bioengineering

MDPI

Article

3D Reconstruction of the Infant Hip Joint from 2D Ultrasound Sweep Sequences for DDH

Jianwei Cui *, Wenyi Zhang, Jiakuan Wang and Yuxiang Dai

Institute of Instrument Science and Engineering, Southeast University, Nanjing 210096, China; 220243765@seu.edu.cn (W.Z.); jkwangyz@126.com (J.W.); 220243737@seu.edu.cn (Y.D.)
* Correspondence: cjw@seu.edu.cn

Abstract

Infant hip ultrasound screening is crucial for the diagnosis of developmental dysplasia of the hip (DDH). However, accurately understanding the complex hip joint anatomy and identifying the Graf standard plane from two-dimensional (2D) ultrasound images alone remains challenging. Therefore, this study developed and evaluated a three-dimensional (3D) visualization method for reconstructing infant hip structures from 2D ultrasound sweep sequences. In this method, a six-degree-of-freedom robotic arm equipped with a 2D linear-array ultrasound probe was used to perform linear scanning of an infant hip ultrasound phantom, synchronously acquiring ultrasound images and the end-effector poses of the robotic arm at 0.2 mm intervals. By fully exploiting information from adjacent frames, BCAS-UNet++, which integrates atrous spatial pyramid pooling and bidirectional convolutional long short-term memory networks, was employed to segment the hip structures. Finally, continuous 3D visualization of the infant hip joint was achieved using an improved MB-Marching Cubes reconstruction algorithm. Experiments based on the phantom dataset yielded an average surface distance of 0.8717 mm, a Hausdorff distance of 2.3713 mm, a Dice coefficient of 0.8536, and an IoU of 0.7446, while the relative errors of the α angle and β angle were 1.04% and 2.83%, respectively. In the preliminary clinical evaluation involving five infants, the mean absolute errors of the α and β angles were 1.14 ± 0.42° and 2.02 ± 0.84°, respectively, with corresponding mean relative errors of 1.70% and 3.23%. Moreover, for the α angle, the 95% limits of agreement ranged from −2.83° to 2.28°, with the systematic bias being close to zero. In conclusion, the proposed method shows promising potential to assist infant DDH screening through 3D visualization and Graf-angle assessment, although further large-scale clinical validation is needed.

Keywords: infant hip ultrasound screening; image segmentation; three-dimensional visualization; continuity

Academic Editors: Laura Brattain, Vivek Singh and Md Mostafa Kamal Sarker

Received: 19 October 2026
Revised: 18 November 2026
Accepted: 23 November 2026
Published: 24 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Ultrasound examination is widely used in infant screening due to its noninvasive nature, ability to provide real-time imaging, and minimal impact on infants [1,2]. Developmental dysplasia of the hip (DDH) is one of the most common congenital joint disorders, primarily characterized by an abnormal relationship between the acetabulum and the proximal femur, encompassing a spectrum of abnormalities ranging from mild acetabular dysplasia to complete hip dislocation [3]. Ultrasound screening is commonly performed at 4–6 weeks after birth, as physiological hip laxity and transient acetabular

immaturity during the early neonatal period may lead to false-positive findings if examinations are performed too early. This timing balances screening reliability with early DDH detection, while timely treatment during the first six months of life is important for achieving favorable outcomes. If not detected and treated in a timely manner, DDH may lead to serious complications, including abnormal gait, chronic pain, and secondary osteoarthritis [4], potentially affecting patients' long-term quality of life [5]. Therefore, early diagnosis and intervention are essential. Currently, the traditional Graf method based on static 2D ultrasound images is widely used in clinical practice [6]. This method requires accurate identification of the standard plane and manual drawing of three reference lines (the baseline, bony roof line, and cartilaginous roof line) to calculate the bony roof angle (α angle) and cartilaginous roof angle (β angle) [7]. DDH is then diagnosed based on these two angles in combination with the infant's age [8]. However, accurately identifying the Graf standard plane remains challenging due to factors such as infant movement, variations in body size, and high operator dependency [9,10]. During clinical screening, physicians typically need to continuously adjust the position of the 2D ultrasound probe while mentally segmenting, interpreting, and reconstructing the 3D anatomical structures of the infant hip joint [11]. This process requires considerable professional skills and clinical experience [12].

