World Wide Web (2023) 26:539–559
https://doi.org/10.1007/s11280-022-01051-0

Multi-task hourglass network for online automatic diagnosis of developmental dysplasia of the hip

Jingyuan Xu[1] · Hongtao Xie[1] · Qingfeng Tan[2] · Hai Wu[1] · Chuanbin Liu[1] · Sicheng Zhang[3] · Zhendong Mao[1] · Yongdong Zhang[1]

Received: 21 December 2021 / Revised: 14 March 2022 / Accepted: 6 April 2022 /
Published online: 4 May 2022

Abstract
Developmental dysplasia of the hip (DDH) is one of the most common diseases in children. Due to the experience-requiring medical image analysis work, online automatic diagnosis of DDH has intrigued the researchers. Traditional implementation of online diagnosis faces challenges with reliability and interpretability. In this paper, we establish an online diagnosis tool based on a multi-task hourglass network, which can accurately extract landmarks to detect the extent of hip dislocation and predict the age of the femoral head. Our method utilizes a multi-task hourglass network, which trains an encoder-decoder network to regress the landmarks and predict the developmental age for online DDH diagnosis. With the support of precise image analysis and fast GPU computing, our method can help overcome the shortage of medical resources and enable telehealth for DDH diagnosis. Applying this approach to a dataset of DDH X-ray images, we demonstrate 4.64 mean pixel error of landmark detection compared to the results of human experts. Moreover, we can improve the accuracy of the age prediction of femoral heads to 89%. Our online automatic diagnosis system has provided service to 112 patients, and the results demonstrate the effectiveness of our method.

Keywords Online automatic diagnosis · Developmental dysplasia of the hip · Multi-task hourglass network

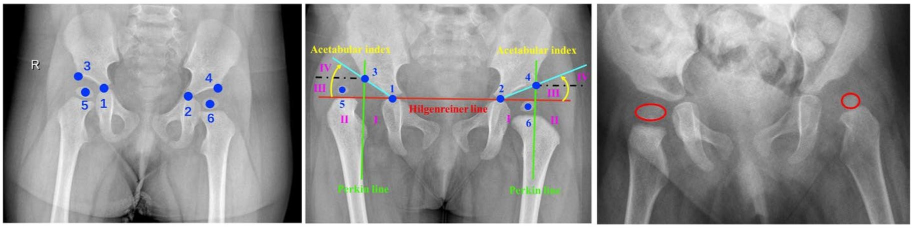
1 Introduction

Medical data analysis provides convenient and high-quality services, which can be used for prescription recommendation, treatment planning, and online diagnosis for many diseases [1–3]. Developmental dysplasia of the hip (DDH) is one of the

✉ Hongtao Xie
htxie@ustc.edu.cn

✉ Qingfeng Tan
tqf528@gzhu.edu.cn

Extended author information available on the last page of the article



(a) DDH image (b) Illustration of DDH clinical diagnosis (c) Delayed development and calcification

Fig. 1 Figure (**a**) shows six landmarks that need to be detected. Landmark 1 6 represent tri-radiate cartilage center (right), tri-radiate cartilage center(right), acetabulum superolateral margin (right), tri-radiate cartilage center (left), tri-radiate cartilage center(left), acetabulum superolateral margin (left) respectively. Figure (**b**) shows the schematic diagram [5] of the clinical DDH diagnosis. We need to detect four landmarks (1, 2, 3, 4) to draw the Hilgenreiner line and Perkin line [5] to divide quadrants shown as I, II, III, IV. When landmark 5 and 6 are detected, the degree of DDH is determined. The yellow arrows denote the acetabular index, which is of significance in diagnosis. Figure (**c**) shows delayed development and calcification of femoral heads due to hip dislocation. The red areas denote the size of femoral heads

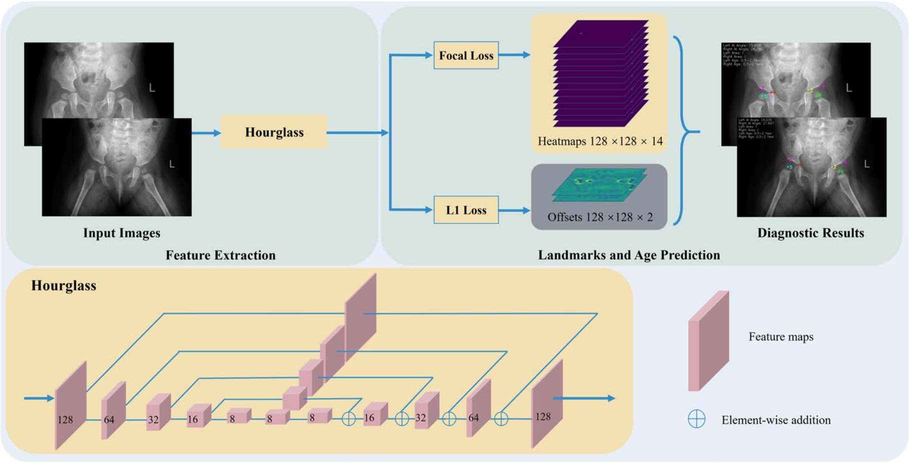


Fig. 2 Illustration of the proposed multi-task hourglass network. The backbone is default hourglass [19]. The overall architecture of this network mainly comprises two components, i.e. the feature extraction section and landmarks and the age prediction section. We use a single network to predict landmarks and the age of femoral heads. Diagnostic results can be seen in the figure

most common skeletal system diseases in infants and children, which attacks with a global incidence ranging from 0.16% to 2.85% [4]. It is a developmental disease in the hip joint and can be diagnosed by observing bone morphology on an X-ray image [5, 6]. Six landmarks on the X-ray image are the key to the quantitative analysis of DDH as show in Figure 1(a),

The acetabular index can be calculated by their included angles based on these landmarks. The rules for evaluating the acetabular index are shown in Figure 2(b), It is a fundamental diagnostic metric for DDH in many respect [7, 8]. It measures the severity of DDH, uses for acetabular modeling, and indicates for DDH treatment [9–12]. In addition, predicting the age of the femoral head is significant. It can provide auxiliary information for the diagnosis of DDH. As we can see in Figure 1(c), the calcification and development of these two femoral heads are different. Based on the development age of the femoral head and the patient's actual age, we can further confirm the condition and give an appropriate treatment plan.

However, due to the shortage of professional doctors to diagnose DDH, many children do not receive timely treatment. Most highly trained experts are generally located in large urban hospitals, while children in rural areas and small cities are difficult to obtain a professional diagnosis [13]. Online automatic diagnosis for DDH has the potential to meet the challenge of insufficient professional treatment by performing medical image analysis. Despite its potential, the reliability and interpretability of online automatic diagnosis remain challenging. The traditional medical image analyzing approach relied on hand-engineered features and multi-step landmark localization. Implementing and improving such an approach requires dedicated design and is computationally expensive [14].

