



Segmentation of measurable images from standard plane of Graf hip ultrasonograms based on Mask Region-Based Convolutional Neural Network

Aysun Sezer¹, Hasan Basri Sezer, MD²

¹Biruni University, Computer Engineering, Istanbul, Türkiye

²Orthopaedics and Traumatology Clinic, Clinique Du Sport Paris, Paris, France

Graf's ultrasonography (US) method is one of the most commonly used imaging techniques for developmental dysplasia of the hip (DDH). It involves an anatomical description, as well as measuring bony and soft tissue coverage in coronal two-dimensional (2D) US images of the hip. After capturing the image, the physician has to decide whether it is in the standard plane or not and the measurability of the image. An image has to contain a straight iliac wing, lower limb of the ilium, and the labrum for it to be classified as measurable. Graf's method is prone to performer and interpretation variability due to the anatomical complexity of the hip structures, which can lead to false classification.^[1] When the anatomical regions are not exactly identified, selection of points for angle calculations may not be expected to be exactly determined, and the image is not accepted for measurements.^[2]

Received: July 11, 2023

Accepted: July 29, 2023

Published online: September 20, 2023

Correspondence: Aysun Sezer, Biruni Üniversitesi, Bilgisayar Mühendisliği, 34010 Zeytinburnu, İstanbul, Türkiye.

E-mail: sezeraysun@gmail.com

Doi: 10.52312/jdrs.2023.1308

Citation: Sezer A, Sezer HB. Segmentation of measurable images from standard plane of Graf hip ultrasonograms based on Mask Region-Based Convolutional Neural Network. *Jt Dis Relat Surg* 2023;34(3):590-597. doi: 10.52312/jdrs.2023.1308.

©2023 All right reserved by the Turkish Joint Diseases Foundation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (<http://creativecommons.org/licenses/by-nc/4.0/>).

ABSTRACT

Objectives: This study proposed a Mask Region-Based Convolutional Neural Network (R-CNN)-based automatic segmentation to accurately detect the measurable standard plane of Graf hip ultrasonography images via segmentation of the labrum, lower limb of ilium, and the iliac wing.

Patients and methods: The study examined the hip ultrasonograms of 675 infants (205 males, 470 females; mean age: 7 ± 2.8 weeks; range, 3 to 20 weeks) recorded between January 2011 and January 2018. The standard plane newborn hip ultrasound images were classified according to Graf's method by an experienced ultrasonographer. The hips were grouped as type 1, type 2a, type 2b, and type 2c-D. Two hundred seventy-five ultrasonograms were utilized as training data, 30 were validation data, and 370 were test data. The three anatomical regions were simultaneously segmented by Mask-R CNN in the test data and defective ultrasonograms. Automatic instance-based segmentation results were compared with the manual segmentation results of an experienced orthopedic expert. Success rates were calculated using Dice and mean average precision (mAP) metrics.

Results: Of these, 447 Graf type 1, 175 type 2a or 2b, 53 type 2c and D ultrasonograms were utilized. Average success rates with respect to hip types in the whole data were 96.95 and 96.96% according to Dice and mAP methods, respectively. Average success rates with respect to anatomical regions were 97.20 and 97.35% according to Dice and mAP methods, respectively. The highest average success rates were for type 1 hips, with 98.46 and 98.73%, and the iliac wing, with 98.25 and 98.86%, according to Dice and mAP methods, respectively.

Conclusion: Mask R-CNN is a robust instance-based method in the segmentation of Graf hip ultrasonograms to delineate the standard plane. The proposed method revealed high success in each type of hip for each anatomic region.

Keywords: Deep learning, developmental dysplasia of the hip, Mask R-CNN, ultrasound.

The diagnostic power of hip US attracts researchers to develop computer-aided diagnosis (CAD) systems to improve the diagnostic efficacy

and objectivity of US in DDH. Hareendranathan et al.^[3] demonstrated the impact of US image quality on the success of the artificial intelligence-based diagnosis. They proposed a scale and a cut-off point of 7/10 to improve the diagnostic capability of the system. Paserin et al.^[4] proposed a recurrent neural network (RNN)-based method to increase three-dimensional (3D) scan accuracy. Liu et al.^[5] proposed a feature fusion attention network to decrease the interpretation errors on fifty 2D ultrasonograms. El-Hariri et al.^[6] proposed a volumetric measuring method to detect key anatomical structures from 3D hip ultrasonograms. Quader et al.^[7] proposed a random forest-based learning technique and a formula to calculate 3D dysplasia metrics. These studies suggest that there is no perfect segmentation method and that there is a need for more precise segmentation of anatomical areas for a CAD system to be helpful in the clinical field.

