to each set of three input slices. As in (Paserin 2017), an adequate 3D scan is defined as one that contains adequate frames spanning a minimum of 5mm (approximately half the typical diameter of the femoral head) with 2mm of inadequate slices preceding and following the adequate slices such that the hip joint is fully captured.
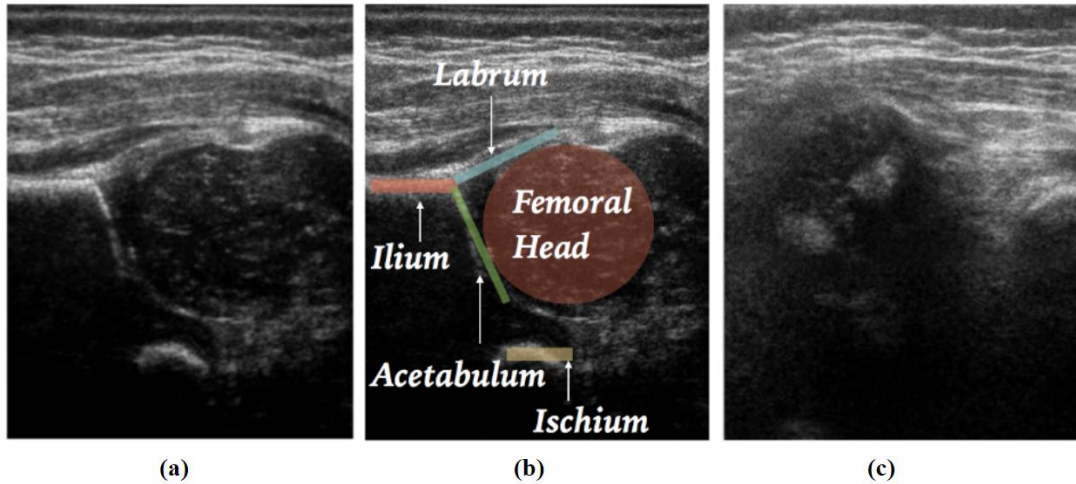

Figure 1: 2D slice examples of **a)** an adequate US slice, **b)** the slice in (a) with the relevant anatomy needed for adequacy annotated, and **c)** an inadequate US slice.

2.3   Network Training

We split our data into a validation set containing 20% of the subjects and used the remaining 80% for training. We employed transfer learning by retraining only the final convolutional layer of the network and performed 5-fold cross validation taking care to avoid mixing images from the same patient between training and validation groups.

3   Results

Classification accuracy: To assess the performance of our classifier, we tested on a dataset of 20 US volumes (10 adequate, 10 inadequate) from five new patients. We achieved 93% accuracy in single frame classification. Since there is currently no gold standard for clinical classification of US volumes, we confirmed that 100% of the volumes classified as adequate by our network were in fact adequate for diagnostic interpretation by calculating 3D dysplasia metrics using the only automatic extraction method available (Quader 2016). The same extraction method was not able to localise the anatomy required for dysplasia metrics in the volumes classified as inadequate by our network.
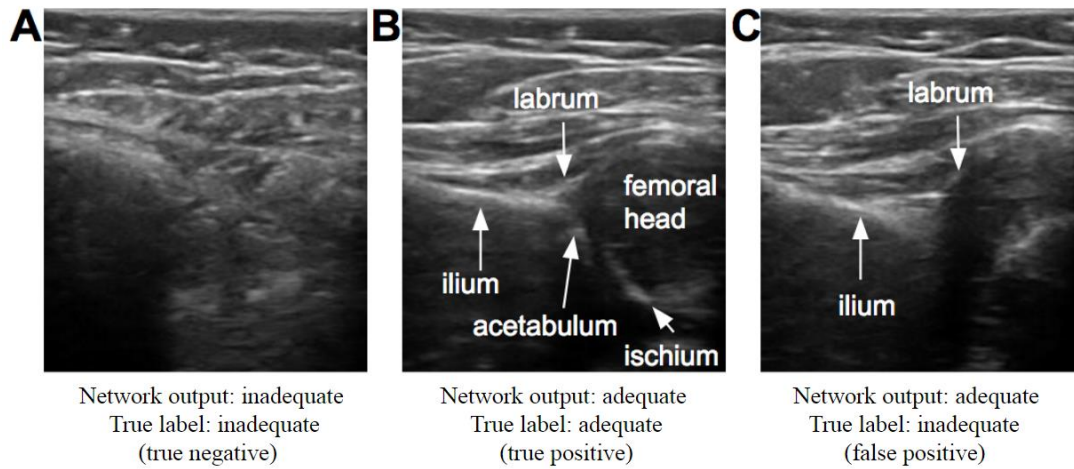
Figure 2: Three test images with their respective network output labels and true labels. Example A shows a true negative case where the network correctly identified an inadequate image. Example B shows a false positive case where the network incorrectly classified an adequate image. Example C shows a false positive case where the network incorrectly classified an inadequate image as adequate, where the image contains the ilium and labrum but no acetabulum or ischium.

Runtime: Our method operates in near real-time, requiring only 2 seconds for classifying one volume.

4 Discussion

In this study, we proposed a CNN-based automated method for classification of 3D US scan adequacy for application to DDH screening and diagnosis. As a result of transfer learning and incorporation of inter-slice information, our implementation performs better than our previous techniques (Paserin 2017) with a 3% increase in accuracy. Our network design achieves a very fast processing time (Quader 2017, Paserin 2017) that is compatible with a clinical workflow. This promising approach may improve acquisitions and assist operators in volume interpretation. In future work, we plan to expand the size of our training set and combine our proposed scan adequacy module with automatic DDH metric extraction to design and test a fully automatic DDH assessment system.

References

Tschauner C, Matthissen H, Hip Sonography with Graf-method in Newborns: Checklists help to avoid mistakes. OUB Journal, pp 7-8, 2012.

Jaremko J, Mabee M, Swami V, Jamieson L, Chow K, Thompson R, Potential for Change in US Diagnosis of Hip Dysplasia Solely Caused by Changes in Probe Orientation: Patterns of Alpha-angle Variation Revealed by Using Three-dimensional US. Radiology 273(3), pp 870-878, 2014.

Quader N, Hodgson A, Mulpuri K, Schaeffer E, Abugharbieh R, Automatic Evaluation of Scan Adequacy and Dysplasia Metrics in 2-D Ultrasound Images of the Neonatal Hip. Ultrasound in Medicine and Biology, 43(6), pp 1252-126, 2017.

Paserin O, Mulpuri K, Cooper A, Hodgson A J, Abugharbieh R, Automatic Near Real-Time Evaluation of 3D Ultrasound Scan Adequacy for Developmental Dysplasia of the Hip, MICCAI Workshop on Clinical Image-Based Procedures (CLIP), pp 124-132, 2017.

Iandola F N, Moskewicz M W, Ashraf K, Han S, Dally W J, Keutzer K, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. arXiv preprint arXiv:1602.07360, pp 46-67, 2016.

Quader N, Hodgson A, Mulpuri K, Cooper A, Abugharbieh R, Towards Reliable Automatic Characterization of the Neonatal Hip Dysplasia from 3D Ultrasound Images. Medical Image Computing and Computer Assisted Interventions, 2016.