Despite the difficulty in diagnosing DDH automatically and effectively, the development of the convolutional neural network has allowed for training an end-to-end model to meet the challenges of medical images [15–18]. In our paper, we propose a multi-task hourglass network to detect the landmarks and predict the age on the hip X-ray image [19]. It is an encoder-decoder structure that can extract features effectively and accurately predict developmental age. First and foremost, our team constructs a professional DDH dataset with the experts' annotations include 9369 hip X-ray images. Secondly, we train and evaluate the proposed multi-task hourglass network on our dataset to detect six landmarks and predict the patients' age. Finally, we deploy our method on the GPU server to provide services through the network for the public and continue to collect DDH data. In this way, our approach can realize a stable and interpretable service for online automatic DDH diagnosis. Therefore, our work shows that a multi-task hourglass network can learn the accurate location of the landmarks in the X-ray images and distinguish between different stages of development. Moreover, with precise image analysis and fast GPU computing, our method can help overcome the shortage of medical resources and enable telehealth for DDH diagnosis. The experiments show that our method achieves 4.64 mean pixel error in landmark detection and accuracy of 89% in age prediction. And our online service has successfully helped 112 children with DDH diagnosis.

In summary, the main contributions of this paper are as follows:

- We propose a novel multi-task hourglass network to investigate the online automatic diagnosis for DDH. Our method achieves accurate landmark detection and age prediction on X-ray images with the powerful feature extracting capabilities and the end-to-end framework, .
- To overcome the shortage of medical resources, we build an online DDH diagnosis platform based on the GPU server. To the best of our knowledge, it is the first work that provides an online service for automatic DDH diagnosis.
- We construct a DDH database with a considerable amount of X-ray images to train and evaluate our model. The success of online diagnosis demonstrates that our model can effectively learn the knowledge from the database.

2 Related work

2.1 Developmental dysplasia of the hip

In 1985, Tonnis first proposed the current standard method to diagnose DDH on an X-ray image [5]. The core of Tonnis's approach is detecting six landmarks (see Figure 1(a)) to estimate the dislocation degrees (see Figure 1(b)) of DDH. According to these landmarks, Hilgenreiner first described the acetabular index in 1986. It provides a direct measurement referred to as the acetabular index angle to diagnose DDH [6].

Many researchers have studied automatic diagnosis methods due to the high incidence of DDH in children [20–23]. AI-Bashir et al. utilize Canny edge detection and Hough transform to detect the landmark on the X-ray image [20]. However, their methods are struggling to detect landmarks among different ages. Sahin et al. propose an object detection methods which uses the landmark as a template to match the target image [21]. However, this method requires massive computation because of the patch-by-patch searching for the best fitting patch. Liu et al. [22] and Bier et al. [23] utilize convolutional neural network to detect hip landmarks on the X-ray image. Yet, they ignore the significance of the age of the femoral head to assist DDH diagnosis. Therefore, the above DDH analysis methods are insufficient to realize an online DDH diagnosis which requires accurate and adequate analysis of the hip X-ray image.

In our work, we propose an end-to-end multi-task hourglass network to diagnose DDH effectively and provide service online, thereby realizing an earlier treatment for improved clinical outcomes.

2.2 Online diagnosis

The advances in computer networking enable online diagnosis as an alternative for medical treatment [13, 24–28]. Sublett et al. propose a distribution system that supports remote ultrasound examinations [13]. They achieve a real-time diagnosis under the limited bandwidth. Hollander et al. suggest that telemedical innovations play a pivotal role in medical service especially during the Covid-19 pandemic [25]. This is because online automatic diagnosis can not only save medical resources but also avoid overexposure. Esteva et al. propose a deep-learning-based method to diagnose skin cancer [27]. They deploy their method on a mobile device to extend the reach of dermatologists outside of the clinic. Long et al. build the AI system with a cloud-based platform for multihospital collaboration [28]. They design the platform to help patients with rare diseases and improve disease management for different hospitals. Their method demonstrates the capabilities of an online automatic diagnosis system to help both the patients and the hospitals.

To sum up, online diagnosis provides patients with a convenient and effective solution for health care. Our work proposes a novel multi-task hourglass network. It facilitates the online service with a GPU server to automatically diagnose DDH, the first work opened for public service.[1]

[1] Online service can be found at http://202.38.69.241:30128/ddh.php

2.3 Multi-task hourglass network

Recent years have witnessed the progress of the deep neural network, which has triggered significant changes in the medical field [3, 29–34]. Newell et al. propose an hourglass network with an encoder-decoder architecture to generate the heatmap for landmark detection [19]. The hourglass network has been successfully applied to many areas such as pose estimation, face recognition and object detection [35–37]. Cai et al. utilize a multi-task hourglass network to expand receptive fields and reconstruct a frame for fall detection [31]. Their method demonstrates the multi-task hourglass network can complete the detection task in complex scenes. Kordon et al. propose a deep-stacked multi-task hourglass network to jointly localize landmarks, predict a tangent line, and perform segmentation on the X-ray image [34]. Their work indicates that the multi-task hourglass network has sound performance in exploring bone morphology.

Here we propose a novel method based on the hourglass network, which can accurately detect landmarks to diagnose the degree of hip dislocation and predict the age of femoral head development. With a large-scale DDH dataset collected from Anhui Provincial Children's hospital, the hourglass network can be adequately trained and appropriately evaluated for diagnosing DDH. Meanwhile, the online service can continue to collect extensive DDH data to asses and refine the multi-task hourglass network.

3 Method

3.1 Ethical approval statement

The Ethics Review Committee of Anhui Province Children's Hospital of China has passed the review of this project. We have submitted the ethics review report as Support Documents. The ethics review report can be found in supplementary material. All data and images used in this study have been desensitized. These data and images used for research do not contain any patient's private information and do not contain information showing the patient's identity. Moreover, these images and data are only used for academic research. This statement was presented by Yongdong Zhang and Jun Sun, who are the directors of this research at the University of Science and Technology of China, and director of the Ethics Committee of Anhui Province Children's Hospital of China, respectively.

3.2 Confirming statement of parents or guardians

We have received written consent from parents or guardians of these children who participated in this study. These data and images used in this study have been desensitized, and this study does not reveal any private information about children. These data and images for research are not intended for any commercial activity.

3.3 Multi-task hourglass network

We propose a novel multi-task hourglass network for online automatic DDH diagnosis. Figure 2 illustrates the overall framework for landmark detection and age prediction.

3.3.1 Network architecture

The proposed hourglass network is an encoder-decoder architecture, as shown in Figure 2. We follow the standard implementation of the hourglass network for the feature extraction section. In specific, to build a multi-task hourglass network, we first utilize the convolutional block followed by the pooling layer to extract the features from images [38]. After four convolutional blocks and pooling layers, we obtain a feature map in a small shape with rich information:

$$f_{i+1} = Pool(Conv(f_i)) \tag{1}$$

where f_i denotes the features of the i-th stage in the encoder, *Conv* and *Pool* represent the convolutional block and pooling layer. The stacked convolutional layers enlarge the reception field of the model and enable detection from a large perspective [39]. Then, we follow the network engineering of the hourglass network and fully convolution network [19, 40], the features are unpsampled and added to the output of former layers by skip connection:

$$\hat{f}_{i+1} = Up(Conv(\hat{f}_i)) + Conv(f_{i+1}) \tag{2}$$

where $\hat{f}_i$ represents the features of the i-th stage in the decoder, *Up* denotes the upsampling layer. In this way, the model can combine the rich context information in the shallow layers with the adequate semantics in the deep layers for landmark detection. In the final stage, we use 1×1 convolutional layer to generate output for DDH diagnosis.

In our proposed multi-task hourglass network, there are three types of the output. First of all, the multi-task hourglass output the heatmap of the landmarks 1 to 4. The heatmaps indicate the probability of the location of these landmarks. Secondly, the model output the heatmap of the femoral head. Since the age of patients has an enormous impact on the morphology of the femoral head, we use four different heatmaps to represent each femoral head in different age periods. Thus, the output of these heatmaps could indicate the age of the patients by the highest value. The third type of output is six offset maps for revising the location of the landmarks. Such creation can refine the location results by learning the difference between the heatmap and the ground-truth landmark.