In recent years, many researchers have begun to apply deep learning methods to infant hip ultrasound images for standard plane detection, anatomical structure segmentation, and measurement of relevant parameters. El-Hariri et al. used a U-Net network with multi-channel inputs to segment the ilium. Their results demonstrated that, compared with manual segmentation, the deep learning-based method improved segmentation accuracy and speed while reducing outliers and failure rates [13]. Zhang et al. introduced a region of interest (ROI) into a fully convolutional network (FCN) to improve acetabular segmentation [14]. Jaremko et al. proposed integrating an AI-assisted workflow into clinical practice and reported a reduction in the number of follow-up scans [15]. In our previous study, we proposed DDH-Net, a fully automated computer-aided method for infant hip ultrasound assessment. By combining YOLOv8-based pose estimation with U-Net image segmentation, DDH-Net enabled automatic identification of the Graf standard plane and measurement of the α and β angles, thereby reducing the dependence of conventional ultrasound examinations on operator experience [16]. However, most existing methods primarily focus on analyzing individual 2D ultrasound images near the standard plane, limiting their ability to interpret complex hip joint structures and spatial anatomical relationships.

Previous studies have demonstrated that 3D ultrasound can provide a more comprehensive representation of the spatial anatomy of the hip joint and reduce the risk of misdiagnosis caused by inaccurate probe positioning [17]. However, conventional 2D linear-array ultrasound probes are still predominantly used in clinical DDH screening, while the clinical adoption of 3D volumetric probes and high-frequency 3D linear-array probes remains limited. Therefore, 3D reconstruction techniques based on 2D ultrasound image sequences have the potential to overcome these limitations and provide a more convenient and intuitive approach to assisting DDH screening.

3D reconstruction of hip joint medical images has been extensively investigated for disease diagnosis, preoperative planning, and personalized treatment. Zheng et al. proposed a reconstruction method based on a statistical shape model (SSM), which enables the reconstruction of a patient-specific 3D pelvic model from a single anteroposterior pelvic radiograph, allowing accurate assessment of acetabular cup orientation after total hip arthroplasty (THA) [18]. Schumann et al. developed a 3D hip reconstruction method based on biplanar X-ray images. By combining landmark-based calibration with hierarchical fitting of statistical shape models, they achieved surface reconstruction of the

acetabulum and proximal femur. The mean surface reconstruction errors for the acetabulum and proximal femur were only 1.06 mm and 1.01 mm, respectively [19]. Wan et al. proposed a fully automated method for 3D reconstruction of the outer surface of the femoral head from computed tomography (CT) images. By generating valley-enhanced images to improve bone boundary contrast and combining Otsu multi-threshold segmentation, 3D morphological operations, and Bayesian decision rules to optimize segmentation results, smooth 3D surface reconstruction of the femoral head and acetabulum was ultimately achieved using the Marching Cubes algorithm, providing reliable technical support for computer-assisted hip surgery planning [20]. Zhang et al. proposed a method for 3D surface reconstruction of the acetabulum from multi-view preoperative X-ray images. By incorporating square-root velocity function (SRVF)-based elastic registration, the method iteratively deformed a standard hemispherical model into a patient-specific acetabular model, providing accurate anatomical information for hip replacement surgery [21]. Pham et al. developed an electromagnetic tracking-based freehand 3D ultrasound reconstruction method. An electromagnetic tracking system was used to record the spatial position and orientation of each 2D ultrasound frame, and volume rendering was subsequently performed to reconstruct the hip joint in three dimensions, enabling quantitative assessment of hip displacement [22]. Hoang et al. proposed a noninvasive method for assessing hip displacement based on 3D reconstruction from conventional 2D ultrasound sequences. The Visualization Toolkit (VTK) was used for 3D reconstruction, while linear interpolation was employed to improve voxel resolution. The maximum relative error in migration percentage (MP) measurement using this ultrasound-based 3D reconstruction method was only 1.16% [23]. Although 3D reconstruction techniques based on high-quality medical imaging modalities such as CT have become relatively mature, 3D reconstruction based on ultrasound remains more challenging. In particular, conventional 2D ultrasound images are affected by speckle noise, blurred tissue boundaries, and nonuniform grayscale distributions, making accurate anatomical segmentation difficult. Furthermore, current methods have several limitations. First, insufficient use of temporal information from consecutive 2D ultrasound frames limits the consistency of anatomical segmentation. Second, manual probe scanning is highly operator-dependent, making standardized image acquisition and accurate spatial positioning difficult. Third, existing 3D ultrasound studies for DDH have mainly focused on hip displacement assessment, with limited integration of detailed 3D anatomical reconstruction and Graf parameter measurement. To address these limitations, an integrated framework combining robotic scanning, spatio-temporal segmentation, 3D reconstruction, and clinical parameter measurement is needed.

