Original Article

Semi-supervised learning in diagnosis of infant hip dysplasia towards multisource ultrasound images

Xuanpeng Li[1]^, Ruixiang Zhang[1], Zhibo Wang[1], Jiakuan Wang[2]

[1]School of Instrument Science and Engineering, Southeast University, Nanjing, China; [2]Department of Orthopedics, Yangzhou Maternal and Child Health Care Service Centre, Yangzhou, China

Contributions: (I) Conception and design: X Li, J Wang; (II) Administrative support: X Li; (III) Provision of study materials or patients: J Wang; (IV) Collection and assembly of data: Z Wang; (V) Data analysis and interpretation: R Zhang, Z Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jiakuan Wang, MD. Associated Professor, Director of Orthopedics, Department of Orthopedics, Yangzhou Maternal and Child Health Care Service Centre, 395 Guoqing Road, Yangzhou 225002, China. Email: jkwangyz@126.com.

Background: Automated diagnosis of infant hip dysplasia is heavily affected by the individual differences among infants and ultrasound machines.

Methods: Hip sonographic images of 493 infants from various ultrasound machines were collected in the Department of Orthopedics in Yangzhou Maternal and Child Health Care Service Centre. Herein, we propose a semi-supervised learning method based on a feature pyramid network (FPN) and a contrastive learning scheme based on a Siamese architecture. A large amount of unlabeled data of ultrasound images was used via the Siamese network in the pre-training step, and then a small amount of annotated data for anatomical structures was adopted to train the model for landmark identification and standard plane recognition. The method was evaluated on our collected dataset.

Results: The method achieved a mean Dice similarity coefficient (DSC) of 0.7873 and a mean Hausdorff distance (HD) of 5.0102 in landmark identification, compared to the model without contrastive learning, which had a mean DSC of 0.7734 and a mean HD of 6.1586. The accuracy, precision, and recall of standard plane recognition were 95.4%, 91.64%, and 94.86%, respectively. The corresponding area under the curve (AUC) was 0.982.

Conclusions: This study proposes a semi-supervised deep learning method following Graf's principle, which can better utilize a large volume of ultrasound images from various devices and infants. This method can identify the landmarks of infant hips more accurately than manual operators, thereby improving the efficiency of diagnosis of infant hip dysplasia.

Keywords: Sonographic image; contrast learning; landmark identification; Semantic segmentation; infants hip dysplasia

Submitted Oct 02, 2023. Accepted for publication Mar 19, 2024. Published online Apr 18, 2024.
doi: 10.21037/qims-23-1384
View this article at: https://dx.doi.org/10.21037/qims-23-1384

^ ORCID: 0000-0001-9320-0658.

Introduction

Developmental dysplasia of the hip (DDH) has an incidence rate of 1.6–28.5%, which is one of the most common musculoskeletal disorders that seriously affects infant health (1). Late diagnoses increase the need for operative intervention and have long term implications for patients and their families (2). Ultrasound, which can penetrate the infant hip and produce different echo strengths, is used to assess its structural abnormalities (3). Moreover, it is radiation-free and cost-effective, making it a primary tool for early DDH diagnosis. Graf introduced a standardized scanning technique for hip sonography examination, involving recognition of standard plane and anatomical structure (4). It categorizes DDH into 4 types and multiple subtypes, providing reliable results with repeatability.

Graf's ultrasound classification is crucial for determining whether infant hips are abnormal. It relies on the standard plane recognition and landmark identification of anatomical structures, which depend on the personal experience of specialized doctors and may lead to measurement variations among doctors.

In recent years, deep neural networks have been explored for their excellent image feature extraction capabilities in ultrasound diagnosis (5-10). Most auxiliary diagnostic methods based on deep learning rely heavily on the pixel-level annotated data (11). To mitigate the dependency of deep learning methods on annotation of various data sources, we introduced contrastive learning techniques based on a Siamese architecture to improve the performance of ultrasound diagnosis among various infants and ultrasound machines.