3.3.2 Heatmap estimation

Based on the multi-task hourglass network, we utilize the heatmap to simultaneously predict the landmark and the age of the femoral head on the DDH X-ray image. Thus, heatmap estimation plays a pivotal role in landmark detection and age prediction. Suppose the input image is $I \in R^{W \times H}$. The proposed multi-task hourglass network predicts the heatmap $\hat{Y}$ from an image I. The following equation is used to generate the target heatmap according to the position of the ground truth landmark, we utilize the Gaussian kernel as following:

$$Y_{xyc} = \exp\left(-\frac{(x - l_{xc})^2 + (y - l_{y_c})^2}{2\sigma^2} \right) \tag{3}$$

where σ is an size-adaptive standard deviation, Y_{xyc} represnets the heatmap of c-th channel, $l_x c$ and $l_y c$ represent the coordinate of the c-th landmark. If these Gaussian labels have overlaps, we select the element-wise maximum $M_{xyc} = \max_{c=1,2,\ldots,C} Y_{xyc}$. As mentioned above, we divide the left femoral head into four categories (0.1-2, 2-6, 6-12 and > 12) according to age information. We divided the patients into four groups based on the morphological characteristics of the femoral head at different ages. To decide the age period of the input X-ray image, the network output four heatmaps for landmarks 5 and 6 (the femoral head) to represent the age periods. The heatmap with the maximum value is the result of the age period. The same operation is applied to the right femoral head. Other landmarks (1, 2, 3, 4) are 4 categories. Thus, the model output 12 different heatmaps for each X-ray image. The proposed model predicts the landmarks by generating $\hat{Y}$ and offset $\hat{O}$. The network can predict different locations and categories of landmarks. We can use the predicted category of landmark to obtain the age of the femoral head.

When our trained model diagnoses a new X-ray image, we extract the peaks in the heatmap for each category to obtain landmarks. We use the categories of these predicted peaks to obtain ages. All outputs are produced directly from the landmark estimation without the need for Non-Maximum Suppression or other post-processing. When using the trained model to detect landmarks and predict ages in an X-ray DDH image, the results are consistent and stable. In the end, we can derive the predicted landmark from the heatmap:

$$< \hat{l}_x, \hat{l}_y >= \arg\max_{x,y}(\hat{Y}) + < \hat{O}_x, \hat{O}_y > \tag{4}$$

where $\hat{l}_x, \hat{l}_y$ represent the predicted landmark. And the age of femoral head can be derived by:

$$A = \arg\max_{c}(\hat{Y}) \tag{5}$$

where A indicates the age range.

3.3.3 Loss function

The loss function plays a pivotal role in optimizing our model. Our model is essentially a one-stage detector. Therefore, we utilize focal loss [41] to optimize the detector and generate the target heatmap in different bone morphologies. The training loss can be formulated as follows:

$$L_l = -\frac{1}{N} \sum_{xyc} \psi_{xyc} (1 - \hat{Y}_{xyc})^{\alpha} \log(\hat{Y}_{xyc}, \tag{6}$$

where

$$\hat{Y}_{xyc} = \begin{cases} \hat{Y}_{xyc} & \text{if} \quad Y_{xyc} = 1 \\ 1 - \hat{Y}_{xyc} & \text{otherwise} \end{cases} \tag{7}$$

and

$$\Psi_{xyc} = \begin{cases} 1 & \text{if} \quad Y_{xyc} = 1 \\ (1 - M_{xyc})^{\beta} & \text{otherwise.} \end{cases} \tag{8}$$

The hyper-parameters (α and β) are adopted in the focal loss [41], and N is the number of landmarks in an image I. We expect N to be 6. We adopt α =2 and β =4 [37, 42] in all experiments.

To recover the error caused by down sampling, we predict a local offset $\hat{O}$ for each landmark. We adopt L1 loss to optimize the offset map:

$$\mathrm{L}_o = \frac{1}{N} \sum_{l} \left[\hat{O}_{\tilde{l}} - \left(\frac{l}{S} - \tilde{l} \right) \right]_1 \tag{9}$$

Finnaly, the overall training loss of the model is

$$L = L_l + \lambda_o L_o. \tag{10}$$

where λ_o is a hyper-parameter to adjust the value of L1 Loss.

4 Experimental results and discussion

We test our trained model on the test dataset with 2119 DDH images and conduct a series of experiments to elaborate on the results of our method. These results contain predictions of landmarks, acetabular index angle, and age in detail.

4.1 Experimental setup

We apply the proposed method to the DDH dataset for landmark detection and age prediction. The network is trained using the Pytorch framework [43] on a Ubuntu workstation equipped with an Intel i7-9700 CPU and an 11GB Nvidia GeForce 1080Ti GPU. During training, the mini batch size is set to 4. Adam grad optimizer is used for updating with the learning rate of 1.25e-4 [44, 45]. We set the total epoches in training as 30.

4.2 Dataset

Our dataset is collected in the clinical routine of Anhui Province Children's Hospital of China and contains all common conditions in clinical cases in 2012-2019. The original medical images were collected from the PACS system of Anhui Province Children's Hospital of China. We converted the original DICOM files to JPG files for labeling. And we extracted children's actual age information from these DICOM files. All landmarks of dataset are labeled by fifteen professional doctors. These doctors are divided into three groups to label landmarks. Three of five doctors in each group label images and the other two doctors check images labeled by these three doctors. If there is an inaccurately labeled image among these three doctors, the image will be re-labeled until it passes the check process. In this way, each image has been labeled three times, and we calculated the average of coordinates of each landmark.

Patients are between 0.1-14 years old. The total number of DDH images is 9369, in which 7250 images are used for training, and the rest 2119 images are for testing.

Table 1 Distribution of test set (9369 images) in our database

	0.1-2 year	2-6 year	6-12 year	> 12 year
Left I	7352	1344	421	25
Left II	127	24	1	0
Left III	26	30	3	2
Left IV	3	11	0	0
Right I	7377	1356	419	25
Right II	96	21	1	0
Right III	23	23	5	2
Right IV	12	9	0	0

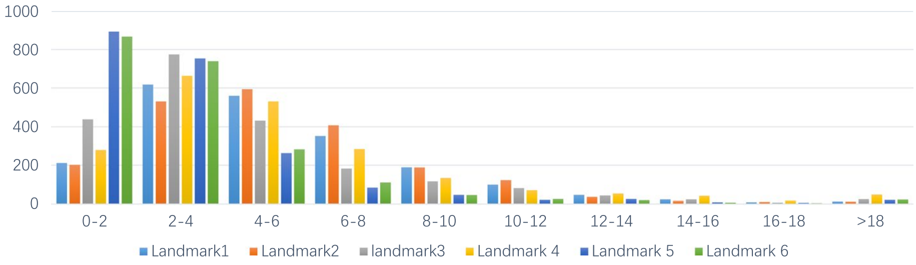


Fig. 3 Distribution the pixel error. The X-axis denotes landmarks mean pixel error interval of all landmarks in an image. The Y-axis represents the number of images. Six different colors denote six landmarks. *Best-viewedincolor*

The distributions of the training and test dataset are shown in Table 1. Because we need to learn the relationship between age and development of femoral heads, the femoral heads of the training dataset are all in the I quadrants. During the training process, we resize the input resolution to 512 × 512. At inference time, we recover the output to the original size to statistically analyze behaviors.

