The Visual Computer
https://doi.org/10.1007/s00371-024-03789-2

RESEARCH

AutoDDH: A dual-attention multi-task network for grading developmental dysplasia of the hip in ultrasound images

Mengyao Liu[1] · Ruhan Liu[2] · Jia Shu[3] · Qirong Liu[4] · Yuan Zhang[1] · Lixin Jiang[1]

Accepted: 29 December 2024

Abstract

Developmental dysplasia of the hip (DDH) in newborns can cause serious long-term adverse effects if not promptly diagnosed and treated. Early intervention via ultrasound screening at 0–6 months is beneficial. However, early DDH diagnosis by ultrasound is complex and requires high-level experience for radiologists. Hence, deep learning for clinically assisted DDH screening is meaningful. The DDH classification task is challenging due to low ultrasound image resolution and difficulty in extracting structural features. We propose a dual-attention multi-task network (AutoDDH) for DDH grading using ultrasound images. It includes a dual-attention module for feature enhancement, a feature fusion module for detail improvement, and a dual-output branch for position embedding and generating outputs of DDH grading and anatomical structure segmentation. With the help of the segmentation task, the average accuracy and AUC of DDH four classifications reached 80.43% and 0.96, outperforming other methods and laying the foundation for DDH intelligent assisted screening. Code available at: https://github.com/Liuruhan/AutoDDH.

Keywords Ultrasound image (US) · Developmental dysplasia of the hip (DDH) · Dual attention · Detection network

1 Introduction

Developmental dysplasia of the hip (DDH) is one of the most common congenital disorders in newborns [1]. Early treatment is crucial as correction becomes more difficult after one year of age. Untreated DDH can lead to serious consequences later in life, including lameness, inconsistent leg length, pain, frequent surgery, osteoarthritis, and disability. In severe cases, artificial hip replacements may even be required in adulthood, and studies have shown that DDH is one of the most common reasons for hip replacement in patients under 60 [2]. Ultrasound is the key means for early DDH screening in newborns aged 0–6 months due to its non-invasive, non-radiation, and low-cost characteristics [3]. Accurate DDH classification diagnosis based on ultrasound images is highly significant.

However, accurate ultrasonic DDH screening faces issues such as slow experience accumulation for radiologists and difficult screening techniques [4, 5]. While many researchers are considering using deep learning technology for automated DDH diagnosis [6–8], existing methods based on binary classification models and key point detection have limitations in clinical practice [9, 10]. These methods have achieved good accuracy through the design of binary classi-

Mengyao Liu and Ruhan Liu equally contributed to this work.

✉ Lixin Jiang
jinger_28@sina.com

Mengyao Liu
mengyao_liu08@163.com

Ruhan Liu
223101@csu.edu.cn

Jia Shu
frmdshujia@sjtu.edu.cn

Qirong Liu
liuqirong5833@link.tyut.edu.cn

Yuan Zhang
columbianzhang@163.com

1 Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

2 Furong Laboratory, Central South University, Hunan, China

3 Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

4 College of Software, Taiyuan University of Technology, Taiyuan, China

Published online: 17 February 2025

fication model and key point detection technology. However, these methods are not effective in clinical practice. First, the model based on binary classification cannot give visual results, and the confidence is challenged in the clinic. In addition, the annotation method based on key point detection is complicated, and the large-scale annotation costs the radiologist a lot of time, so it is difficult to realize. Therefore, how to use a small amount of structure segmentation labeling and a large number of classification labeling to develop a multi-task learning model to achieve accurate and visualized deep learning networks has become an important research problem in the field of automated DDH screening.

To address these challenges, we propose AutoDDH, which combines dual-attention mechanisms and multi-task learning for optimal accuracy. There are four types of ultrasound image in hip: normal hip development (Type I), mild DDH (Types IIa and IIb), severe DDH (Type IIc), and hip dislocation (Types III and IV) [11]. The DDH classification standard and examples are shown in Fig. 1. In summary, our contributions are in the following ways:

- We proposed a dual-attention multi-task network (AutoDDH) for grading four types of DDH and segmenting seven key structures of the hip joint. The average accuracy of AutoDDH is 80.43%, average F1 score is 81.17%, and average AUC is 0.96.
- We adopted a two-stage training method and multi-task loss function. In the training process, we first conducted segmentation training to learn anatomical structure information, and then conducted multi-task learning to improve classification performance (average F1 increase by 5.57%).
- Based on the NHBS-Net composed of dilated ResNet, the two-channel attention mechanism, and the feature fusion module, we further integrated the location coding to integrate the spatial location information better and improve the segmentation effect of details.

The paper is structured as follows: Sect. 2 reviews relevant literature, Sect. 3 details our proposed method, Sect. 4 describes the experiment setup and results, along with comparison and analysis, and Sect. 5 introduces the discussions and conclusions.

2 Related work

This section primarily analyzes the application of deep learning-based models in DDH ultrasound images from the following two related aspects: identification and segmentation of hip bone and cartilage structures in newborns by ultrasound, and automated ultrasound-based DDH screening and diagnostic grading.

2.1 Key structure segmentation in DDH

Ultrasound-based segmentation and identification of bone cartilage structure of hip joint in newborn serve as an important foundation for subsequent screening, diagnosis, and grading. Previous researchers have conducted extensive work on the detection and segmentation of anatomical structures

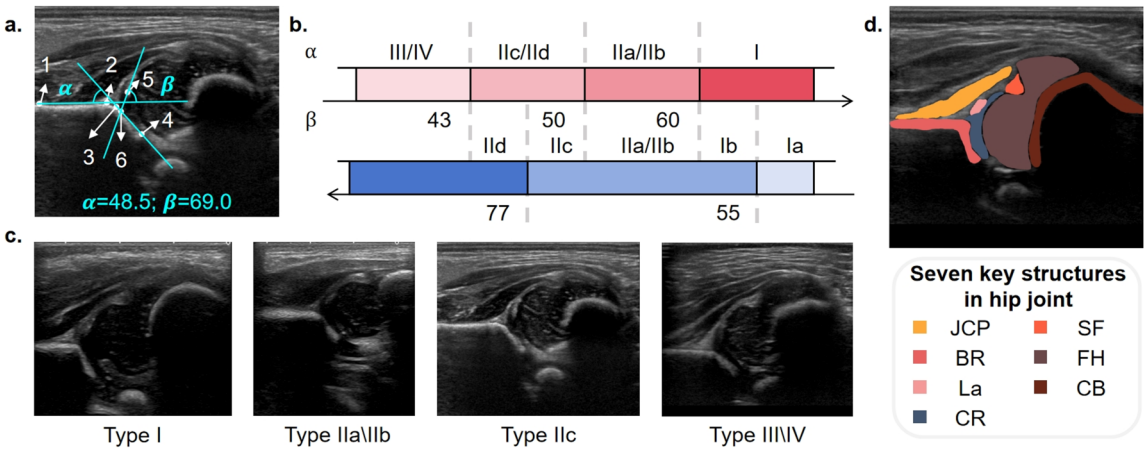

Fig. 1 Methods and examples of ultrasound diagnosis and grading measurement of DDH. **a.** Schematic diagram of six key points of grading measurement in standard ultrasonic image. 1—bony roof (left), 2—bony roof (right), 3—bony acetabular roof, 4—lower limb-plane, 5—middle of the labrum, and 6—turning point of the bony rim. **b.** The standard of ultrasound diagnosis and grading of DDH is based on the degree of α and β angles. **c.** An example diagram of the four DDH classifications. **d.** Seven key anatomical structures in neonatal hip joint, including joint capsule & perichondrium (JCP), bony roof (BR), labrum (La), cartilaginous roof (CR), synovial fold (SF), femoral head (FH), and chondro-osseous border (CB)