Recently, Chen et al.^[8] proposed the YOLOv3-tiny method to automatically segment 2D images statically and dynamically and a scoring system to evaluate the success of the method in finding the measurable images in type 1 and type 2 hips. They used a semantic segmentation approach and reported 1.71° deviation of alpha and 2.40° deviation of beta angles. Lee et al.^[9] developed a Mask R-CNN-based deep learning model for DDH diagnosis using a segmentation and key-point multidetection approach other than Graf's method with an overall success rate of 84.37%.

Semantic segmentation is a process that assigns each pixel in an image to either the foreground or the background without differentiating between instances of the same class. The goal of semantic methods as Yolo versions, YOLACT and SSD is to create regions of interest in the form of rectangles that contain the object without delineating the real target borders and with a confidence interval surrounding it. On the other hand, instance-level based segmentation (e.g., Mask R-CNN) focuses on target boundary detection, where various objects in an image are identified. This allows a precise localization of objects and pixel-level accuracy, enabling a more detailed analysis and understanding of the scene.^[10,11]

A complete CAD system should be capable of deciding the measurable standard plane. Therefore, precise recognition of basic anatomical structures is indispensable, followed by angular measurements and, finally, the classification. Sezer et al.^[12,13] proposed an automatic CAD system based on CNN that was capable of angular measurements

and another one free from angle measurements, demonstrating the high capacity of the deep learning methods to classify hip ultrasonograms according to Graf's classification system, with 93.90 and 97.70% accuracy, respectively. Moreover, they successfully enhanced the classification accuracy rate by reducing the presence of speckle noise inherent in US images.

This study proposed a Mask R-CNN-based automatic segmentation to accurately detect the measurable standard plane of Graf hip US images via segmentation of the labrum, lower limb of the ilium, and the iliac wing. The hypothesis is that the instance-based method would accurately and simultaneously delineate the three key anatomical structures to augment the reliability and consistency of DDH assessments.

PATIENTS AND METHODS

The study was conducted at the Şişli Hamidiye Etfal Training and Research Hospital with hip US images of DDH patients recorded between January 2011 and January 2018. The images of newborns aged 0 to 6 months that met Graf's standard plane criteria were considered for inclusion. Images with probe tilt and rotation artifacts were excluded. The study population consisted of both normal and affected hips at various stages of DDH. Ultrasound images of 447 healthy infants and 228 infants with DDH were included in the analysis for a total of 675 participants (205 males, 470 females; mean age: 7±2.8 weeks; range, 3 to 20 weeks).

An orthopedic expert with over 15 years of experience in using Graf's hip US collected and evaluated the dataset. The hip US images were recorded using a 7.5 MHz linear-array probe and the Logiq 200 US device (GE Healthcare Inc., Chicago, IL, USA), which has a 60-mm field of view. The images were captured in a vertical orientation and in the right hip projection, following Graf's^[14] guidelines. Each image had a size of 640×480 pixels and was in the right hip coronal projection. A suitable image was selected for the evaluation, ensuring that the necessary anatomical structures were visible, and the correct orientation was maintained for accurate angle calculations. All images were annotated using the LabelMe tool^[15] by an experienced orthopedist (Figure 1).