4.3 Landmark detection

We analyze pixel error of these detected six landmarks on the entire test dataset with 2119 DDH X-ray images. The calculation method of pixel error is $error = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. (x_i,y_i) denotes landmarks labeled by doctors, and (x_j,y_j) denotes landmark predicted by our network. As shown in Figure 3, the horizontal axis represents pixel error, and we divide the pixel error range into 10 intervals (0-2, 2-4, 4-6, 6-8, 8-10, 10-12, 12-14, 14-16, 16-18, > 18 pixels, We can see in the figure that the pixel error mainly concentrates in 0-2, 2-4, 4-6 intervals. In addition, we can also see that the detection results of landmark 5 and landmark 6 are more concentrated in the low error interval (0-2, 2-4) than landmarks 1 to 4, which indicates that the overall detection accuracy of landmark 5 and landmark 6 are higher than other landmarks. In specific, landmarks 1 to 4 have pixel errors around five while landmarks 5 and 6 have smaller pixel errors. It is because landmarks 5 and 6 locate at the center of the femoral head, which is more distinguishable. However, landmarks 5 and 6 have greater SD for localization. This is because

Table 2 Mean±SD and median error of landmark (1, 2, 3, 4, 5, 6) detection in pixels

	Landmark 1		Landmark 2		Landmark 3		Landmark 4		Landmark 5		Landmark 6	
	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median
Overall	5.5 ± 3.5	4.8	5.6 ± 3.6	5.0	4.6 ± 3.7	3.6	5.5 ± 4.3	4.4	3.4 ± 9.6	2.3	3.4 ± 7.1	2.4
Left I	5.6 ± 3.2	5.0	5.7 ± 3.1	5.2	4.3 ± 3.3	3.4	5.0 ± 3.5	4.2	3.4 ± 10.0	2.3	3.3 ± 7.4	2.3
Left II	3.6 ± 2.3	3.0	3.5 ± 2.0	3.4	6.9 ± 5.4	5.6	8.8 ± 5.5	7.0	2.8 ± 2.3	2.3	4.2 ± 4.8	2.9
Left III	5.7 ± 9.1	3.2	7.7 ± 10.5	4.3	7.1 ± 5.8	5.2	11.3 ± 9.5	8.0	4.6 ± 7.8	2.4	4.4 ± 5.3	2.8
Left IV	4.4 ± 2.2	4.7	3.3 ± 1.9	3.5	7.054 ± 4.0	5.7	9.0 ± 4.5	10.8	1.7 ± 0.9	1.6	2.9 ± 1.1	2.8
Right I	5.5 ± 3.2	4.9	5.7 ± 3.4	5.2	4.1 ± 3.1	3.4	5.3 ± 4.0	4.3	3.1 ± 3.1	3.2	3.3 ± 3.3	2.4
Right II	4.0 ± 2.9	3.5	3.8 ± 3.0	3.5	8.4 ± 5.9	7.6	6.5 ± 4.8	5.3	3.9 ± 7.6	2.2	3.2 ± 2.7	2.6
Right III	7.1 ± 9.6	5.0	5.2 ± 7.3	3.6	8.6 ± 5.8	6.9	7.8 ± 5.1	7.3	3.3 ± 3.2	2.0	4.1 ± 4.8	2.7
Right IV	4.9 ± 2.3	4.4	4.1 ± 3.0	4.1	10.1 ± 7.6	8.2	10.1 ± 6.5	9.3	21.6 ± 88.2	1.8	16.1 ± 62.3	1.9
0.1-2 Year	5.1 ± 2.9	4.6	5.3 ± 2.9	4.8	4.1 ± 3.3	3.3	4.8 ± 3.4	4.0	2.7 ± 10.7	2.0	2.8 ± 7.8	2.0
2-6 Year	5.6 ± 3.4	4.9	5.5 ± 3.2	5.1	5.5 ± 4.1	4.3	6.9 ± 5.6	5.2	4.4 ± 5.2	3.3	4.3 ± 3.4	3.6
6-12 Year	8.0 ± 4.8	7.2	8.1 ± 6.0	6.7	7.0 ± 4.7	6.0	8.3 ± 5.3	7.3	7.1 ± 5.1	5.9	6.8 ± 4.9	5.7
> 12 Year	19.1 ± 18.8	8.8	18.0 ± 17.1	11.7	7.6 ± 6.2	5.7	10.4 ± 7.8	6.8	15.2 ± 7.9	12.4	17.5 ± 8.9	18.3

The *lmk* in the table denotes landmark. Overall denotes the result of all the 2119 images. The Left II, Left III, Left IV, Right I, Right II, Right III, Right IV denote different types of DDH. 0.1-2 Year in table represents sub-dataset which only contains images whose real ages are in 0.1-2 years. The similar explanation for other years intervals the femoral heads have different morphological shapes at different ages, which could affect the detection of the landmark.

In the test dataset with 2119 images, based on the label of fifteen professional doctors, there are 1892 images whose left femoral head center (landmark 6) locates in the I quadrant. There are 152 images whose landmark 6 are in the II quadrant. There are 61 images whose landmark 6 are in the III quadrant, and 14 landmark 6 are in the IV quadrant. For the right femoral head centers (landmark 5), 1927 images are in the I quadrant, 118 in the II quadrant, 53 in the III quadrant, and 21 in the IV quadrant. We compare pixel error in these different sub-test datasets, and the results are in Table 2. The comparison results included Mean±SD and median of pixel error. We can see that the average pixel errors of these six landmarks are 5.5, 5.6, 4.6, 5.5, 3.4, and 3.4, respectively. The overall pixel error is relatively low for these different sub-test datasets.

The test dataset with 2119 images can be roughly divided into four categories based on the age information collected from the Anhui Province Children's Hospital of China. These results of landmark detection in different ages are in Table 2. On the whole, the pixel error of older age will be slightly larger. This is because as the age becomes larger, mainly at 6-12 and > 12, the degree of bone calcification becomes larger. The feature richness of landmarks in the image is reduced with higher age, thereby reducing the accuracy of landmark detection. Especially for images older than 12, pixel errors of landmark 1 and landmark 6 become larger, which is caused by two following reasons. The first is that degree of bone calcification becomes deeper. The second is that the amount of images older than 12 is small in the training and testing dataset.

4.4 Acetabular index detection

Based on the landmarks predicted by the proposed network, we calculated the left and right acetabular index angles for every image in all test dataset with 2119 images. The results are shown in Table 3. The average error between the predicted and labeled left acetabular index angle is 2.759°, and the median error is 2.175°. The average error between the predicted and the labeled right acetabular index angle is 2.793°, and the median error is 2.211°. We analyze the error of the acetabular index angle prediction in detail, as shown in Figure 3. It suggests that the angle error mainly concentrates at 0-2°, 2-4°, and 4-6°. We also separately count the images of left femoral head center (landmark 6) at I, II, III, and IV, respectively. The number of images is 1892, 152, 61, and 14, respectively. The average and median error of these four types of images are shown in Table 3. Similarly, we also analyze angle error on images whose center of the right femoral head (landmark 5) locate in I, II, III, and IV.

As children growing, their hips continue to develop, and the degree of bone calcification will gradually deepen. Children of different ages have different acetabular index angles. In order to detail the performance of angle prediction by the designed network, we analyze the statistical acetabular index angle error of different ages in the test dataset, as shown in Table 4. Overall, for children of different ages, the error between predicted and labeled acetabular index angle is about 3° to 4°, which is relatively stable.