in ultrasonic images. Early studies mainly focused on artificially designed features to segment bone structures with high echo, such as the iliac bone. However, it was difficult to achieve satisfactory results for other cartilage structures with significant feature differences and identification difficulties [12–14]. de Luis-Garcia et al. [15] segmented the femoral head and bony rim based on energy function and texture information to identify the anatomical structure of the hip. In addition, Quader et al. [16] proposed using confidence-weighted structured phase symmetry (CSPS) feature to segment different bone structures of the hip joint in 3D ultrasound images to improve the segmentation efficiency of bone structures. Pandey et al. [17] proposed the shadow-peak (SP) method to further simplify the bone shadow feature extraction method, which has a certain improvement in accuracy and speed compared with CSPS.

Recently, deep learning method has also achieved outstanding results in the field of hip joint structure segmentation in neonatal ultrasound images. These methods can be well extended to large-scale data and have demonstrated certain accuracy and robustness for cartilage structures [18, 19]. El-Hariri et al. compared artificial design features such as U-Net, SP, and CSPS with single-channel and multi-channel inputs and observed that deep learning methods performed better than artificial features [18]. Moreover, in our previous study, we proposed seven key neonatal hip bone-cartilage structure segmentation models, NHBS-Net. By designing feature enhancement and fusion modules, we improved the accuracy of the segmentation model in edge recognition, reduced the error rate of bone-cartilage structure segmentation, and made the model highly robust for different data scales [19].

2.2 Ultrasound-based DDH screening methods

In recent years, researchers have focused on the study of DDH intelligent diagnosis and screening models based on ultrasonic images and have achieved remarkable results. Deep learning methods provide an important tool for the development of automatic DDH diagnosis models for ultrasonic images [9, 10, 20]. Some researchers consider using binary classification methods to study DDH diagnostic models. For example, Gong et al. [9] proposed a deep exclusive regularization machine based on two-stage meta-learning for the development of a DDH binary classification auxiliary diagnosis model for ultrasonic images. Experimental verification was conducted, and an accuracy rate of more than 0.85 was obtained.

Another group of researchers explored the use of anatomical structure segmentation and key point detection to achieve DDH automatic line diagnosis and grading. Sezer et al. [20] established a fully automatic computer-aided diagnosis system based on a convolutional neural network. It carried out automatic classification diagnosis based on the standard plane on the basis of recognizing three necessary anatomical and diagnostic elements in DDH ultrasound images. Additionally, Shen Bozhi et al. also published an assisted screening method and screening system for hip joints in newborns, which was used to calculate the measurement angles in standard ultrasound images. Chen et al. [10]developed a deep neural network tool that can automatically label the five key points involved in the Graf guideline [11]. It realizes the relevant angle measurement and DDH diagnostic classification. However, the above methods did not consider the relationship between DDH classification and the relative location distribution of anatomical structures. On one hand, the characteristic information and spatial location correlation of structures could not be fully utilized. On the other hand, the detected key points might conflict with the location of structures, resulting in low accuracy. Therefore, considering the anatomical characteristics of DDH classification has more important significance.

2.3 Multi-task learning

In the past few years, significant correlations have often been demonstrated in medical tasks targeting the same images. For instance, in the classification and segmentation of breast cancer based on ultrasound images [21] and chest X-ray-based pneumonia diagnosis and lesion area segmentation [22]. Multi-task learning methods exhibit great potential in medical imaging analysis for disease detection and diagnosis. The combination of key structure segmentation and hierarchical diagnosis of ultrasonic image DDH with the multi-task learning method also holds high research significance.

3 Method

The proposed dual-attention multi-tasking network architecture is shown in Fig. 2. We used dilated ResNet50 as the network backbone to extract multi-scale features. Based on the dual-attention and feature fusion module architecture in [19], we enhanced and fused extracted features. We used location coding to emphasize spatial position relationships and improve segmentation and classification accuracy. Location-encoded features are fed into two output branches for segmentation and classification results. Additionally, we propose a training strategy based on segmentation pre-training. First, the segmentation branch helps the network understand anatomical features and structure positions. Then, the joint network is trained to improve multi-task learning performance. Our network has fewer parameters and can achieve optimal segmentation classification performance compared to others.

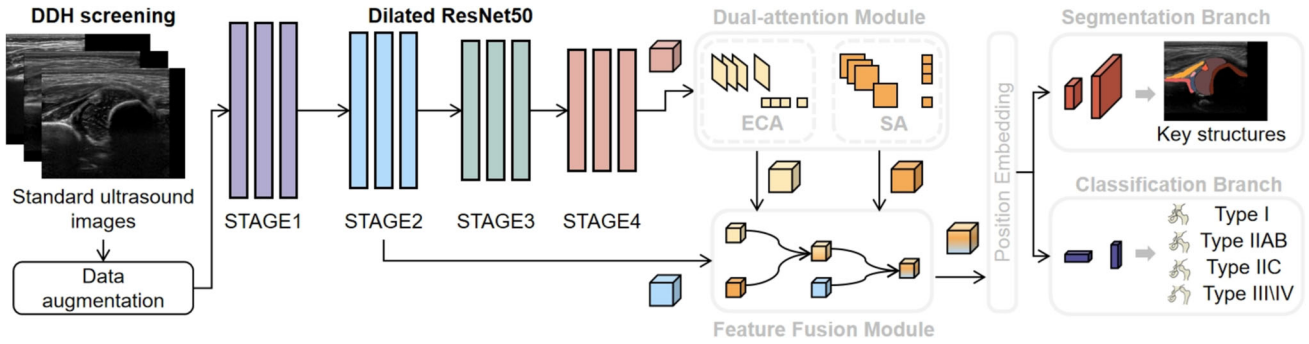

Fig. 2 Overview of AutoDDH. The AutoDDH model consists of four modules: network backbone based on dilated ResNet50, dual-attention module, feature fusion module, and dual-output branch for segmentation and classification. The ultrasonic image is augmented by the image enhancement method of gray-scale transformation and then input into the network backbone to obtain the low-dimensional features and high-dimensional features. The high-dimensional features are enhanced by DAM, and then fused with the low-dimensional features by FFM. The resulting fusion features enhance the spatial position information of the features by location coding. Finally, the task is segmented and classified by two output branches, respectively

3.1 DDH segmentation and classification

Our proposed multi-task model aims to achieve intelligently assisted DDH classification using anatomical information, thereby aiding in enhancing the efficiency of clinical DDH screening. DDH ultrasound screening has more complex ultrasonic diagnostic criteria. Table 1 summarizes the detailed description of four Graf's classification types of ultrasound diagnosis [11], including normal, mild DDH, severe DDH, and hip dislocation.