Methods

Sample selection

A total of 493 infants who underwent hip ultrasound examinations were selected for this study at the Department of Orthopedics of Yangzhou Maternal and Child Health Care Service Centre between 2021 and 2022. The age of these infants ranged from 30 to 90 days. In the experiment, a total of 4,437 hip ultrasound images were collected. Among these, 1,479 images strictly adhered to the requirements of standard plane based on the Graf's method. These images were reviewed and categorized by 6 orthopedists with extensive clinical experience. The study was approved by the Research Ethics Committee of Southeast University and Yangzhou Maternal and Child Health Care Service Centre. The requirement for individual consent for this retrospective analysis was waived. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Our approach

The components employed in this method are illustrated in *Figure 1*. Initially, ultrasound data were collected and partially annotated to construct the dataset. A subset of the dataset was annotated with semantic information of landmarks for anatomical structures. A feature pyramid network (FPN) (12) was pretrained using a contrastive learning approach via a Siamese network. Then, the model with 2 downstream task branches was trained by using the pretrained feature maps and labeled data. The downstream tasks consisted of standard plane recognition and landmarks identification

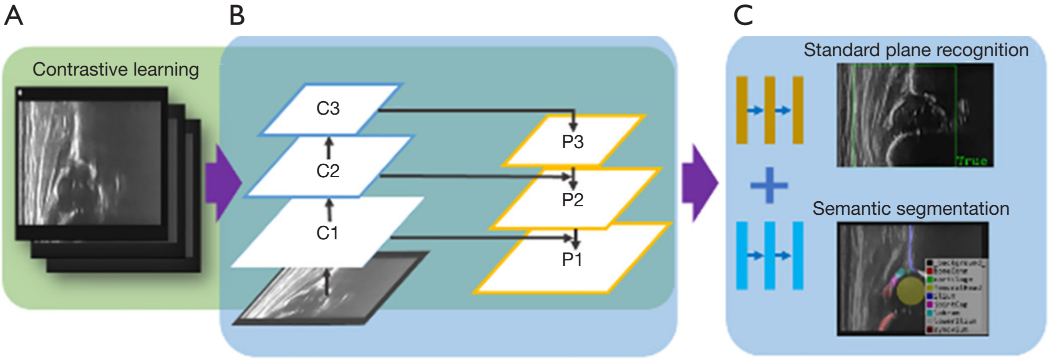

Figure 1 Framework of our proposed method. (A) Contrastive learning; (B) feature pyramid network module; (C) downstream tasks including standard plane recognition and semantic segmentation.

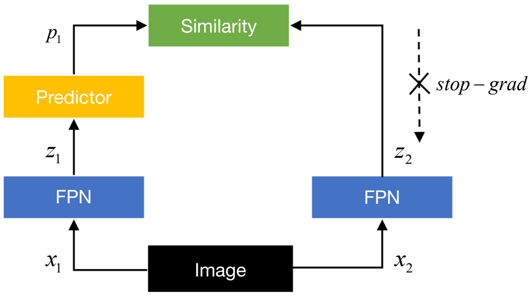

Figure 2 The pretraining process. FPN, feature pyramid network.

based on semantic segmentation. Finally, α and β angles were calculated based on the landmarks. The final results were presented to assist doctors in the diagnosis process.

Contrastive self-supervised learning

In order to fully leverage a large amount of unlabeled data, this approach utilizes contrastive learning for pre-training. Contrastive learning, as a novel machine learning technique, guides the model to learn common features among unlabeled data by teaching it the similarities and dissimilarities between data features. Its advantage lies in its ability to make full use of a large amount of unlabeled data by creating a proxy task where custom pseudo-labels are treated as training signals, and the learned representations can then be applied to downstream tasks (11).

Currently, there are several commonly used strategies of contrastive learning:

SimCLR (13): it generates positive samples using data augmentation and uses other data within the same batch as negative samples.

SwAV (14): it uses clustering to create positive and negative samples, avoiding the need for large batch sizes or MemoryBank.

BYOL (15): it uses BatchNormalization to eliminate the need for searching negative samples during training and prevents training collapse.

SimSiam (16): it employs the expectation-maximization (EM) strategy to control the gradient propagation during training, preventing training collapse more effectively and being easier to implement.

Contrastive learning proxy tasks typically encourage models to bring similar data pairs closer while pushing dissimilar data pairs apart. SimCLR directly uses co-occurring negative samples within the current batch, requiring a large batch size and resulting in significant memory consumption during training. SwAV combines clustering algorithms with neural networks to encode input information for contrast, integrating it into cluster centers, avoiding the use of large batch sizes but significantly increasing computational complexity during training. BYOL does not use negative sample pairs; it directly predicts the output of 1 view from another view of an image. BYOL is essentially a momentum encoder twin network and cannot entirely prevent model training collapse. Compared to the methods above, SimSiam is a kind of Siamese architecture, which directly maximizes the similarity of 1 image's 2 views, using neither negative pairs nor a momentum encoder. It works with typical batch sizes and does not rely on large-batch training (16).