We use all the images in the test dataset for analysis of sensitivity/ specificity. We have unified situations that landmark 5 (landmark 6) is in these quadrants II, III, IV as DDH in left (right) hip, and in quadrant I as non-DDH. The statistical results are as Table 5 and Figure 5.

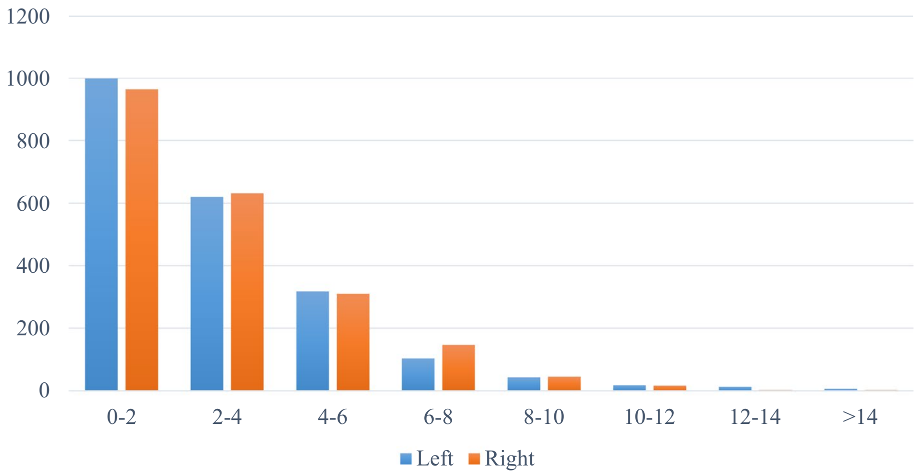


Fig. 4 Distribution of acetabular index angle error of DDH X-ray images. X axis denotes mean acetabular index angle error interval of two acetabular index angles in an image. Y axis represents numbers of images. Two different colors denote left acetabular index and right acetabular index angle. *Bestviewedincolor*

Table 3 Mean±SD and median error of acetabular index angle in left and right hips

	Mean±SD	Median
Left	2.759°±2.486°	2.175°
Left I	2.603°±2.176°	2.094°
Left II	3.398°±2.635°	3.012°
Left III	6.098°±6.062°	3.993°
Left IV	2.402°±1.836°	2.417°
Right	2.793°±2.275°	2.211°
Right I	2.644°±2.111°	2.132°
Right II	4.126°±2.974°	3.637°
Right III	4.821°±3.658°	4.087°
Right IV	3.864°±2.486°	3.010°

Left and right in table denotes left and right acetabular index angle of all test dataset with 2119 images. Left I in table represents sub-dataset which only contains images whose landmark 5 are in I quadrants. The similar explanation for other sub-dataset

Table 4 Mean±SD and median error of acetabular index angle. 0.1-2 Year in table represents sub-dataset which only contains images whose real ages are in 0.1-2 years

	Left Acetabular index		Right Acetabular index	
	Mean±SD	Median	Mean±SD	Median
0.1-2 Year	2.631°±2.138°	2.146°	2.771°±2.227°	2.205°
2-6 Year	2.916°±2.967°	2.088°	2.770°±2.291°	2.151°
6-12 Year	3.452°±3.187°	2.622°	2.946°±2.269°	2.530°
> 12 Year	7.583°±8.801°	4.343°	5.580°±5.930°	2.882°

The similar explanation for other years interval

Table 5 Analysis of sensitivity and specificity on test dataset with 2119 X-ray images

	Sensitivity	Specificity
Left (0.1-2)	69/156	1394/1399
Right (0.1-2)	66/131	1421/1424
Left (2-6)	43/65	352/354
Right (2-6)	35/53	364/366
Left (6-12)	2/4	133/133
Right (6-12)	3/6	130/131
Left (> 12)	0/2	6/6
Right (> 12)	0/2	6/6

Left in table denotes left hip. Right in table denotes right hip. 0.1-2 denotes patients whose age are between 0.1-2 years old. Similar explanations for other age intervals

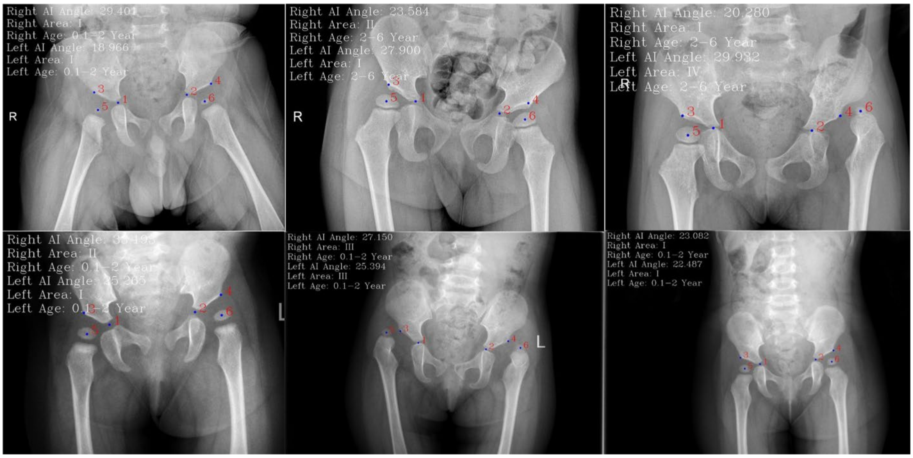


Fig. 5 Example landmark detection results of DDH. Landmarks and number represent the outputs of our method. Text on the left-top of the image is diagnostic report, include the information of left acetabular index angle, Left acetabular index angle, quadrant of landmark 5, quadrant of landmark 6, age of left femoral head and age of left femoral head. *Bestviewedincolor*

4.5 Age prediction

While predicting landmarks in the DDH X-ray image, we also predict the age of the femoral head. Under normal circumstances, the development of bilateral femoral heads in children is roughly the same. However, when the left (right) hip joint dislocates, the development of the left (right) femoral head will lag behind the average developmental age.

By predicting the developmental age of the femoral head, we can analyze whether there is a delay in the femoral head in children. When the ages of the femoral heads on both sides are inconsistent, it can also assist doctors in confirming the situation of hip dislocation. We analyze the accuracy of the prediction of the femoral head ages on the left and right sides in Table 6. When the age of the femoral head predicted by the proposed network is consistent with the actual age, the prediction is considered to be

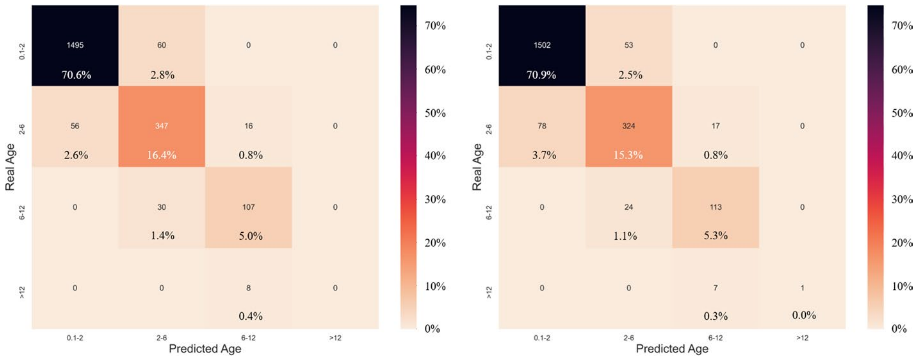


Fig. 6 Distribution of age prediction of femoral heads on the test dataset (2119 images, The left figure indicates the age prediction of left femoral head and vice versa. Numbers in figure denote number of images