3.2 Data augmentation

Since the ultrasound images obtained from different participating infants vary in size, and the classification of DDH depends on the position relationship between anatomical structures in the images. Consequently, image augmentation methods such as stretching and scaling, which alter the shape and position of anatomical structures in ultrasonic images, are not suitable for DDH ultrasonic images. In image preprocessing, we employed the method of equal ratio transformation to first transform all ultrasound images to the same height, and then fill the width with zero value to ensure that each image has the same length and width without changing the relative position of the anatomical structure. Further, in image augmentation, we selected inversion, gray linear change, gamma transform, and other augmentation methods to enrich the diversity of input without affecting the key position information.

3.3 Pre-trained dilated ResNet50

For the selection of the network backbone, we adopted dilated ResNet50 [23] as the backbone of the feature extractor. The dilated ResNet architecture can enable the back convolutional layer to maintain a larger feature maps size while keeping the number of parameters unchanged and the field of view of the convolutional layer at each stage unchanged. This is beneficial for the detection of small targets such as La and SF. When constructing AutoDDH, we took advantage of the pre-trained weights on ImageNet and performed fine-tuning on our dataset based on that weight. The weight transfer considers only the parameters of the four convolution stages used for feature extraction.

3.4 Dual-attention module

The attention mechanism is capable of enhancing features, highlighting more useful and crucial ones. Models based on the attention mechanism have also achieved remarkable results in medical image processing. In our previous research work [19], NHBS-Net based on dual attention has yielded excellent results in the segmentation of seven key structures of hip joints in infants. In this study, we employed the previous dual-attention module (DAM) to enhance the features extracted by dilated ResNet50 [23]. The implementation details of the dual-attention module are shown in Fig. 3. The output of dilated ResNet50 is fed into DAM to obtain position attention maps (PAMs) and channel attention maps (CAMs). The two-channel attention mechanism can make the model pay more attention to the structure-related features, which has advantages for the subsequent multi-tasks of key point detection and structure segmentation.

3.5 Feature fusion module and output branches

AutoDDH uses a two-level feature fusion module to perform feature fusion after different enhancements and fusion of high and low dimensional features. The structural details of the fusion module used are shown in Fig. 4. Through the

Table 1 Graf's grade [11] for developmental dysplasia of the hip using ultrasound

Grades	Description	α angle range	β angle and cartilaginous roof	Age range
Type I	Normal	$\geq 60°$	Type Ia: $< 55°$ Type Ib: $> 55°$ Cover the femoral head well	Any age
Type IIa/IIb	Mild DDH	$50° - 59°$	$55° - 77°$ Cover the femoral head	Type IIa: < 3 mths Type IIb: > 3 mths
Type IIc/IId	Severe DDH	$43° - 49°$	$< 77°$ Type IIc: Cover the femoral head Type IId: Decentered hip with a displaced cartilage roof	Any age
Type III/IV	Hip dislocation	$< 43°$	Labrum pressed upward or downward	Any age

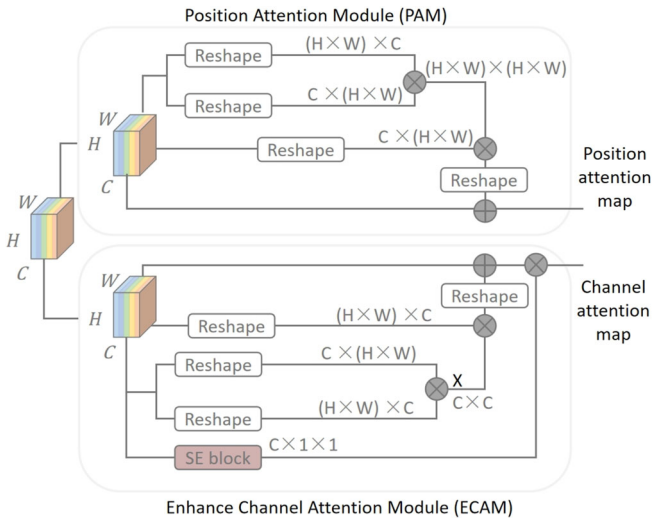

Fig. 3 The architecture of the dual-attention module. The dual-attention module contains two-path attention named enhance channel attention module position attention module. In two attention paths, different paths enhance important information in channel and location relatively

FFM module, two attention feature maps can be fused well, and high-level feature maps and low-level feature maps can be fused to obtain better feature representation.

After that, the extracted high-level features (feature fusion maps $F \in \mathbb{R}^{C\times H\times W}$) are reshaped to flatten features $R \in \mathbb{R}^{C\times N}$ ($N = H \times W$). The flatten features R are fed into a 1-D embedding layer Emb to obtain position embedding features $P \in \mathbb{R}^{C\times N}$. Finally, we resize the position embedding features P to the same size as F.

The input of two output branches (segmentation branch and classification branch) is the sum of position embedding feature P and high-level features F. We use two convolutional layers and an interpolation layer to generate segmentation results, and an average pooling layer and a linear layer to generate classification results.

3.6 Training strategy of AutoDDH

To take full advantage of the relevance of segmentation and classification tasks, we designed a two-stage training strategy (TTS) using clinical diagnostic logic, as shown in Fig. 5. First, seven kinds of key anatomical structure segmentation data were used to train the feature extraction, enhancement, fusion backbone and segmentation branches, so that the AutoDDH model could extract the relative features and position relationships of different anatomical structures, and save the model parameters with the best segmentation accuracy.

After that, the AutoDDH model is fine-tuned based on the parameter weights mention below. Specifically, the segmented data of seven structures were processed, and 2 anatomical structures closely related to diagnosis were retained: BR and La. The processed segmentation labels are supervised jointly with the classification results, and the AutoDDH model is trained again. At this time, both the segmentation and classification output branches are trained to obtain the final segmentation and classification performance.