A CAD system based on deep learning was developed to delineate the three anatomical structures of newborn hips. In the light of variations of high inter- and intraobserver reliability, the proposed

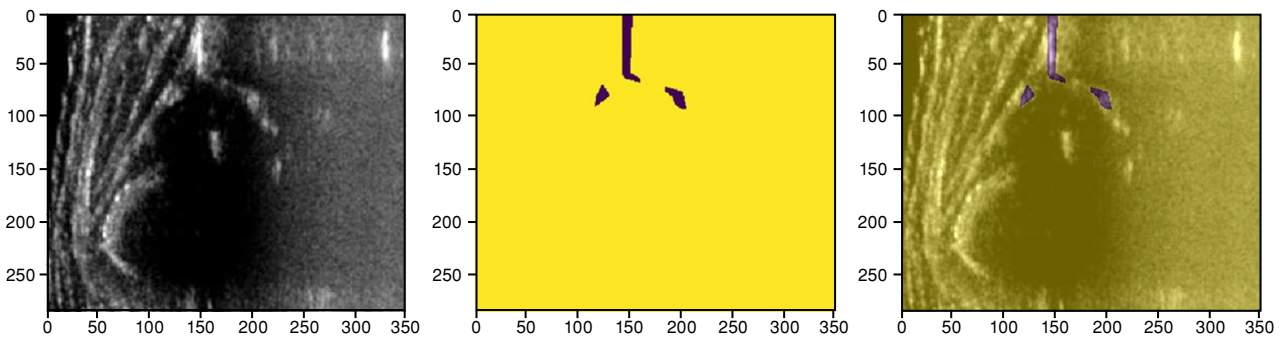


FIGURE 1. Demonstration of (a) original US, (b) manual annotation of tree anatomical regions on the original image, and (c) extracted manually segmented regions.

US: Ultrasonography.

CAD system aimed to overcome these challenges by enabling multidetection and enhancing the accuracy of DDH diagnosis.

Mask R-CNN

A total of 675 US images were gathered and annotated for segmentation purposes. Out of these, 275 images were randomly selected for training, 30 were picked for validation, and the remaining 370 were chosen for testing. The implementation of Mask R-CNN was done using Keras and Tensorflow, and the training process was performed on an NVIDIA V100 graphics processing unit (Nvidia Corp., Santa Clara, CA, USA) for 1000 epochs. The Adam optimizer was utilized with a learning rate of 0.001.

To evaluate the quality of the predicted segmentation masks, we employed the Dice

coefficient and mean average precision (mAP) score. The Dice coefficient measures the similarity between the predicted mask and the ground truth mask. Additionally, the mAP score is calculated as the average precision under the precision-recall curve, averaged for each object in an image at various recall levels. The precision-recall curves and average precision scores for all 370 images in the test dataset were averaged. Segmentation performance was calculated by using Dice and mAP metrics for each hip class and each region of interest (ROI).

All instances were segmented by the proposed method for each class and each ROI (Figure 2). The accuracy of anatomical region segmentation in US images was assessed by calculating the intersection over union (IOU) for an IOU threshold of 0.5. Intersection over union measures the overlap