Table 6 Accuracy of age prediction for both left and right femoral heads

	Overall	Left I	Left II	Left III	Left IV	Right I	Right II	Right III	Right IV
Left	89.5%	90.6%	84.9%	72.1%	71.4%	90.4%	85.6%	73.6%	66.7%
Right	89.1%	90.5%	86.8%	60.7%	42.9%	89.9%	87.3%	75.5%	66.7%

Numbers in table are expressed as a percentage of accuracy

Table 7 Accuracy of age prediction of left and right femoral heads in different age intervals

Age	0.1-2 Year	2-6 Year	6-12 Year	> 12 Year
left	96.1%	82.8%	78.1%	0.0%
right	96.6%	77.3%	82.5%	12.5%

Numbers in table are expressed as a percentage of accuracy

accurate. From Table 6, we can see that the accuracy of the predicted age of the left femoral head is 89.5%, and the accuracy of the right femoral head is 89.1%.

Similarly, we also analyze the accuracy of the sub-test dataset of the femoral head center in different quadrants. Table 7 shows the accuracy for children of different ages. We can find that the accuracy of age prediction of children aged 0.1-2 is the highest, and the accuracy of the left and right sides is 96.1% and 96.6% respectively. The age prediction accuracy of children aged 2-12 years is not high because the amount of images of this age interval in the training dataset is relatively small. We can see in the table that the accuracy of age greater than 12 years old is 0.0% and 12.5%. The reason is that images whose ages are older than 12 are scarce in the dataset.

To further illustrate the results of our network's age prediction, we compare our predicted age to the actual developmental age of the femoral head. The results of the age prediction of the left and right femoral heads are shown in Figure 6. Taking the prediction of the left femoral head as an example, we can find in Figure 4 that there are 1555 children with true ages between 0.1-2, and the predicted results are 1495 cases within 0.1-2 years old, and 60 images belong to 2-6 years old. To the 419 images with real ages

Table 8 Comparison between our proposed model and the template matching method of landmark (1, 2, 3) [21] on the test dataset (2119 images)

	lmk 1		lmk 2		lmk 3	
	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median
Our Method	5.5 ± 3.5	4.8	5.6 ± 3.6	5.0	4.6 ± 3.7	3.6
Template matching [21]	19.1 ± 79.4	5.4	22.9 ± 103.0	5.5	29.8 ± 100.2	8.5

The error of landmark detection is measured in pixels

Table 9 Comparison between our proposed model and the template matching method of landmark (4, 5, 6) [21] on the test dataset (2119 images)

	lmk 4		lmk 5		lmk 6	
	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median
Our Method	5.5 ± 4.3	4.4	3.4 ± 9.6	2.3	3.4 ± 7.1	2.4
Template matching [21]	45.0 ± 176.6	6.5	37.6 ± 95.4	7.8	33.9 ± 146.4	5.5

The error of landmark detection is measured in pixels

Table 10 Mean±SD and median error of acetabular index detection of our proposed model and template matching method [21] on the test dataset (2119 images)

	Left Acetabular index		Right Acetabular index	
	Mean±SD	Median	Mean±SD	Median
Our Method	2.759°±2.486°	2.175°	2.793°±2.275°	2.211°
Template matching [21]	4.754°±5.164°	3.415°	5.981°±6.503°	4.721°

Error of acetabular index detection is measured in angle

of 2-6 years, 347 images are predicted to be consistent with real age. 56 images are predicted in 0.1-2 years and 16 images are belong to 6-12 years.

4.6 Compared to the traditional method

The traditional landmark detection method is represented by template matching [21]. This method can detect the landmarks, but it cannot predict bone age. Tables 8 and 9 show the template matching method results in landmark detection and acetabular index detection. We can see that our method outperforms the traditional method by a large margin. This phenomenon indicates the effectiveness of deep learning based approach for hip dislocation detection. There are three reasons attributed to this phenomenon. First, multi-task hourglass network uses a large number of convolution kernels to extract features, which captures more details of the image. Second, the traditional method for landmark detection is unstable because the local diversity of the landmarks makes it challenging to describe with simple templates [46]. Third, deep neural networks are nonlinear and optimized with the gradient descent method which can effectively fit the target function (Table 10).

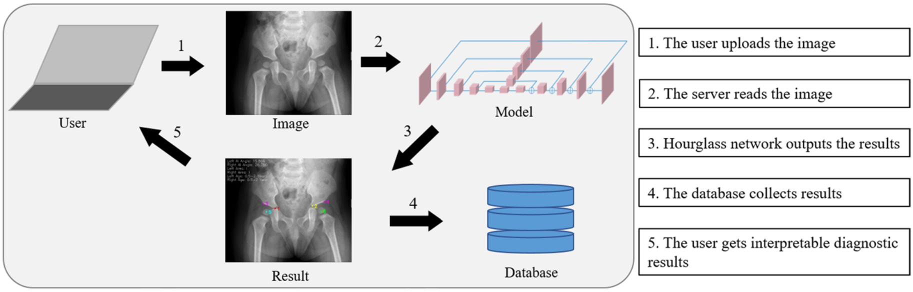

User
1
Image
2
Model
3
5
Result
4
Database
1. The user uploads the image
2. The server reads the image
3. Hourglass network outputs the results
4. The database collects results
5. The user gets interpretable diagnostic results

Fig. 7 Illustration of the online service of our method. When potential patients upload their hip X-ray image for DDH diagnosis, the GPU server is processing the image immediately. The model provides comprehensive evaluation by our multi-task hourglass network and saves the image in the database. If the result from our platform is phase I to IV, a notification is triggered and an a notification is sent to the users. Thus, patients are informed that their condition and get an early diagnosis

Using deep learning to detect landmarks in X-ray images is a new method to solve the challenge of insufficient professional doctors, especially in remote and poor areas. The method based on deep learning to diagnose DDH can help more children in the world. Typically, a professional doctor judges a DDH X-ray for about 5 minutes. And the template matching method consume near 30 seconds to process one image. Our proposed method can detect 20 X-ray images in 1 minute. Based on the deep learning method, the detection speed is very fast, and label errors by different doctors can be reduced.

4.7 Performance of online service

The online service of automatic diagnosis is illustrated in Figure 7. In our online web system, we use a server with the setting of CPU Intel Xeon E5-2630 @ 2.2GHz and GPU Titan XP 12GB. The operating system is Ubuntu 18.04, and apache2 is used to support the web application. The storage of the database is 1TB. The workflow of the online diagnosing is: (1) the user uploads the image to the server via the web address, (2) the server reads the image and sends it to the model, (3) the proposed multi-task hourglass network process the image and output the results, (4) the database collects the image and the results, (5) the user gets the interpretable diagnostic results from the website.

We have collected 112 DDH cases from our website. Noted that all of the collected data have been desensitized and only used for academic purposes. We compare the performance between the images that collected from our web service platform and the images from the test set as shown in Table 11. We can see that our method has a similar performance in both sets. This phenomenon demonstrates that the proposed hourglass network can generalize the new data.