3.7 Loss function

For segmenting seven anatomical structures of the hip joint, a segmentation loss function combined with focal loss and

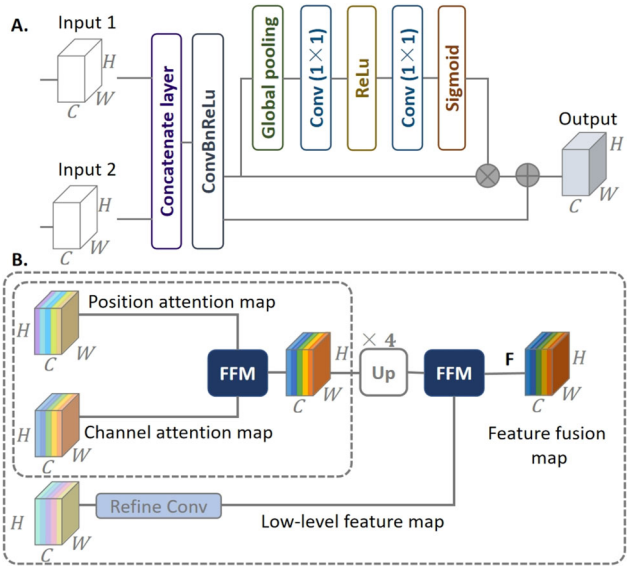

Fig. 4 Structure figure of feature fusion module. A. Structure details of feature fusion block (FFB) using in fusing the location and channel feature maps. B. The architecture of feature fusion module. First, a feature fusion block uses PAMs and CAMs to generate fused dual-attention maps that have the same dimension of location or channel feature maps, and the dual-attention maps are interpolated to 16 times the original (width×4, height×4). Another feature fusion block is applied to fuse dual-attention maps in the above and low-level feature maps. The output of stage 2 of dilated ResNet50 goes through a refined convolution layer to produce a low-level feature map with the same number of channels as CAMs and PAMs

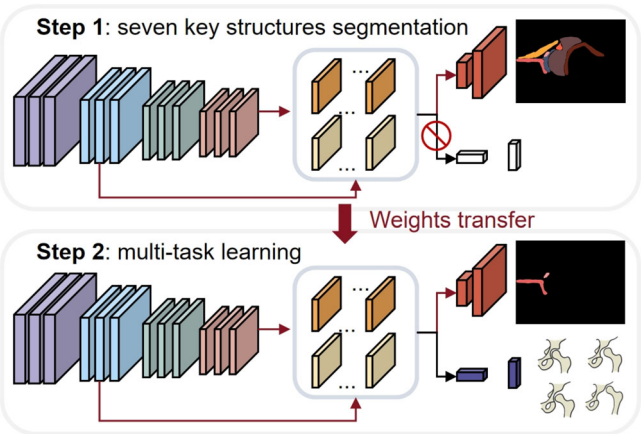

Fig. 5 Schematic diagram of training strategy. Two steps are used in training. First, the feature extraction, enhancement, and fusion modules are trained through seven key structures segmentation task. In this step, only segmentation branch is trained. After that, the whole AutoDDH network is trained by two tasks: segment two diagnosis-related anatomical structures and grade 4 types of DDH. Before training, we transferred the weights of best model in step 1

cross-entropy loss is used. Details of the segmentation loss function can be found in [19].

Further, in segmentation and classification of multi-task learning, the overall loss $\mathbb{L}_{tot}$ is as follows:

$$\mathbb{L}_{tot} = \mathbb{L}_{seg} + \theta \cdot \mathbb{L}_{cls} \tag{1}$$

$\mathbb{L}_{seg}$ takes the same loss form as in [19], but changes the number of classes from seven to two. $\mathbb{L}_{cls}$ is the classification loss, and θ is the weight of classification loss.

There is a serious class imbalance problem in DDH grading, with very few severe DDH and hip dislocation, so we consider using focal loss as the classification loss to solve the class imbalance problem. The classification loss function is as follows:

$$\begin{aligned}\mathbb{L}_{cls} &= \alpha_t(1-p_t)^{\gamma} log(p_t) \\ &= \sum_i -\alpha_t^i(1-p_t)^{\gamma} log(p_t)\end{aligned} \tag{2}$$

where $\alpha_t = \{\ldots, \alpha_t^i, \ldots, \alpha_t^i, \ldots, \alpha_t^{C_N}\}$ is the weighted parameters of different DDH types, C_N is four representing four classes of DDH, p_t represents the output probability distribution, γ is the weight of difficult samples.

4 Experiments

4.1 Dataset

Our study analyzed data from two datasets. The first, the neonatal hip ultrasound (NHU) dataset, included 563 ultrasound images from 271 infants, segmented into seven key structures by experienced radiologists. These images were categorized into three DDH grades: normal, mild, and severe. The NHU dataset was split into a training set (400 images), validation set (53 images), and test set (110 images). For more details, see Reference [19].

The second dataset expanded on the first by adding ultrasound data from Renji Hospital, Shanghai Jiao Tong University School of Medicine, collected between December 2020 and June 2022. It included infants aged 0 to 6 months with high-risk factors for DDH. The study was ethically approved, and images were taken using 5/7.5MHz linear ultrasound transducers set at 40-55 mm depth. Each infant had 1-10 DICOM-formatted images taken. This dataset contained 4184 images from 778 individuals, divided into training (2685 images), validation (671 images), and test sets (828 images). There was no overlap in the training, validation, and testing sets between the two datasets. See Fig. 6 for the detailed division and distribution of data.

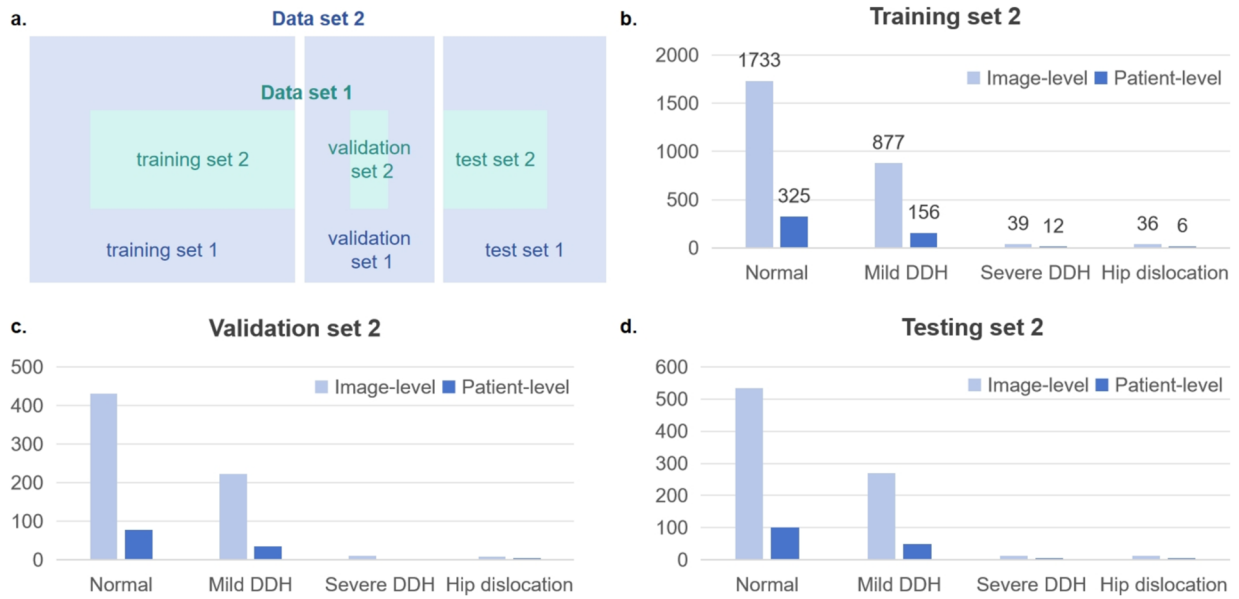

Fig. 6 Relationship between datasets and data distribution diagram of dataset 2. a. Relationship between dataset 1 and dataset 2. **a.** Distribution of DDH types in training set 2. **c.** Distribution of DDH types in validation set 2. **d.** Distribution of DDH types in testing set 2

4.2 Implementation details

Our study carried out all experiments on an Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz CPU and two NVIDIA GeForce RTX 3090 GPUs running on the Ubuntu Linux platform. The proposed network, AutoDDH, and other state-of-the-art methods employed the same loss function design and grid search strategy for hyperparameters. To determine suitable hyperparameters, we explored optimal values of batch size (ranging from 2 to 32), learning rates (from 10^{-3} to 10^{-7}), and the weight of focal loss (from 0.1 to 5). We used the Adam optimizer for the selection of the optimizer. All methods were trained for 30 epochs. The model that achieves the best performance on the validation set is retained as the final model, and the model performance is verified on the test set.