This study establishes a comprehensive technical framework integrating controlled-pose data acquisition, sequential ultrasound image segmentation, 3D reconstruction, and Graf parameter measurement, as shown in Figure 1. The main contributions are as follows:

(1) A robot-assisted ultrasound data acquisition method based on six-degree-of-freedom linear scanning is proposed. This method employs a robotic arm equipped with a conventional 2D ultrasound probe to perform controlled linear scanning, while synchronously recording ultrasound images and the corresponding end-effector poses at intervals of 0.2 mm. The proposed method reduces dependence on manual probe manipulation, improves the standardization of ultrasound data acquisition, and provides spatially ordered ultrasound image sequences for subsequent 3D reconstruction of the infant hip joint.

(2) BCAS-UNet++ segmentation network based on the fusion of spatio-temporal features: We propose the BCAS-UNet++ network, which integrates atrous spatial pyramid pooling (ASPP) with a bidirectional convolutional long short-term memory (Bi-ConvLSTM) network. Building upon multi-scale spatial feature extraction, the network further utilizes contextual information from five consecutive ultrasound

frames, comprising the current frame, two preceding frames, and two subsequent frames, to enhance segmentation accuracy and inter-frame consistency.

(3) MB-Marching Cubes 3D reconstruction algorithm based on cubic triangular Bézier surface interpolation: This algorithm optimises the traditional Marching Cubes algorithm using cubic triangular Bézier surface interpolation to more accurately reconstruct structural features in regions with significant variations in curvature; it also incorporates connectivity optimisation to eliminate interference from isolated fragments, thereby enhancing the surface smoothness and structural fidelity of the 3D reconstruction results.

(4) Real-world clinical ultrasound data of infant hips were collected, and the clinical diagnostic results were compared with the measurements obtained using the 3D reconstruction method. This validated the feasibility of the method described in this paper.

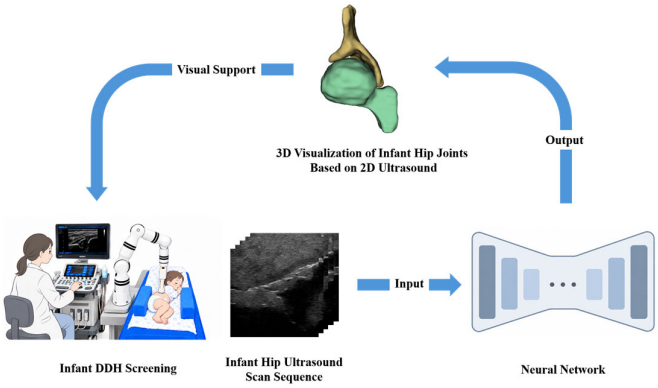


Figure 1. Workflow of 3D visualization-assisted DDH screening based on 2D ultrasound image sequences. A conventional 2D ultrasound probe used in routine clinical examinations was mounted on the end-effector of a robotic arm to acquire ultrasound image sequences. The 3D visualization images of the infant hip joint were reconstructed using the proposed method. The blue arrows indicate the workflow and data flow.