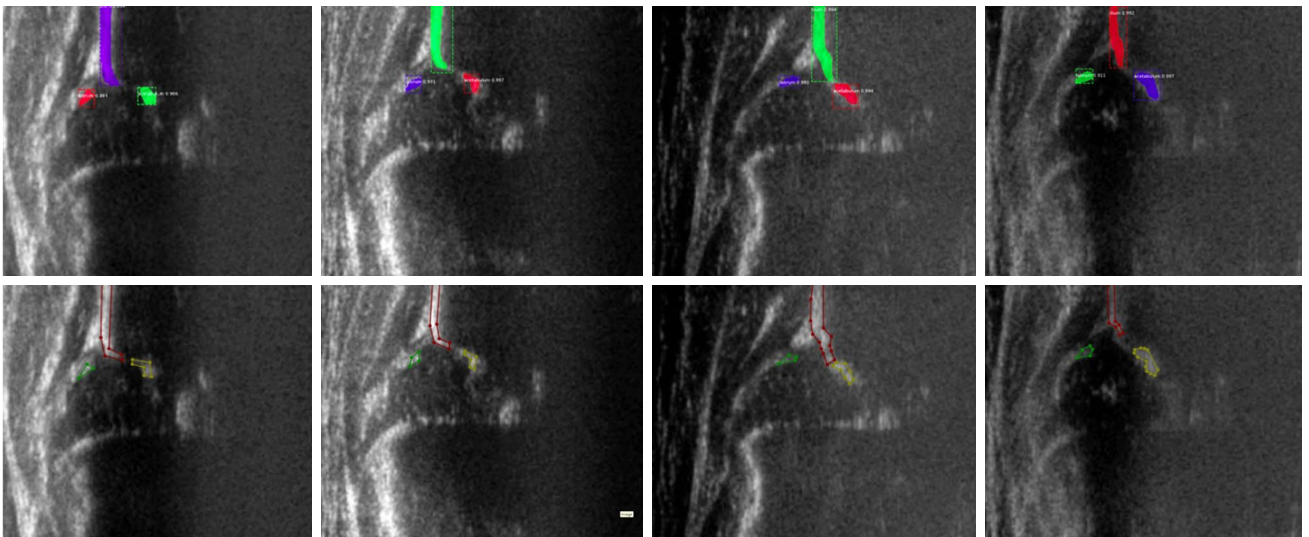


FIGURE 2. The first row demonstrates segmentation results of the iliac wings, labrum, and acetabulum based on Mask R-CNN for type 1 hips, type 2a-b hips, type 2c and D, respectively, and the second row demonstrates their manual segmentation by an expert.

R-CNN: Mask Region-Based Convolutional Neural Network.

between the predicted bounding box or mask and the ground-truth bounding box or mask. It is calculated as the ratio of the area of intersection between the two regions to the area of their union. For a prediction to be considered correct (true positive), its IOU with the corresponding ground-truth region needed to be ≥ 0.5 . Conversely, predictions with an IOU < 0.5 were classified as false positives. This evaluation was performed by comparing the predicted segmentation masks against the manual delineation provided by an expert.

Deep learning-based methods have demonstrated high performance in the field of object detection and segmentation. These methods can be broadly categorized into one-stage and two-stage approaches.

In one-stage object detection methods, a single convolutional neural network is used for both localization and classification tasks. This approach allows for object detection to be performed in a single step, resulting in a simpler and faster model architecture, such as YOLO versions, EfficientDet, and RetinaNet.

Two-stage object detection methods, also known as region-based methods, consist of two stages: region proposal and classification. In the region proposal stage, potential regions of interest are generated, while in the classification stage, these regions are assigned with labels. The specific technique for region proposal can vary depending on the method employed. For example, R-CNN and Fast R-CNN utilize the selective search technique, while Faster R-CNN employs a Region Proposal Network to accelerate this process. During the region proposal stage, the input image is evaluated, and potential regions are identified based on this evaluation. In the second stage, a convolutional neural network is used to classify the proposed regions.^[10,11]

In summary, instance-level-based segmentation relies on target detection, where diverse objects within an image are identified and categorized. To further enhance the capabilities of pixel-level detection, an extension to R-CNN called Mask R-CNN was introduced. Mask R-CNN enables instance-based segmentation by providing both class labels and pixel-wise masks. It effectively addresses the challenge of class imbalance and demonstrates robustness in simultaneously segmenting multiple instances within an image, regardless of their scale and orientation.

Unlike Faster R-CNN, which primarily focuses on classification and bounding box recognition, Mask

R-CNN incorporates an additional mask head. This mask head predicts an object mask for each ROI. Consequently, Mask R-CNN provides an extra output in the form of a binary mask, in addition to class labels and bounding box offsets.

Mask R-CNN has the ability to successfully perform segmentation on images, even with small objects, such as the labrum and acetabulum in hip US, despite the presence of speckle noise. Its performance is significantly higher compared to single-stage approaches (YOLOACT++, YOLO versions, and SparseInst) and semantic based segmentation methods (U-Net, CapsNet, and SegCaps), but it operates at a slower speed.^[9,11,16,17]

RESULTS

To evaluate the quality of the predicted segmentation masks, we employed the Dice coefficient and mean average precision (mAP) score. The Dice coefficient measures the similarity between the predicted mask and the ground truth mask. Additionally, the mAP score is calculated as the average precision under the precision-recall curve, averaged for each object in an image at various recall levels. The precision-recall curves and average precision scores for all 370 images in the test dataset were averaged. Segmentation performance was calculated by using Dice and mAP metrics for each hip class and each ROI.