4.3 Evaluation metrics

In segmentation task, the performance of AutoDDH network and other segmentation networks was evaluated using Dice similarity coefficient (DSC), the DSC metric can be calculated as follows:

$$DSC = 1 - \frac{2\sum_{pixels} y_{true} y_{pred}}{\sum_{pixels} {y_{true}}^2 + \sum_{pixels} {y_{pred}}^2} \tag{3}$$

where y_{true} is the ground truth segmentation label for each class and y_{pred} is the prediction figure for each class.

The performance of AutoDDH and other classification models was evaluated using accuracy, F1-score, recall, κ, and AUCs of Receiver operating characteristic (ROC) curves. The accuracy, F1-score, recall, and κ can be defined as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$re = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 \times \frac{pre \times rec}{pre + rec} \tag{6}$$

$$\kappa = \frac{accuracy - p_e}{1 - p_e} \tag{7}$$

where TP is the true positive samples, TN is the true negative samples, FP is the false positive samples, FN is the false negative samples, $pre = \frac{TP}{TP+TP}$ is the precision rate, and p_e is the sum of the samples of the actual and predicted quantities for all categories divided by the square of the total number of samples.

4.4 Experimental results

In the experimental results, we first compare the performance of AutoDDH model with other classification models and multi-task models on DDH classification. The models compared include: ResNet-50 [24], DenseNet-121 [25], dilated ResNet-50 [23], and visual transformer (ViT) [26], BiSeNet [27], AUNet [28], and UNet [29]. Then, we further compared the segmentation performance of AutoDDH model with other models on the anatomically relevant diagnosis. Finally, we conducted ablation experiments to explore the

Table 2 Comparison classification performance of our AutoDDH and other state-of-the-art methods in image level, including classification networks (ResNet50 [24],DenseNet-121 [25], dilated ResNet-50 [23], and visual transformer (ViT) [26]) and multi-task learning networks (BiSeNet [27], AUNet [28], and UNet [29])

Method's types	Model	acc (%)	re (%)	F1 (%)	κ
classification	ResNet50	73.07%	65.16%	62.24%	0.4476
	DenseNet121	72.95%	63.28%	60.31%	0.4468
	DilatedResNet50	74.88%	73.82%	69.03%	0.4815
	ViT	73.91%	67.66%	66.34%	0.4590
multi-task learning*	BiSeNet	77.42%	79.45%	74.89%	0.5368
	UNet	76.93%	77.18%	72.26%	0.5291
	AUNet	78.50%	80.24%	75.60%	0.5615
	Ours (AutoDDH)	**80.43%**	**87.01%**	**81.17%**	**0.5994**

* Each network is the segmentation model added with a classification output header (one average pooling layer and one linear output layer)

influence of module design and training strategies on model performance in AutoDDH.

4.4.1 Comparison performance of DDH grading

In order to evaluate the validity of the proposed model structure, we conducted an experiment comparing the results of AutoDDH in DDH four classifications with the current popular network models. First, we compared it to four popular classification networks. Compared with models such as ResNet-50, DenseNet-121, dilated ResNet-50 and ViT, the proposed AutoDDH network has improved in accuracy, recall rate, F1 score, and κ value achieved better results. Compared with the best results in the classification model, accuracy was increased by 5.55%, recall rate by 13.19%, F1 score by 12.14%, and κ value by 0.1179. Furthermore, our AutoDDH network also achieves better results than the multi-task model (BiSeNet [27], AUNet [28], and UNet [29]). Compared with the best results in the multi-task model, accuracy was increased by 1.93%, recall rate by 6.77%, F1 score by 5.57%, and κ value by 0.0379.

Table 2 shows the picture-level training results for each model on the dataset. As can be seen from the table, dilated ResNet-50 achieved the best performance among all the compared classification models when using the classification model, and was significantly better than other models in four indicators. AutoDDH also adopts the network backbone based on dilated ResNet-50 and achieves the optimal accuracy, which is higher than 80% in accuracy, recall rate and F1 score. We then use the ROC curve to show the accuracy of the AutoDDH across the four DDH classes (see Fig. 7). The average AUC of AutoDDH is 0.96. It can be seen from the above results that the proposed AutoDDH model is obviously superior to other existing models in DDH classification.

Further, we obtain a confusion matrix for the top six precision models on the test set (as shown in Fig. 8). As shown in Fig. 8, compared with other networks, our AutoDDH obtains the best four types of accurate predictions, with the accuracy results of 84.2%, 71.5%, 92.3% and 100.0%. The main inaccurate predictions came from the mild DDH class, which is consistent with clinical experience: the identification of mild lesions is often the most difficult problem to solve in clinical practice.

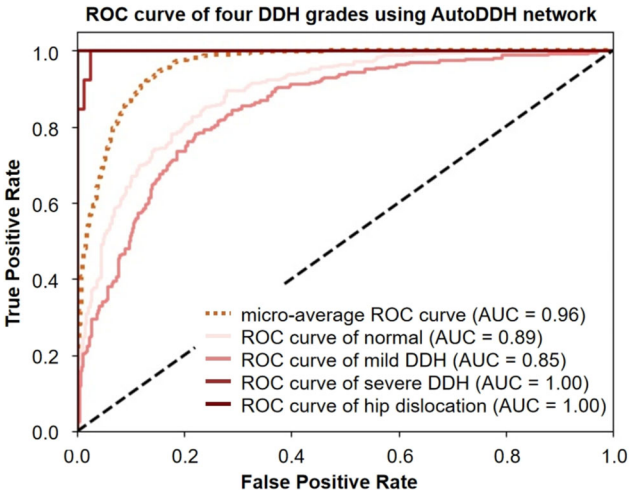

Fig. 7 The ROC curves of the four DDH types (normal, mild DDH, severe DDH, and hip dislocation) in AutoDDH model

4.4.2 Comparison performance of DDH key structure segmentation

After using the two-stage multi-task learning training strategy, AutoDDH not only has the ability to output DDH classification, but also can improve the segmentation of the two structures related to diagnosis (labrum and bony rim). The results of segmenting the final two structures are shown in Table 3. As can be seen from Table 3, AutoDDH also achieved the best segmentation effect in the segmentation of labrum and bony rim. AutoDDH achieved 85.05% DSC on labrum and 89.48% DSC on bony rim. The final average segmentation accuracy is 1.35% higher than the best model in other methods. In addition, we also show an example of segmentation of AutoDDH in detail and overall compared to