Meanwhile, we illustrate the performance of online diagnosis in Table 12. It shows the results of acetabular index detection of the online diagnosis. The results suggest that our method increased the mean angle error by around 0.3° when measuring online data. Thus, our approach can still perform well in measuring the acetabular index for the online data.

There are two potential social impacts of the proposed services. From a positive point of view, the online system could accelerate the diversification of diagnostic modes. Our system provides another way for patients to learn about their health conditions. From an opposing point of view, online diagnosis can also lead to misdiagnosis and may bear

Table 11 Mean±SD and median error of landmark (1, 2, 3) detection of our proposed model on the test dataset (2119 images) and online diagnosis (112 images)

	Landmark 1		Landmark 2		Landmark 3		Landmark 4		Landmark 5		Landmark 6	
	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median	Mean±SD	Median
Test data	5.5 ± 3.5	4.8	5.6 ± 3.6	5.0	4.6 ± 3.7	3.6	5.5 ± 4.3	4.4	3.4 ± 9.6	2.3	3.4 ± 7.1	2.4
Online data	5.0 ± 3.2	4.3	5.9 ± 3.8	5.0	4.8 ± 3.8	3.8	5.2 ± 4.0	4.0	3.2 ± 3.6	2.4	3.2 ± 2.6	2.5

Error of landmark detection is measured in pixels

Table 12 Mean±SD and median error of acetabular index detection of our proposed model on the test dataset (2119 images) and online diagnosis (112 images)

	Left Acetabular index		Right Acetabular index	
	Mean±SD	Median	Mean±SD	Median
Test dataset	2.759°±2.486°	2.175°	2.793°±2.275°	2.211°
Online Diagnosis	3.188°±2.807°	2.710°	3.026°±2.913°	2.295°

Error of acetabular index detection is measured in angle

more criticism than a doctor because the online diagnosis system is open for public use, which could amplify the influence of the mistake. Nevertheless, the widespread use of our model in the future will continue to expand the DDH database. At the same time, we will carry on refining and evaluating our method for better performance for online automatic diagnosis.

5 Conclusion

We propose a novel approach based on the multi-task hourglass network to detect hip dislocation. Online Diagnosing X-ray images of hip dislocation based on deep learning is a new attempt. We design a novel network that simultaneously predicts the locations of six landmarks in the X-ray image and predicts the age of the femoral head. The detection of landmarks is of great significance for medical diagnosis, and accurate medical diagnosis requires precise landmarks localization. The proposed network achieves an average pixel error of the six landmarks are 4.64. The prediction of the developmental age of the femoral head is equally important. Under normal circumstances, the age of femoral head development is nearly the same as the child's actual age. Experiments show that the accuracy of the proposed network is 89.3% for the femoral heads.

For the future work, there are some weaknesses in our algorithms that need to be improved.

1) The pixel error should be minor for accurate DDH diagnosis. Although the average pixel error of these six landmarks is small, the pixel error of some images is significant. The reason is that the diversity of bone morphology in these X-ray images is complex.
2) There are very few patients older than 12 years in the database. Therefore, the trained model cannot thoroughly learn the bone age information in these images. With the expansion of online data in the future, we will add images of children older than 12 years to improve the accuracy of age prediction.

Acknowledgements This work is supported by the National Nature Science Foundation of China (62022076, 61972105, 61976008, U19A2057), the China Postdoctoral Science Foundation (2021M703081), the Fundamental Research Funds for the Central Universities.

Declarations

Conflict of Interests The authors declare that there have no conflict of interests.

References

1. Cheng, L., Shi, Y., Zhang, K.: Medical treatment migration behavior prediction and recommendation based on health insurance data. World Wide Web **23**(3), 2023–2042 (2020). https://doi.org/10.1007/s11280-020-00781-3
2. Zhang, Y., Ou, W., Shi, Y., Deng, J., You, X., Wang, A.: Deep medical cross-modal attention hashing. World Wide Web, 1–18. https://doi.org/10.1007/s11280-021-00881-8 (2021)
3. Yue, L., Tian, D., Chen, W., Han, X., Yin, M.: Deep learning for heterogeneous medical data analysis. World Wide Web **23**(5), 2715–2737 (2020). https://doi.org/10.1007/s11280-019-00764-z
4. Dezateux, C., Rosendahl, K.: Developmental dysplasia of the hip. Lancet **369**(9572), 1541–1552 (2007). https://doi.org/10.1016/S0140-6736(07)60710-7
5. Tönnis, D.: Indications and time planning for operative interventions in hip dysplasia in child and adulthood. Z. Orthop. Ihre. Grenzgeb. **123**(4), 458–461 (1985)
6. Thieme, W.T., Thiersch, J.B.: Translation: Hilgenreiner on congenital hip dislocation. J. Pediatr. Orthop **6**(2), 202–214 (1986)
7. Harris, N.H., Lloyd-Roberts, G., Gallien, R.: Acetabular development in congenital dislocation of the hip: with special reference to the indications for acetabuloplasty and pelvic or femoral realignment osteotomy. J. Bone Joint Surg. Br. Vol. **57**(1), 46–52 (1975)
8. Lindstrom, J.R., Ponseti, I., Wenger, D.R.: Acetabular development after reduction in congenital dislocation of the hip. J. Bone Joint Surg. Am. Vol. **61**(1), 112–118 (1979)
9. Gaffney, B.M., Hillen, T.J., Nepple, J.J., Clohisy, J.C., Harris, M.D.: Statistical shape modeling of femur shape variability in female patients with hip dysplasia. J. Orthop. Res.®; **37**(3), 665–673 (2019). https://doi.org/10.1002/jor.24214
10. El-Sayed, M., Ahmed, T., Fathy, S., Zyton, H.: The effect of dega acetabuloplasty and salter innominate osteotomy on acetabular remodeling monitored by the acetabular index in walking ddh patients between 2 and 6 years of age: short-to middle-term follow-up. J. Child.'s Orthop. **6** (6), 471–477 (2012). https://doi.org/10.1007/s11832-012-0451-x
11. Ertürk, C., Altay, M.A., Isikan, U.E.: A radiological comparison of salter and pemberton osteotomies to improve acetabular deformations in developmental dysplasia of the hip. J. Pediatr. Orthop. B **22**(6), 527–532 (2013). https://doi.org/10.1097/BPB.0b013e32836337cd
12. Roposch, A., Ridout, D., Protopapa, E., Nicolaou, N., Gelfer, Y.: Osteonecrosis complicating developmental dysplasia of the hip compromises subsequent acetabular remodeling. Clin. Orthop. Relat. Res.®; **471**(7), 2318–2326 (2013). https://doi.org/10.1007/s11999-013-2804-2
13. Sublett, J.W., Dempsey, B.J., Weaver, A.C.: Design and implementation of a digital teleultrasound system for real-time remote diagnosis. In: Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems. https://doi.org/10.1109/CBMS.1995.465413, pp 292–298 (1995)
14. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M.K., Pei, J., Ting, M.Y.L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., Shi, A., Zhang, R., Zheng, L., Hou, R., Shi, W., Fu, X., Duan, Y., Huu, V.A.N., Wen, C., Zhang, E.D., Zhang, C.L., Li, O., Wang, X., Singer, M.A., Sun, X., Xu, J., Tafreshi, A., Lewis, M.A., Xia, H., Zhang, K.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–11319 (2018). https://doi.org/10.1016/j.cell.2018.02.010
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp 234–241. Springer, Cham (2015)
16. Xu, J., Xie, H., Liu, C., Yang, F., Zhang, S., Chen, X., Zhang, Y.: Hip landmark detection with dependency mining in ultrasound image. IEEE Trans. Med. Imaging **40**(12), 3762–3774 (2021). https://doi.org/10.1109/TMI.2021.3097355
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2016.90, pp 770–778 (2016)
18. Wang, B., Qi, G.-J., Tang, S., Zhang, L., Deng, L., Zhang, Y.; Automated pulmonary nodule detection: high sensitivity with few candidates. In: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2018), LNCS 11071, pp 759–767, Sep. 16–20, Granada, Spain (2018)
19. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. https://doi.org/10.1007/978-3-319-46484-8_29, pp 483–499. Springer, Cham (2016)