All instances were segmented by the proposed method for each class and each ROI (Figure 2). The accuracy of anatomical region segmentation in US images was assessed by calculating the intersection over union (IOU) for an IOU threshold of 0.5. Intersection over union measures the overlap between the predicted bounding box or mask and the ground-truth bounding box or mask. It is calculated as the ratio of the area of intersection between the two regions to the area of their union. For a prediction to be considered correct (true positive), its IOU with the corresponding ground-truth region needed to be ≥ 0.5 . Conversely, predictions with an IOU < 0.5 were classified as false positives. This evaluation was performed by comparing the predicted segmentation masks against the manual delineation provided by an expert.

The hip types according to Graf were determined and patients were divided into three groups as type 1, type 2a-b and type 2c-D. Four hundred forty-seven hips were type 1, 175 hips were type 2a-b, and 53 hips were type 2c-D.

TABLE I

Segmentation results of each hip type based on Mask R-CNN with mAP for an IoU of 0.5 and Dice metric

Hip type	α angle	β angle	Total number of dataset	Average success rate (with IoU of 0.5) (%)	Average success rate (Dice metric) (%)
1	$\geq 60^\circ$	-	447	98.46	98.73
2a and 2b	$50^\circ - 59^\circ$	$55^\circ - 77^\circ$	175	96.57	96.85
2c and D	$43^\circ - 49^\circ$	-	53	95.89	96.13

R-CNN: Mask Region-Based Convolutional Neural Network; IoU: Intersection over union.

TABLE II

Segmentation results of each anatomical region (iliac wings, labrum, and acetabulum) based on Mask R-CNN with mAP for an IoU of 0.5 and Dice metric

Anatomical region	Number of test dataset	Average success rate (with IoU of 0.5) (%)	Average success rate (Dice metric) (%)
Iliac wing	370	98.25	98.86
Labrum	370	97.72	98.02
Acetabulum	370	94.91	95.27

R-CNN: Mask Region-Based Convolutional Neural Network; mAP: Mean average precision; IoU: Intersection over union.

Average success rates according to hip types in the whole data were 96.95 and 96.96% according to Dice and mAP methods, respectively. According to the Dice method, the average success rates for type 1, type 2a-b, and type 2c-D hips were 98.46, 96.57, and 95.89%, respectively. According to the mAP method, the average success rates for type 1, type 2a-b, and type 2c-D hips were 98.73, 96.85, and 96.13%, respectively (Table I).

Average success rates with respect to anatomical regions in the whole data were 97.20 and 97.35% according to Dice and mAP methods, respectively. According to the Dice method, average success rates for the iliac wing, labrum, and acetabulum were 98.25, 97.72, and 94.91%, respectively. According to

the mAP method, average success rates for the iliac wing, labrum, and acetabulum were 98.86, 98.02, and 95.27%, respectively (Table II).

DISCUSSION

This study demonstrates the segmentation success of the deep learning in Graf's hip US images to delineate the standard cuts. The three baseline anatomical structures of Graf's standard plane, namely iliac wing, deepest point of the acetabulum, and the labrum, were simultaneously segmented with high success (Figure 2). The proposed Mask R-CNN-based segmentation performed an overall success of 96.95 and 96.96% according to Dice and mAP methods, respectively.

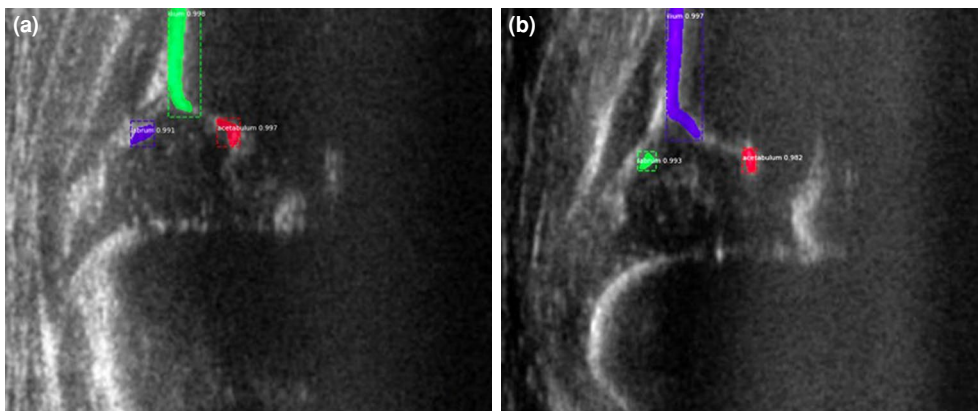


FIGURE 3. Demonstrates correct segmentation in case of (a) anatomically normal and (b) anatomically distorted iliac wings.