In this study, the contrastive learning scheme based on SimSiam was used for the optimization of feature extractor parameters following these steps:

(I) Considering the following loss function:

$$L(\theta,\eta)=l_2\left[F_\theta(x)-F_\eta(x)\right] \qquad [1]$$

where F is the feature extraction network with network parameters θ; Γ is the data enhancement function; F_η is the optimal parameter network. Therefore, in order for the network to learn the optimal extraction capability, we need to solve θ and η simultaneously to minimize L.

(II) According to the *EM* algorithm, the process of minimizing L follows these 2 alternating steps:

$$\theta^t \leftarrow \arg\min_\theta L\left(\theta,\eta^{t-1}\right) \qquad [2]$$

$$\eta^t \leftarrow \arg\min_\eta L\left(\theta^t,\eta\right) \qquad [3]$$

where t is the index of the alternating solution. When solving θ^t, the parameter η^{t-1} is controlled by stopping gradient descent; when solving η^t, θ^t is the optimal parameters under the current data enhancement method Γ.

At the same time, the predictor was improved based on both the selected backbone network and the output feature map. The convolutional layer in FPN was used to replace the original fully connected layer, so that it could process both feature vectors and feature maps, as shown in *Figure 2*.

The pre-training process was divided into 4 steps: (I) perform data augmentation in different ways to obtain multiple pairs of positive samples x_1,x_2; (II) input x_1,x_2 into the feature extractor and obtain the feature maps z_1,z_2; (III) since the input x_1,x_2 are positive samples, we hope that the information extracted by the feature extractor should be

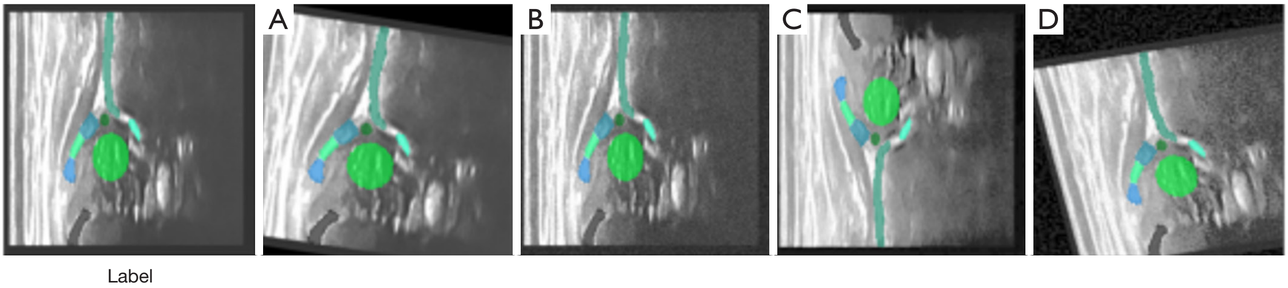

Figure 3 Data augmentation: (A) affine transformation; (B) Gaussian noise and random contrast; (C) image compression and random cropping; (D) stochastic combination.

consistent. Then, p_1 is obtained by the predictor with z_1; (IV) minimize the distance between p_1 and z_1, and estimate the parameters of the feature extractor by controlling the gradient propagation.

A loss function based on cosine similarity was used as:

$$D\left(p_1, z_2\right) = -\frac{p_1 \cdot z_2}{\left\|p_1\right\|_2 \times \left\|z_2\right\|_2} \quad [4]$$

Combined with the EM algorithm, the loss function was defined as:

$$\text{Loss} = D\left(p_1, stopgrad\left(z_2\right)\right) \quad [5]$$

Among them, the encoder on x_1 receives the gradient from p_1 for update, and the encoder on x_2 does not update the parameters from z_2 due to gradient stop. In order to realize the alternate update iteration of the encoder parameters, z_2 is also processed through the predictor to obtain p_2. The loss function is expanded to the following formula to update the encoder parameters on x_2 as follows:

$$\text{Loss} = \frac{1}{2} D\left(p_1, stopgrad\left(z_2\right)\right) + \frac{1}{2} D\left(p_2, stopgrad\left(z_1\right)\right) \quad [6]$$