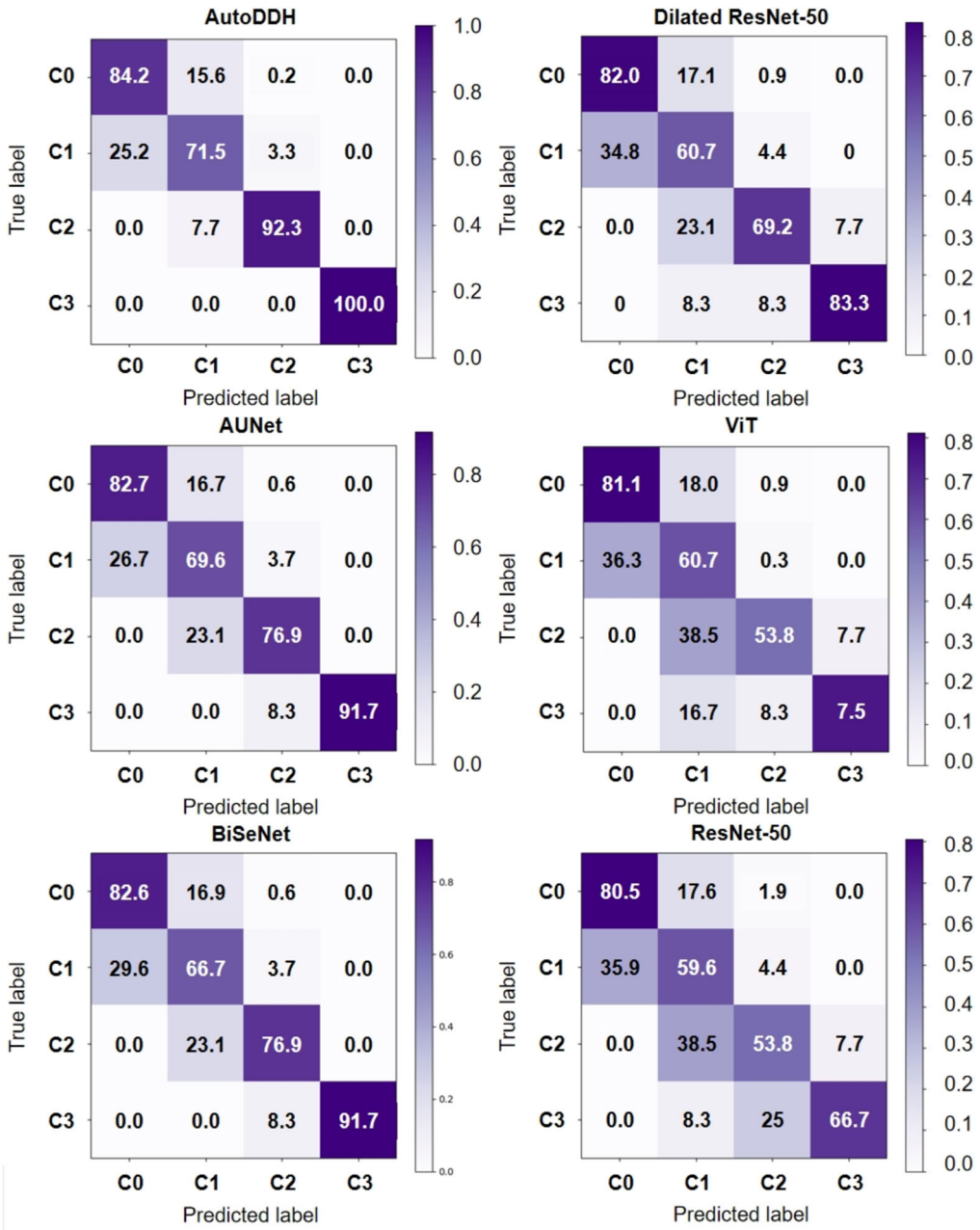

Fig. 8 The confusion matrixes of top-6 accuracy models, including AutoDDH, AUNet [28], BiSeNet [27], dilated ResNet-50 [23], ViT [26], and ResNet-50 [24]. C0: normal, C1: mild DDH, C2: severe DDH, and C3: hip dislocation

other methods (as shown in Fig. 9). It can also be seen from Fig. 9 that AutoDDH has a good detail segmentation ability.

4.4.3 Ablation study

In order to find out the contribution of each module and training strategy in AutoDDH to the improvement of model accuracy, we conducted sufficient ablation experiments to explore. To do this, we conducted two experiments. The first experiment explored how TTS helped improve classi-

Table 3 Comparison segmentation performance of our AutoDDH and other state-of-the-art methods, including BiSeNet [27], AUNet [28], and UNet [29], in two diagnosis-related key structure segmentations

Model	DSC (%)		
	Labrum	bony roof	average
BiSeNet	80.92%	86.74%	83.83%
UNet	84.39%	87.44%	85.92%
AUNet	83.29%	88.10%	85.69%
Ours (AutoDDH)	**85.05%**	**89.48%**	**87.27%**

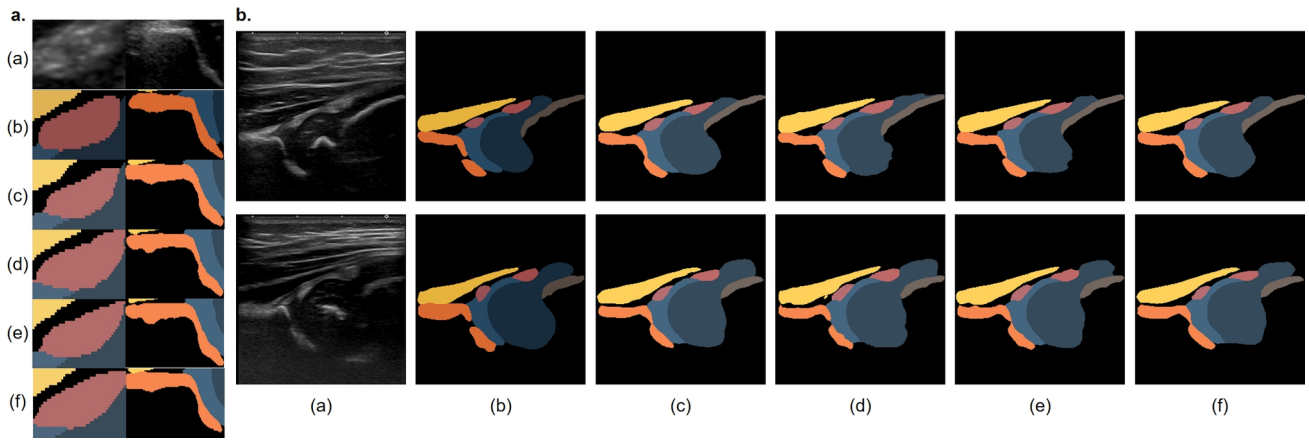

Fig. 9 Segmentation results of BiSeNet [27], **AUNet** [28], **UNet** [29], **and AutoDDH compared with ground truth. a.** An example of segmentation details in two diagnosis-related key structure segmentations. **b.** Two examples of segmentations in seven key structures. (a) ultrasound images, (b) ground truth, (c) segmentations of UNet, (d) segmentations of AUNet, (e) segmentations of BiSeNet, and (f) segmentations of AutoDDH