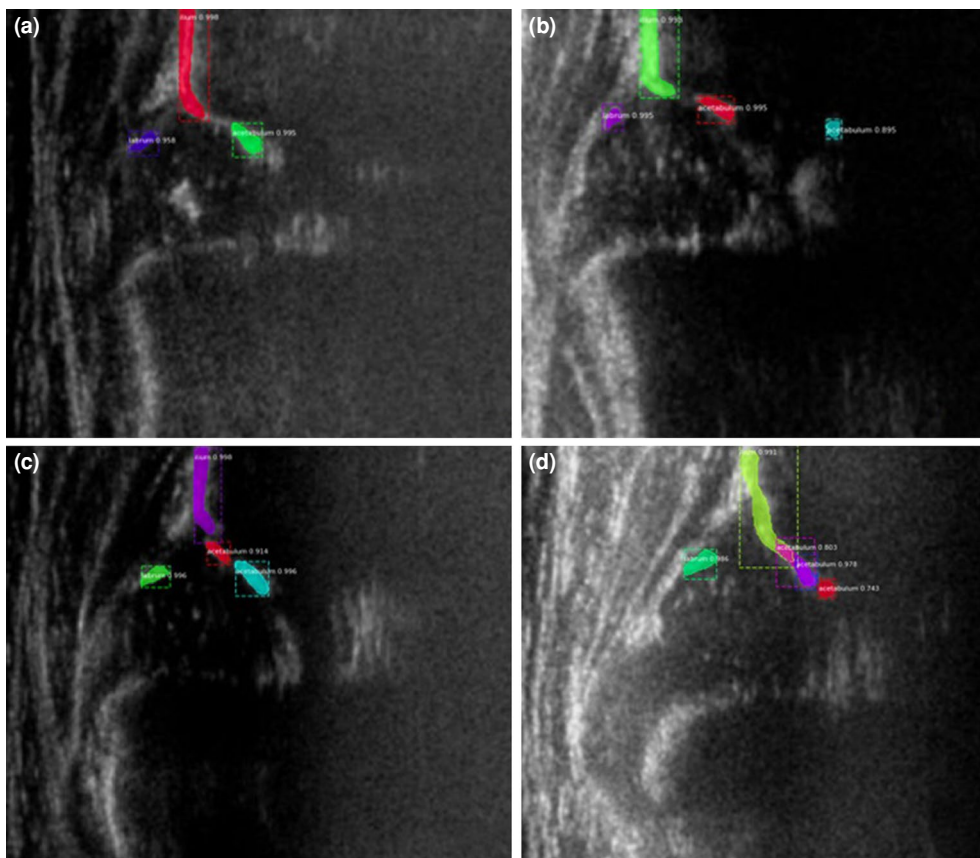


FIGURE 4. Demonstrates different cases of segmentation of the lower limb of the iliac bone. (a) Correct segmentation; (b) false segmentation of lower limb of ilium due to an intrapelvic structure; (c, d) false segmentation with double ROI. ROI: Region of interest.

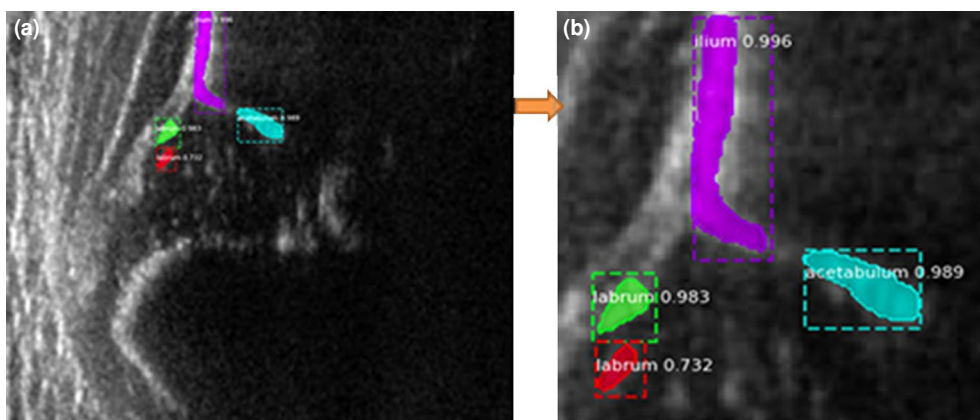


FIGURE 5. (a) Demonstration of only segmentation error of labrum. (b) The venous sinusoid in the femoral head was mistaken for the labrum. A threshold of 73.3% would be sufficient to eliminate this error.