Data augmentation

In order to improve robustness of model, we further adopted random online data augmentation during the training process as follows: (I) affine transformation: Linear transformation and translation are performed on the original image. It does not change the collinearity of each pixel in the original image and the proportion of the contour. (II) Gaussian noise and random contrast are added to sonographic images. (III) Image compression and random cropping: It simulates the differences in the locations of landmarks in different images. (IV) Stochastic combination: randomly select and combine the above methods. Examples of the original sonographic images and the enhanced sonographic images are shown in *Figure 3*.

Dataset construction

The dataset contained 2,958 images of non-standard plane and 1,479 images of standard plane during the training and testing. Among the standard plane images, there were 400 samples with labels for semantic segmentation.

In order to verify the effect of each module in our method, we used 1,079 unlabeled data in the pre-training step, 320 labeled data as the training set, and 80 labeled data as the test set for the landmark identification. All standard plane images and non-standard plane images are used in the evaluation of standard plane recognition. Data allocation is shown in *Table 1*.

The annotated data contained pixel-level semantic landmark information of anatomical structures. All images and landmarks were subject to strict quality control to ensure the authenticity and reliability of model training. The diagram of anatomical structure annotation is shown in *Figure 4*. In terms of Graf's method, a sonographic image should be decided as a standard plane by the principle that 3 landmarks of anatomical structure should appear simultaneously, including the lower limb of the bony ilium in the depth of the acetabular fossa, the mid portion of the acetabular roof, and the acetabular labrum.

Training

We used ResNeXt-FPN-50 as the backbone network in the experiment (12). The model was implemented via the PyTorch framework. First, unlabeled data were used to perform comparative-learning-based pre-training on the backbone network on a server with 4 NVIDIA 3080Ti graphics cards, and 300 epochs were trained with a learning

Table 1 Data split in the experiment

Classification	Dataset	Pre-training	Downstream tasks	Standard plane recognition
Standard plane	Training set	1,079 (unlabeled)	320 (labeled)	–
	Test set	–	80 (labeled)	1,479
Non-standard plane	Training set	–	–	–
	Test set	–	–	2,958
Total		1,079	400	4,437

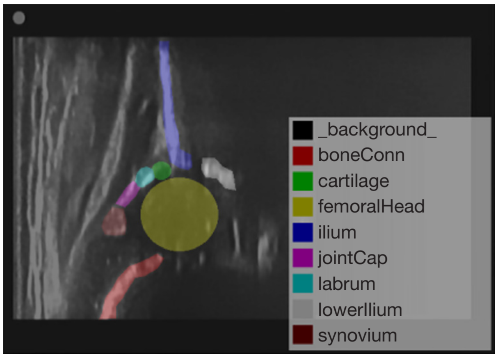

Figure 4 Schematic diagram of anatomical structure of infant acetabulum. boneConn, the junction of cartilage and bone; cartilage, hyaline cartilage preformed acetabular roof; femoralHead, femoral head; jointCap, joint capsule; labrum, acetabular labrum; ilium&lowerIlium, bony part of acetabular roof; synovium, synovial fold.

rate of 0.001. Then, semantic segmentation model and landmark identification were trained for 100 epochs with a learning rate of 0.001. The entire training process took 10 hours.

Evaluation metrics

In this study, we evaluated the effect of contrastive learning and compare our method (“Ori”) with other methods, including fully convolutional network (FCN) (17), Unet (18), and deeplabv3 (19). In this experiment, the following metrics were selected:

(I) Dice similarity coefficient (DSC): it is used to measure the similarity of 2 sets, with a range of 0 to 1. The larger the value is, the more similar the 2 sets are. It is often used to calculate the similarity of closed regions.

$$\mathrm{DSC}(X,Y)=\frac{2|X\cap Y|}{|X|+|Y|}=\frac{2TP}{2TP+FP+FN} \quad [7]$$

where X represents the predicted point set, and Y represents the labeled point set.