20. Al-Bashir, A.K., Al-Abed, M., Sharkh, F.M.A., Kordeya, M.N., Rousan, F.M.: Algorithm for automatic angles measurement and screening for developmental dysplasia of the hip (ddh). In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). https://doi.org/10.1109/EMBC.2015.7319854, pp 6386–6389. IEEE (2015)
21. Korman, S., Reichman, D., Tsur, G., Avidan, S.: Fast-match: Fast affine template matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2013.302, pp 2331–2338 (2013)
22. Liu, C., Xie, H., Zhang, S., Xu, J., Sun, J., Zhang, Y.: Misshapen pelvis landmark detection by spatial local correlation mining for diagnosing developmental dysplasia of the hip. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1109/TMI.2020.3008382, pp 441–449. Springer (2019)
23. Bier, B., Goldmann, F., Zaech, J.-N., Fotouhi, J., Hegeman, R.A., Grupp, R., Armand, M., Osgood, G.M., Navab, N., Maier, A.K., Unberath, M.: Learning to detect anatomical landmarks of the pelvis in x-rays from arbitrary views. Int. J. Comput. Assist. Radiol. Surg., 1–11. https://doi.org/10.1007/s11548-019-01975-5(2019)
24. Craig, J., Petterson, V.: Introduction to the practice of telemedicine. J. Telemed. Telecare **11** (1), 3–9 (2005). https://doi.org/10.1177/1357633X0501100102. PMID: 15829036
25. Hollander, J.E., Carr, B.G.: Virtually perfect? Telemedicine for COVID-19. New Engl. J. Med. **382**(18), 1679–1681 (2020). https://doi.org/10.1056/NEJMp2003539
26. Ekeland, A.G., Bowes, A., Flottorp, S.: Effectiveness of telemedicine: A systematic review of reviews. Int. J. Med. Inform. **79**(11), 736–771 (2010). https://doi.org/10.1016/j.ijmedinf.2010.08.006
27. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056
28. Long, E., Lin, H., Liu, Z., Wu, X., Wang, L., Jiang, J., An, Y., Lin, Z., Li, X., Chen, J., et al: An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat. Biomed. Eng. **1** (2), 1–8 (2017). https://doi.org/10.1038/s41551-016-0024
29. Laina, I., Rieke, N., Rupprecht, C., Vizcaíno, J.P., Eslami, A., Tombari, F., Navab, N.: Concurrent segmentation and localization for tracking of surgical instruments. In: International Conference on Medical Image Computing and Computer-assisted Intervention. https://doi.org/10.1007/s11280-019-00764-z, pp 664–672. Springer (2017)
30. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using cnns. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1007/978-3-319-46723-8_27, pp 230–238. Springer (2016)
31. Cai, X., Li, S., Liu, X., Han, G.: Vision-based fall detection with multi-task hourglass convolutional auto-encoder. IEEE Access **8**, 44493–44502 (2020). https://doi.org/10.1109/ACCESS.2020.2978249
32. Xu, Z., Huang, Q., Park, J., Chen, M., Xu, D., Yang, D., Liu, D., Zhou, S.K.: Supervised action classifier: Approaching landmark detection as image partitioning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. https://doi.org/10.1007/978-3-319-66179-7_39, pp 338–346. Springer (2017)
33. Zhang, J., Liu, M., Shen, D.: Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. IEEE Trans. Image Process. **26**(10), 4753–4764 (2017). https://doi.org/10.1109/TIP.2017.2721106
34. Kordon, F., Fischer, P., Privalov, M., Swartman, B., Schnetzke, M., Franke, J., Lasowski, R., Maier, A., Kunze, H.: Multi-task localization and segmentation for x-ray guided planning in knee surgery. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. https://doi.org/10.1007/978-3-030-32226-7_69, pp 622–630. Springer, Cham (2019)
35. Wan, J., Lai, Z., Liu, J., Zhou, J., Gao, C.: Robust face alignment by multi-order high-precision hourglass network. IEEE Trans. Image Process. **30**, 121–133 (2021). https://doi.org/10.1109/TIP.2020.3032029
36. Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q.: Centernet: Keypoint triplets for object detection. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV. https://doi.org/10.1109/ICCV.2019.00667, pp 6568–6577 (2019)
37. Law, H., Deng, J.: Cornernet: Detecting objects as paired keypoints. In: Proceedings of the European Conference on Computer Vision (ECCV). https://doi.org/10.1007/978-3-030-01264-9_45, pp 734–750 (2018)

38. Liu, A.-A., Su, Y.-T., Nie, W.-Z., Kankanhalli, M.: Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(1), 102–114 (2017)
39. Li, G., Xu, F., Li, H., Yuan, Y., An, M.: Dra-odm: a faster and more accurate deep recurrent attention dynamic model for object detection. World Wide Web. https://doi.org/10.1007/s11280-021-00971-7 (2021)
40. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2015.7298965, pp 3431–3440. IEEE Computer Society, Los Alamitos, CA, USA (2015)
41. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **42**(2), 318–327 (2020). https://doi.org/10.1109/TPAMI.2018.2858826
42. Zhou, X., Wang, D., Krähenbühl, P.: Objects as points. https://doi.org/10.48550/arXiv.1904.07850 (2019)
43. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
44. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015)
45. Xu, N., Zhang, H.-W., Liu, A.-A., Nie, W.-Z., Su, Y.-T., Nie, J., Zhang, Y.-D.: Multi-level policy and reward-based deep reinforcement learning framework for image captioning. IEEE Trans. Multimedia **22**(5), 1372–1383 (2019)
46. Deng, L., Tang, S., Fu, H., Wang, B., Zhang, Y.: Spatiotemporal breast mass detection network (MD-Net) in 4D DCE-MRI images. In: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2019), LNCS 11767, pp 271–279, Oct. 13–17, Shenzhen, China (2019)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jingyuan Xu[1] · Hongtao Xie[1] · Qingfeng Tan[2] · Hai Wu[1] · Chuanbin Liu[1] · Sicheng Zhang[3] · Zhendong Mao[1] · Yongdong Zhang[1]

Jingyuan Xu
xjy2012@mail.ustc.edu.cn

Hai Wu
wuh@mail.ustc.edu.cn

Chuanbin Liu
liucb92@ustc.edu.cn

Sicheng Zhang
398667629@qq.com

Zhendong Mao
zdmao@ustc.edu.cn

Yongdong Zhang
zhyd73@ustc.edu.cn

[1] School of Information Science and Technology, University of Science and Technology of China,
Jinzhai Road, Hefei 230026, Anhui, China

[2] Cyberspace Institution of Advanced Technology, Guangzhou University,
Guangzhou 511442 Guangdong, China

[3] Anhui Provincial Children's Hospital, Wangjiang Road, Hefei 230022, Anhui, China