The regional success rate of Mask R-CNN on the US images was 98.86, 98.02, and 95.27% for the segmentation of the iliac wing, labrum, and acetabulum, respectively, with the dice metric,

whereas the regional success rate was 98.25, 97.72, and 94.91% for an IOU of 0.5 according to mAP. Average success rate values measured by both the Dice coefficient and mAP metrics were found to be

very close across different hip types and individual anatomical regions.

The highest achievements in both metrics were observed for type 1 hips and the iliac region (Tables I, II).

According to Graf,^[2] the questionability of Graf's method is mainly due to angle measurement errors, which are mostly due to incorrect determination of the anatomical structures. Determining the turning point of concavity of acetabulum to convexity, the lower limb of the ilium, and labrum are crucial to finding angles. The proposed system performed a very high success rate in newborn hip US images even in the presence of anatomical variations and neighboring confounders (Figure 3). The lower limb of ilium may be mistaken for ligamentum teres and nearby intrapelvic structures (Figure 4). The anatomical regions that are in the vicinity resembling the labrum are the synovial fold, ischiofemoral ligament, and the proximal perichondrium. In this study, the only error in the labrum region was due to a vascular sinusoid in the femoral head rather than other anatomical structures (Figure 5).^[2,12]

Out of 370 test images, the proposed method delineated more than the ROIs in only seven images (Figures 3, 4). Only one image was confounded in the labrum area, and six demonstrated double acetabular region when there was no threshold. In case of the double assignment of a ROI, the correct anatomical region was found at the highest probability. Although the rate of errors was low, this study revealed a threshold in range of 73.3% for the labrum and 91.5% for the lower limb of ilium to eliminate errors (Figures 4, 5). Above these threshold values, the proposed system becomes foolproof and can thus be safely utilized as a guide by clinicians.

This study proposed the use of instance-based segmentation with Mask R-CNN, different from other CAD systems. Different from one-stage methods, Mask R-CNN provides exact borders of an object rather than a rectangular bounding box that preserves a certain gap between the object and the ROI.^[8] Mask R-CNN is a two-stage approach that is more accurate but slower due to its complexity and resulting computational cost.^[10,11]

There is only one study in the literature that employed a two-stage method in hip US images; this study focused on key points and an imaging principle other than Graf's method.^[9] To our knowledge, this is the first study using the Mask R-CNN method to segment three anatomical regions to find the standard plane of Graf.^[18] The segmented images are suitable for angle measurements.

The main limitation of this study is the number of patients in the whole data and in the increased dysplasia group (type 2c or D). This limitation is difficult to overcome due to the rarity of these patients, for which a solution might be through data augmentation.^[12]

In conclusion, Mask R-CNN-based segmentation of anatomical structures of newborn hips from Graf's US images to delineate measurable images provided high success. The segmentation performance stayed in a very narrow range regardless of hip types and anatomical regions. Nonetheless, the highest success rates were acquired in type 1 hips and the iliac region. The proposed algorithm is expected to perform even better when provided with higher resolution and lower noise images. Future research could examine the introduction of precise mobile applications in the field to guide performers.