(II) Hausdorff distance (HD): it is sensitive to the segmented boundaries, used to measure the distance between 2 edge point sets and reflect the similarity between 2 contours.

$$h(A,B)=\max_{a\in A}\left\{\min_{b\in B}\left\{d(a,b)\right\}\right\} \quad [8]$$

$$\mathrm{HD}(A,B)=\max\left\{h(A,B),h(B,A)\right\} \quad [9]$$

where $d(.)$ is the distance between 2 points; A and B are the contour point sets; a and b are the points in the contour point sets in pixel.

(III) In addition, the metrics, accuracy, precision, and recall are used to evaluate the performance of standard plane recognition.

Results

The comparison of landmark identification between different models on DSC and HD is shown in *Table 2* and *Table 3*, respectively. The method with contrastive learning, denoted as “+CL”, had a mean DSC of 0.7873 and a mean HD of 5.0102 in landmarks identification, whereas the method without contrastive learning had a mean DSC of 0.7734 and a mean HD of 6.1586. These values were obtained from all tested images (80 samples). The average performance of the model pretrained based on the contrastive learning was better, as shown in *Figure 5*. In *Figure 6*, taking the bony part of acetabular roof as an example, the feature extraction networks based on contrastive learning could better extract features and ultimately perform better on semantic segmentation of landmarks.

Table 2 Comparison of landmark identification between different network models on DSC

Method	BoneConn	Cartilage	FemoralHead	Ilium	JointCap	Labrum	LowerIlium	Synovium	Avg
Unet	0.6851	0.0000	0.8761	0.7519	0.1654	0.5696	0.3792	0.5860	0.5016
FCN	0.7175	0.0647	0.8976	0.6802	0.5711	0.7356	0.5382	0.7059	0.6138
Deeplabv3	0.7431	0.3077	0.8995	0.7586	0.7045	0.7604	0.6951	0.8143	0.7104
Ori	0.8086	0.5593	0.8602	0.8701	0.7382	0.7652	0.8192	0.7663	0.7734
+CL	0.8636	0.5653	0.8738	0.8854	0.7269	0.8141	0.7928	0.7765	0.7873

DSC, Dice similarity coefficient; BoneConn, the junction of cartilage and bone; cartilage, hyaline cartilage preformed acetabular roof; FemoralHead, femoral head; JointCap, joint capsule; Labrum, acetabular labrum; Ilium&lowerIlium, bony part of acetabular roof; Synovium, synovial fold; Avg, average; Unet, the U-net model; FCN, fully convolutional network; Deeplabv3, the Deeplab v3 model; Ori, the original model; +CL, the original model with contrastive learning.

Table 3 Comparison of landmark identification between different network models on HD

Method	BoneConn	Cartilage	FemoralHead	Ilium	JointCap	Labrum	LowerIlium	Synovium	Avg
Unet	15.7367	142.4057	11.8173	17.9590	14.1463	9.7828	20.0939	16.3002	31.0302
FCN	6.6909	84.3727	3.4329	7.3286	3.9515	3.3489	8.9629	2.8928	15.1227
Deeplabv3	4.0278	16.7851	3.4537	8.2089	3.0556	2.8471	7.4389	2.6112	6.0535
Ori	7.1437	9.6309	8.1311	4.6308	3.1921	3.4713	5.0061	8.0626	6.1586
+CL	5.0624	5.6635	8.3374	6.5300	5.9187	3.0374	2.3865	3.1459	5.0102

HD, Hausdorff distance; BoneConn, the junction of cartilage and bone; cartilage, hyaline cartilage preformed acetabular roof; FemoralHead, femoral head; JointCap, joint capsule; Labrum, acetabular labrum; Ilium&lowerIlium, bony part of acetabular roof; Synovium, synovial fold; Avg, average; Unet, the U-net model; FCN, fully convolutional network; Deeplabv3, the Deeplab v3 model; Ori, the original model; +CL, the original model with contrastive learning.

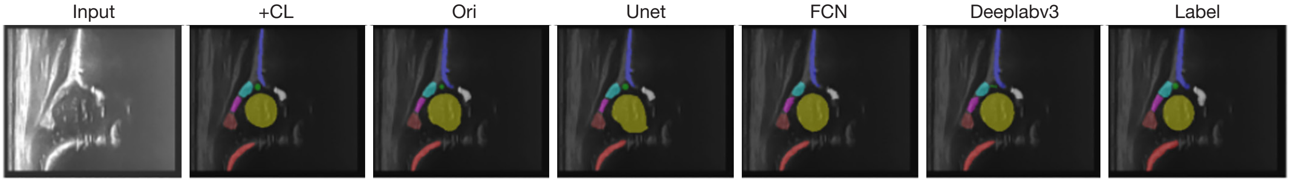

Figure 5 Comparison between various models and our proposed method. +CL, the original model with contrastive learning; Ori, the original model; Unet, the U-net model; FCN, fully convolutional network; Deeplabv3, the Deeplab v3 model.