Table 4 Classification comparison shown in ablation study results, including five kinds of experiments. "Baseline" shows the evaluation metrics for the original dilated ResNet-50 model

Model	accuracy (%)	re (%)	F1 (%)	κ
DilatedResNet50	74.88%	73.82%	69.03%	0.4815
Baseline + DAM	77.90%	77.74%	73.82%	0.5453
Baseline + DAM + FFM	78.38%	81.73%	75.08%	0.5590
Baseline + DAM + FFM +PE (AutoDDH)	80.43%	87.01%	81.17%	0.5994

"Baseline + DAM" represents the evaluation metrics for the dilated ResNet-50 backbone which added a DAM module. "Baseline + DAM + FFM" represents the evaluation metrics for the dilated ResNet-50 backbone which added a DAM and FFM modules. "Baseline + DAM + FFM + PE" illustrates the evaluation metrics for the dilated ResNet-50 backbone which added DAM, FFM and position embedding

fication accuracy.Table 4 shows the changes of classification evaluation indicators with and without TTS model. As can be seen from Fig. 10, using TTS can significantly improve the classification performance of the model. Based on TTS, AutoDDH's accuracy in classification increased by 2.90%, recall rate by 5.59%, F1 score by 6.22%, and κ value by 5.87%. This results show that according to the diagnostic logic, the method of phased training can better integrate relevant information and improve the performance of classification.

Further, we carried out the second experiment. In experiment 2, the influence of each module in AutoDDH on performance after TTS is adopted is explored. The experimental results are shown in Table4. It can be seen from the experimental results that all the modules have improved the results after joining. Through DAM, the accuracy rate increased by 0.72%, recall rate by 2.29%, F1 score by 3.03%, and κ value by 1.24%. Through FFM, the accuracy rate was increased by 0.48%, recall rate by 3.99%, F1 score by 1.26%, and κ value by 1.37%. Through PE, the accuracy rate increased by 2.05%, recall rate by 5.28%, F1 score by 6.08%, and κ value by 4.04%. It can be seen from the experimental results that the introduction of position embedding

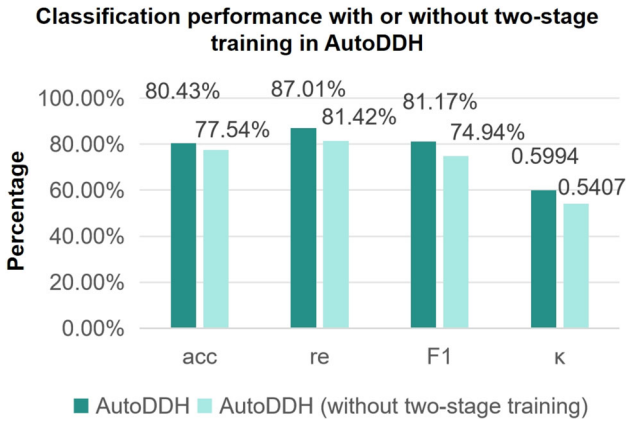

Fig. 10 Bar chart for comparison of classification performance with or without two-stage training strategy in AutoDDH network

makes the classification result improve most obviously. This also supports the view that the introduction of location features can make the model obtain more accurate predictions in classification.

5 Discussions and conclusions

In this study, we developed a dual-attention-based multi-task network and the corresponding two-stage training method (AutoDDH). AutoDDH introduces a dual-attention module that enhances feature recognition, a feature fusion module for comprehensive analysis, and a position coding module for spatial awareness, leading to superior diagnostic accuracy and segmentation of key structures in DDH. This system not only outperforms existing models but also holds high clinical value by enabling accurate visual DDH ultrasound diagnosis from standard images, which is crucial for early intervention and treatment planning.

The two-stage training method of AutoDDH ensures the network's adaptability and generalization, which is essential for clinical use. The potential integration of AutoDDH into clinical procedures could streamline the screening process for DDH, aiding doctors in making quicker and more accurate diagnoses. This could lead to earlier interventions, improved patient outcomes, and reduced healthcare burdens.

While AutoDDH shows promise, challenges such as clinical validation, image quality variability, and system robustness against ultrasound video must be addressed. Future research will focus on overcoming these to ensure AutoDDH's reliability and safety in diverse clinical settings. In conclusion, AutoDDH represents a great advancement in AI-aided DDH diagnosis, with the potential to revolutionize patient care through automation and accuracy. Continued development is necessary to fully realize its clinical potential.

Acknowledgements This work was supported by the National Natural Science Foundation of China (62402530, 82171936), the Shanghai Jiao Tong University "Jiao Tong University Star" plan key Project of the Medical and Industrial Cross (YG2022ZD007), the Program of Shanghai Academic/Technology Research Leader (23XD1403100), the Science and Technology Commission of Shanghai Municipality (23DZ2202200), and the Key Discipline Construction Project of Jiading District Health System (XK202403).

Author Contributions M. L. and R. L. were responsible for drafting the main manuscript text, while Y. Z. and L. J. oversaw the research activities. Data collection and method implementation were conducted by R. L., M. L., J. S., and Q. L., with all authors participating in the review of the manuscript.