Ethics Committee Approval: The study protocol was approved by the University of Health Sciences Hamidiye Etfal Training and Research Hospital Clinical Research Ethics Committee (date: 13.12.20216, no: 1344). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: Idea/concept, design, analysis and/or interpretation, literature review, writing the article, critical review, references and fundings, materials: A.S., H.B.S.; Control/supervision: A.S.; Data collection and/or processing: H.B.S.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Atalar H, Arıkan M, Tolunay T, Günay C, Bölükbaşı S. The infants who have mature hip on ultrasonography but have risk factors of developmental dysplasia of the hip are required radiographic examination. *Jt Dis Relat Surg* 2021;32:598-604. doi: 10.52312/jdrs.2021.413.
2. Graf R. Hip sonography: 20 years experience and results. *Hip Int* 2007;17 Suppl 5:S8-14.
3. Hareendranathan AR, Zonoobi D, Mabee M, Cobzas D, Punithakumar K, Noga ML, et al. Toward automatic diagnosis of hip dysplasia from 2D ultrasound. *ISBI* 2017;982-5.
4. Paserin O, Mulpuri K, Cooper A, Hodgson A, Garbi R. (2018). Real time RNN based 3D ultrasound scan adequacy for developmental dysplasia of the hip. *International Conference on Medical Image Computing and Computer-Assisted Intervention* 2018. doi: 10.1007/978-3-030-00928-1_42.

5. Liu R, Liu M, Sheng B, Li H, Li P, Song H, et al. NHBS-Net: A feature fusion attention network for ultrasound neonatal hip bone segmentation. *IEEE Trans Med Imaging* 2021;40:3446-58. doi: 10.1109/TMI.2021.3087857.
6. El-Hariri H, Hodgson AJ, Mulpuri K, Garbi R. Automatically delineating key anatomy in 3-D ultrasound volumes for hip dysplasia screening. *Ultrasound Med Biol* 2021;47:2713-22. doi: 10.1016/j.ultrasmedbio.2021.05.011.
7. Quader N, Hodgson AJ, Mulpuri K, Cooper A, Garbi R. 3-D ultrasound imaging reliability of measuring dysplasia metrics in infants. *Ultrasound Med Biol* 2021;47:139-53. doi: 10.1016/j.ultrasmedbio.2020.08.008.
8. Chen T, Zhang Y, Wang B, Wang J, Cui L, He J, et al. Development of a fully automated graf standard plane and angle evaluation method for infant hip ultrasound scans. *Diagnostics (Basel)* 2022;12:1423. doi: 10.3390/diagnostics12061423.
9. Lee SW, Ye HU, Lee KJ, Jang WY, Lee JH, Hwang SM, et al. Accuracy of new deep learning model-based segmentation and key-point multi-detection method for ultrasonographic developmental dysplasia of the hip (DDH) screening. *Diagnostics (Basel)* 2021;11:1174. doi: 10.3390/diagnostics11071174.
10. Zhang Z, Zhang X, Lin X, Dong L, Zhang S, Zhang X, et al. Ultrasonic diagnosis of breast nodules using modified faster R-CNN. *Ultrason Imaging* 2019;41:353-67. doi: 10.1177/0161734619882683.
11. Bharati P, Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: A survey. *Computational Intelligence in Pattern Recognition* 2020;657-68. doi: 10.1007/978-981-13-9042-5_56.
12. Sezer A, Sezer HB. Deep convolutional neural network-based automatic classification of neonatal hip ultrasound images: A novel data augmentation approach with speckle noise reduction. *Ultrasound Med Biol* 2020;46:735-49. doi: 10.1016/j.ultrasmedbio.2019.09.018.
13. Sezer H, Sezer A. Automatic segmentation and classification of neonatal hips according to Graf's sonographic method: A computer-aided diagnosis system. *Applied Soft Computing* 2019;82. doi: 105516. 10.1016/j.asoc.2019.105516..
14. Graf R. Hip sonography: Diagnosis and management of infant hip dysplasia. Berlin/Heidelberg: Springer; 2006. p. 6-16.
15. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: A database and web-based tool for image annotation. *Int J Comput Vis* 2008;77:157-73. doi: 10.1007/s11263-007-0090-8
16. Viedma IA, Alonso-Caneiro D, Read SA, Collins MJ. OCT retinal and choroidal layer instance segmentation using mask R-CNN. *Sensors (Basel)* 2022;22:2016. doi: 10.3390/s22052016.
17. El Gayar N. A Review of Capsule Networks in Medical Image Analysis. *Proceedings of ANNPR* 2022;13739:65.
18. Atik OŞ. Which articles do the editors prefer to publish? *Jt Dis Relat Surg* 2022;33:1-2. doi: 10.52312/jdrs.2022.57903.