Furthermore, the results of standard plane classification on the model "Ori+CL" were as follows: accuracy of 95.4%, precision of 91.64%, and recall of 94.86%, compared to the model without contrastive learning which had an accuracy of 93.66%, precision of 84.32%, and recall of 91.68%. A big difference on the precision value was apparent because our method via contrastive learning could improve the capacity of identification of image details. The corresponding confusion matrix is as shown in *Table 4* and *Table 5*. The receiver operating characteristic (ROC) curve of the model "Ori+CL" is illustrated in *Figure* 7. The corresponding area under the curve (AUC) was 0.982.

By means of 1,500 ultrasound images (500 positive samples and 1,000 negative samples), we implemented comparison experiments between the proposed algorithm and the professional manual operators in terms of false positive ratio (FPR) and false negative ratio (FNR), which were expressed as follows.

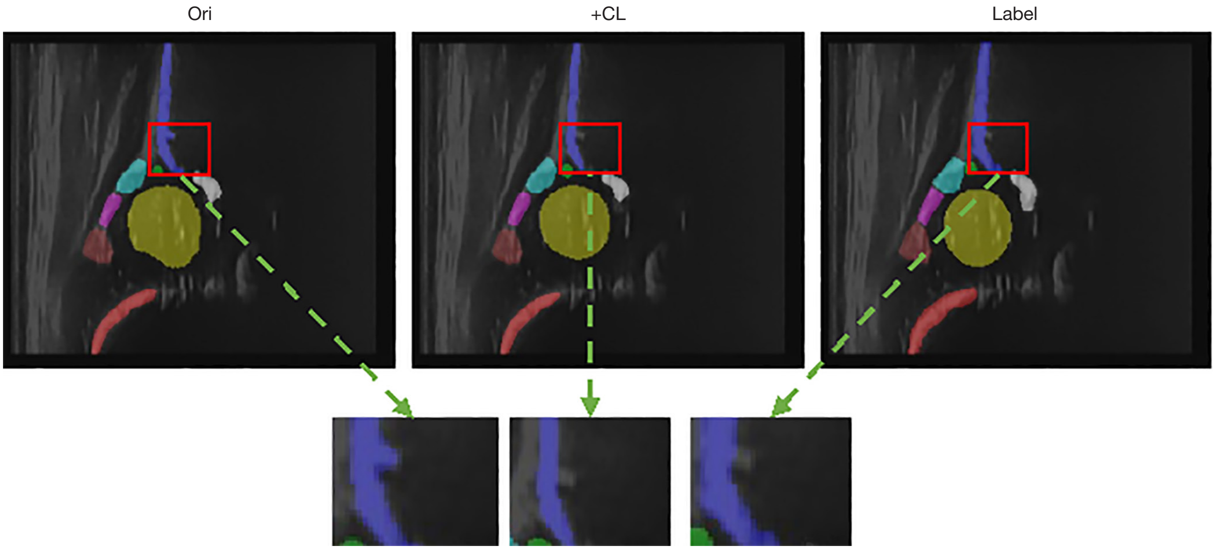

Figure 6 Comparison of semantic segmentation about effects of contrastive learning. +CL, the original model with contrastive learning; Ori, the original model.

Table 4 Confusion matrix of standard plane recognition with contrastive learning

Actual value	Predicted value	
	Positive	Negative
Positive	1,403	76
Negative	128	2,803

Table 5 Confusion matrix of standard plane recognition without contrastive learning

Actual value	Predicted value	
	Positive	Negative
Positive	1,356	123
Negative	252	2,706

Table 6 Comparison between our algorithm and manual operators

Methods	FPR (%)	FNR (%)
Ours	4.3	5
Manual operators	7.1	9.4

FPR, false positive ratio; FNR, false negative ratio.

$$\mathrm{FPR} = \frac{FP}{FP + TN} \quad [10]$$

$$\mathrm{FNR} = \frac{FN}{TP + FN} \quad [11]$$

where FP denotes the number of false positive samples, TN means the number of true negative samples, FN represents the number of false negative samples, and TP indicates the number of true positive samples.