Data availability statement No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Schaeffer, E. K., Ihdi Study Group, Mulpuri, K.: Developmental dysplasia of the hip: Addressing evidence gaps with a multicentre prospective international study. Med. J. Aust., **208**(8), 359–364 (2018)
2. Chan, A., McCaul, K.A., Cundy, P.J., Haan, E.A., Byron-Scott, R.: Perinatal risk factors for developmental dysplasia of the hip. Arch. Dis. Child. Fetal Neonatal Ed. **76**(2), F94-100 (1997)
3. O'Beirne, J.G., Chlapoutakis, K., Alshryda, S., Aydingoz, U., Baumann, T., et al.: International interdisciplinary consensus meeting on the evaluation of developmental dysplasia of the hip. Ultraschall Med., **40**(4), 454–464 (2019)
4. Graf, R., Mohajer, M., Plattner, F.: Hip sonography update quality-management, catastrophes - tips and tricks. Med. Ultrason. **15**(4), 299–303 (2013)
5. Aarvold, A., Perry, D.C., Mavrotas, J., Theologis, T., Katchburian, M., et al.: The management of developmental dysplasia of the hip in children aged under three months: a consensus study from the British society for children's orthopaedic surgery. Bone Joint J. **105–B**(2), 209–214 (2023)
6. Huang, Shan, Liu, Xiaohong, Tan, Tao, Menghan, Hu., Wei, Xiaoer, et al.: Transmrsr: transformer-based self-distilled generative prior for brain MRI super-resolution. Vis. Comput. **39**(8), 3647–3659 (2023)
7. Al-Jebrni, A.H., Ali, S.G., Li, H., Lin, X., Li, P., et al.: Sthy-net: A feature fusion-enhanced dense-branched modules network for small thyroid nodule classification from ultrasound images. Vis. Comput. **39**(8), 3675–3689 (2023)
8. Dai, Ling, Liang, Wu., Li, Huating, Cai, Chun, Qiang, Wu., et al.: A deep learning system for detecting diabetic retinopathy across the disease spectrum. Nat. Commun. **12**(1), 3242 (2021)
9. Gong, Bangming, Shi, Jing, Han, Xiangmin, Zhang, Huan, Huang, Yuemin, et al.: Diagnosis of infantile hip dysplasia with b-mode ultrasound via two-stage meta-learning based deep exclusivity regularized machine. IEEE J. Biomed. Health Inform. **26**(1), 334–344 (2022)
10. Chen, Yueh-Peng., Fan, Tzuo-Yau., Chu, Cheng-Cj., Lin, Jainn-Jim., Ji, Chin-Yi., et al.: Automatic and human level Graf's type identification for detecting developmental dysplasia of the hip. Biomed. J. **47**(2), 100614 (2024)
11. Graf, R.: The diagnosis of congenital hip-joint dislocation by the ultrasonic combound treatment. Arch. Orthop. Trauma Surg. **97**(2), 117–133 (1980)
12. Kolb, Alexander, Chiari, Catharina, Schreiner, Markus, Heisinger, Stephan, Willegger, Madeleine, et al.: Development of an electronic navigation system for elimination of examiner-dependent factors in the ultrasound screening for developmental dysplasia of the hip in newborns. Sci. Rep. **10**(1), 16407 (2020)
13. Komatsu, Masaaki, Sakai, Akira, Dozen, Ai., Shozu, Kanto, Yasutomi, Suguru, et al.: Towards clinical application of artificial intelligence in ultrasound imaging. Biomedicines **9**(7), 720 (2021)
14. Mabee, M.G., Hareendranathan, A.R., Thompson, R.B., Dulai, S., Jaremko, J.L.: An index for diagnosing infant hip dysplasia using 3-D ultrasound: the acetabular contact angle. Pediatr. Radiol. **46**(7), 1023–1031 (2016)
15. de Luis-García, R., Alberola-López, C.: Parametric 3D hip joint segmentation for the diagnosis of developmental dysplasia. Conf. Proc. IEEE Eng. Med. Biol. Soc **2006**, 4807–4810 (2006)
16. Quader, N., Hodgson, A.J., Mulpuri, K., Schaeffer, E., Abugharbieh, R.: Automatic evaluation of scan adequacy and dysplasia metrics in 2-D ultrasound images of the neonatal hip. Ultrasound Med. Biol. **43**(6), 1252–1262 (2017)

17. Pandey, P.U., Quader, N., Guy, P., Garbi, R., Hodgson, A.J.: Ultrasound bone segmentation: A scoping review of techniques and validation practices. Ultrasound Med. Biol. **46**(4), 921–935 (2020)
18. El-Hariri, Houssam, Mulpuri, Kishore, Hodgson, Antony J., Garbi, Rafeef: Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound. In Proc. MICCAI **11765**, 12–20 (2019)
19. Liu, Ruhan, Liu, Mengyao, Sheng, Bin, Li, Huating, Li, Ping, et al.: NHBS-Net: A feature fusion attention network for ultrasound neonatal hip bone segmentation. IEEE Trans. Med. Imaging **40**(12), 3446–3458 (2021)
20. Sezer, A., Sezer, H.B.: Deep convolutional neural network-based automatic classification of neonatal hip ultrasound images: A novel data augmentation approach with speckle noise reduction. Ultrasound Med. Biol. **46**(3), 735–749 (2020)
21. Wang, Junxia, Zheng, Yuanjie, Ma, Jun, Li, Xinmeng, Wang, Chongjing, et al.: Information bottleneck-based interpretable multitask network for breast cancer classification and segmentation. Medical Image Anal. **83**, 102687 (2023)
22. Zhang, Xin, Han, Liangxiu, Sobeih, Tam, Han, Lianghao, Dempsey, Nina, et al.: Cxr-net: A multitask deep learning network for explainable and accurate diagnosis of COVID-19 pneumonia from chest x-ray images. IEEE J. Biomed. Health Inform **27**(2), 980–991 (2023)
23. Yu, F., Koltun, V., Funkhouser, T. A.: Dilated residual networks. In: Proceedings of IEEE CVPR, pp. 636–644 (2017)
24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of IEEE CVPR, pp. 770–778 (2016)
25. Huang, Gao, Liu, Zhuang, van der Maaten, Laurens, Weinberger, Kilian Q.: Densely connected convolutional networks. In Proc. IEEE CVPR, pages 2261–2269 (2017)
26. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: Proc, ICLR (2021)
27. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., et al.: Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: Proceedings ECCV, volume 11217 of Lecture Notes in Computer Science, pages 334–349 (2018)
28. Sun, Hui, Li, Cheng, Liu, Boqiang, Liu, Zaiyi, Wang, Meiyun, et al.: AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. Phys. Med. Biol. **65**(5), 055005 (2020)
29. Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas: U-net: Convolutional networks for biomedical image segmentation. In Proc. MICCAI **9351**, 234–241 (2015)

Mengyao Liu is currently pursuing a Ph.D. degree in the Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai. Her current research interests include musculoskeletal ultrasound diagnosis and treatment, medical image analysis, and artificial intelligence ultrasound diagnosis.

Ruhan Liu received her Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai. Currently, she is a lecturer at Central South University, and affiliated with Furong Laboratory. Her research interests include medical image analysis and medical signal processing.

Jia Shu received her B.S. degree in Biomedical Engineering from Nanjing Medical University, Nanjing, China, in 2023. She is currently pursuing a Ph.D. in Computer Science and Technology at Shanghai Jiao Tong University, Shanghai, China. Her research interests include medical data analysis and modeling.

Qirong Liu is currently pursuing a B.S. degree in College of Software, Taiyuan University of Technology. His research interests include sport medicine, computer vision and medical ultrasound image processing.

Yuan Zhang received his M.D. degree in Pediatrics from Fudan University, Shanghai, China, in 2022. Currently, he is a doctor in Department of Ultrasound Medicine, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. His major interests include musculoskeletal diseases, ultrasound evaluation of inflammatory bowel disease and congenital malformation.

Lixin Jiang received a Ph.D. degree from Shanghai Jiao Tong University School of Medicine, Shanghai, China, in 2006. He is currently the director of the Department of Ultrasound Medicine, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. He has been awarded the honorary title of Shanghai Academic/Technology Research Leader and has contributed more than 60 articles and coauthored 12 books. His research interests include sports medicine ultrasound, high-intensity focused ultrasound (HIFU) treatment of tumors, ultrasound diagnosis of endocrine and metabolic diseases, ultrasound evaluation of rheumatic immune diseases and artificial intelligence in ultrasound medicine.