The concrete comparison results are provided in *Table 6*. From this table, we can see that the performance of our algorithm was better than that of manual operators. Specifically, the FPR of the proposed method decreased by 39.4%, and the FNR of the manual operators was 9.4%, which is far higher than that of our algorithm (5%). Obviously, the established model can assist doctors in diagnosis to improve medical efficiency.

Discussion

In recent years, deep neural networks have been gradually explored in hip ultrasound screening of infants, due to their excellent image feature extraction capabilities. In

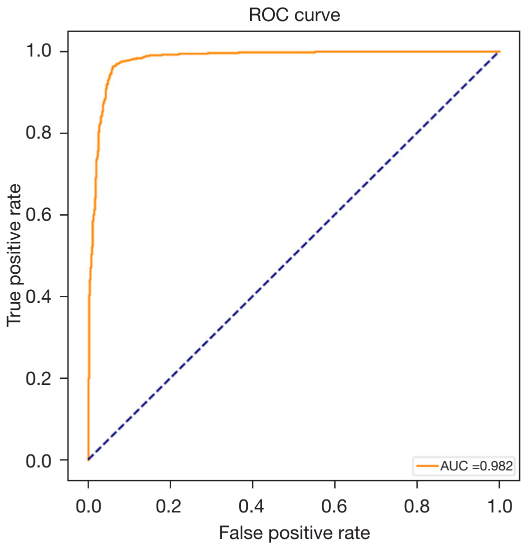

Figure 7 The ROC curve of our proposed model. ROC, receiver operating characteristic; AUC, area under the curve.

2016, Golan *et al.* introduced a deep convolutional neural network (CNN) for the first time (5). They used CNN to divide the flat iliac bone and lower limbs to automatically calculate the Graf's alpha angle in infant sonographic images. In 2017, Hareendranathan *et al.* proposed a method to automatically segment the acetabulum bone and derive geometric indices of hip dysplasia (6). In 2018, Zhang *et al.* introduced region of interest (ROI) into a fully CNN to improve the recognition accuracy of acetabulum (7). In 2019, Sezer *et al.* input the segmented image segments into CNN to directly diagnose the patient's hip development (8). In the same year, El-Hariri *et al.* proposed a feature-based deep learning method that utilizes the multi-channel input of U-Net to achieve iliac segmentation (9). The methods used in these studies require a large amount of preliminary data annotation work in practical applications. This study applied contrastive learning to DDH ultrasound examination. It was found that our pre-training method can effectively moderate the demands of data annotation, and improve the accuracy of ultrasound detection, especially when there are multiple data sources. These findings are of great significance to guide clinical treatment.

In this paper, we propose a semi-supervised learning framework for assisting DDH diagnosis, by using data enhancement and contrastive learning to deal with the problem of distribution shift due to ultrasound data from various objects and machines. A backbone network based on the FPN structure is used to effectively identify landmarks at different scales. The performance of detection is improved in terms of accuracy and robustness. The contrastive learning method in this article has a mean DSC of 0.7873 and a mean HD of 5.0102 in landmark identification. The mean DSC without the contrast learning method is 0.7734 and the mean HD is 6.1586. As a reference, the metrics of FCN, Unet, and deeplabv3 are worse than our proposed method. By using this method, the accuracy of standard plane recognition is 95.4%, the precision is 91.64%, and the recall is 94.86%.

This study shows that the contrastive self-supervised learning method and the FPN structure can effectively extract the features of sonographic images and improve the performance of analysis and measurement. The pre-training method can effectively reduce the need for data annotation and lower the threshold for using deep learning in DDH inspection. This solution enables automated measurement and evaluation of sonographic images. It can also promote large-scale DDH screening, help patients detect and treat early-stage diseases in time, and improve prognosis. This model can assist doctors in diagnosis via a visible interpretation and automated measurement, improve medical efficiency, and has important clinical significance. At present, our method is in the process of clinical verifications.

This study had some limitations: (I) in terms of standard plane acquisition, this model can only provide simple prompts based on the identified landmarks. Further work is needed on how to automatically obtain standard images from the examination process to better serve the clinic. (II) In terms of downstream task, the training data used by this model only includes type I and type II hip data. More data on dislocated hips need to be collected and annotated for further research. (III) Artificial intelligence (AI)-based methods can only provide assistance to doctors. In actual clinical work, doctors can obtain diagnosis data from various patients to improve diagnostic accuracy.

Conclusions

We have proposed a semi-supervised deep learning method following Graf's principle, which can better utilize past ultrasound examination data to more accurately identify the landmarks of infant hips, thereby improving the efficiency of diagnosis of infant hip dysplasia. Early screening of DDH by auxiliary doctors is of great significance and has broad clinical application prospects.

Acknowledgments

Funding: This study was supported by the Research Foundation of Jiangsu Provincial Commission of Health (No. ZD2022049).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-1384/coif). All authors report that this study was supported by the Research Foundation of Jiangsu Provincial Commission of Health (No. ZD2022049) from Jiangsu Provincial Commission of Health, China. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was approved by the Research Ethics Committee of Southeast University and Yangzhou Maternal and Child Health Care Service Centre. The requirement for individual consent for this retrospective analysis was waived. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Gulati V, Eseonu K, Sayani J, Ismail N, Uzoigwe C, Choudhury MZ, Gulati P, Aqil A, Tibrewal S. Developmental dysplasia of the hip in the newborn: A systematic review. World J Orthop 2013;4:32-41.
2. Nicholson A, Dunne K, Taaffe S, Sheikh Y, Murphy J. Developmental dysplasia of the hip in infants and children. BMJ 2023;383:e074507.
3. Yan H, Du L, Liu J, Yang X, Luo Y. Developmental retardation of femoral head size and femoral head ossification in mild and severe developmental dysplasia of the hip in infants: a preliminary cross-sectional study based on ultrasound images. Quant Imaging Med Surg 2023;13:185-95.
4. O'Beirne JG, Chlapoutakis K, Alshryda S, Aydingoz U, Baumann T, Casini C, et al. International Interdisciplinary Consensus Meeting on the Evaluation of Developmental Dysplasia of the Hip. Ultraschall Med 2019;40:454-64.
5. Golan D, Donner Y, Mansi C, Jaremko J, Ramachandran M. Fully automating Graf's method for DDH diagnosis using deep convolutional neural networks. International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis International Workshop on Deep Learning in Medical Image Analysis 2016:130-41.
6. Hareendranathan AR, Zonoobi D, Mabee M, Cobzas D, Punithakumar K, Noga M, Jaremko JL. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, Australia, 2017:982-5.
7. Zhang Z, Tang M, Cobzas D, Zonoobi D, Jagersand M, Jaremko JL. End-to-end detection-segmentation network with ROI convolution. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018:1509-12.
8. 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.
9. El-Hariri H, Mulpuri K, Hodgson A, Garbi R. Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, 2019:12-20.
10. Hu X, Wang L, Yang X, Zhou X, Xue W, Cao Y, Liu S, Huang Y, Guo S, Shang N, Ni D, Gu N. Joint Landmark and Structure Learning for Automatic Evaluation of Developmental Dysplasia of the Hip. IEEE J Biomed Health Inform 2022;26:345-58.
11. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, Tang J. Self-supervised learning: generative or contrastive. IEEE Transactions on Knowledge and Data Engineering 2023;35:857-76.
12. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017:2117-25.

13. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. ICML'20: Proceedings of the 37th International Conference on Machine Learning, 2020:1597-607.
14. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020:9912-24.
15. Grill JB, Strub F, Altché F, Tallec C, Richemond P, Buchatskaya E, Doersch C, Pires BA, Guo ZD, Azar MG, Piot B, Kavukcuoglu K, Munos R, Valko M. Bootstrap your own latent: A new approach to self-supervised learning. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020:21271-84.
16. Chen X, He K. Exploring simple siamese representation learning. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021:15750-8.
17. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:640-51.
18. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham, 2015:234-41.
19. Chen LC, ZhuangbilityY, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y. editors. Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11211. Springer, Cham, 2018:801-18.

Cite this article as: Li X, Zhang R, Wang Z, Wang J. Semi-supervised learning in diagnosis of infant hip dysplasia towards multisource ultrasound images. Quant Imaging Med Surg 2024;14(5):3707-3716. doi: 10.21037/qims-23-1384