The remainder of this paper is organised as follows: Section 2 provides a detailed description of the system's overall architecture and technical principles; Section 3 presents the experimental validation and performance evaluation of the proposed approach; Section 4 contains the discussion; and Section 5 presents the conclusions.

2. Materials and Methods

2.1. Ultrasonic Data Acquisition System

This paper describes the design and construction of an ultrasound image acquisition system based on a six-degree-of-freedom robotic arm; the overall hardware architecture is shown in Figure 2a. It primarily comprises a six-degree-of-freedom robotic arm (RM65-B, RealMan Intelligent Technology (Beijing) Co., Ltd., Beijing, China, with a repeatability of ±0.05 mm), a 2D linear array ultrasound probe (model D8c, Guangdong Qiyou Medical Co., Ltd., Foshan, Guangdong, China), with a frequency range of 6–11 MHz and an image resolution of 1920 × 1200), and a main controller (a computer equipped with an Intel® Core™ i7-10750H CPU). The robotic arm and the main controller exchange data at high speed via the TCP/IP protocol to ensure real-time control.

The infant hip ultrasound phantom (as shown in Figure 2b) was created by extracting and reconstructing a digital 3D model comprising multiple structures—including the femoral head, acetabulum and ilium—from high-resolution CT/MRI DICOM data using Mimics 25.0 software (Materialise NV, Leuven, Belgium), followed by surface smoothing with

Geomagic Wrap 2023 (3D Systems, Inc., Rock Hill, SC, USA); hydrogels with varying compositions, tailored to different tissues, were then employed to simulate the acoustic properties of the actual hip joint structure; finally, a removable multi-chamber mould was fabricated using SLA photopolymerisation 3D printing, and the phantom was completed through a combination of gradient photothermal curing, vacuum degassing and calcium ion gradient diffusion processes.

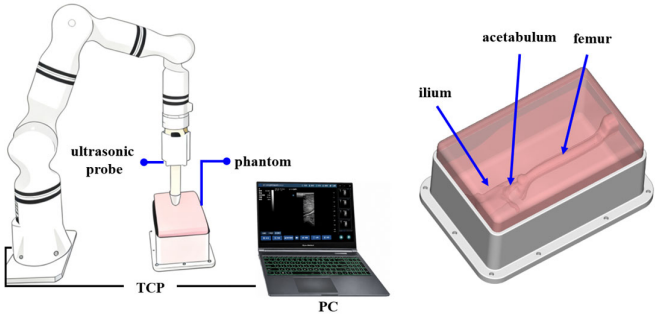


Figure 2. Ultrasonic data acquisition system. (**a**) System hardware architecture; (**b**) ultrasound phantom.

The 0.2 mm scanning interval was determined through analysis of multiple ultrasound sweep sequences, taking into account the five-frame input configuration of BCAS-UNet++. The analysis showed that larger intervals reduced anatomical continuity between adjacent frames, whereas smaller intervals increased information redundancy and computational cost. Accordingly, the ultrasound probe was configured to acquire one frame every 0.2 mm, with the robotic arm's end-effector pose recorded simultaneously.

2.2. Ultrasound Image Segmentation Network

Ultrasound images of infants' hips are generally characterised by significant noise, blurred tissue edges and uneven grey-scale distribution, as shown in Figure 3. This paper aims to achieve 3D reconstruction of infants' hips based on segmentation results, which imposes stricter requirements on segmentation accuracy and continuity. Therefore, this paper proposes the BCAS-UNet++ segmentation network, which is based on the fusion of spatio-temporal features.

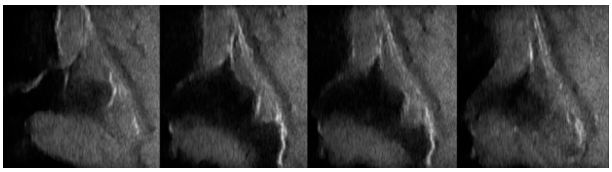


Figure 3. Sequence of 2D ultrasound images.

2.2.1. BCAS-UNet++ Network Framework

The BCAS-UNet++ network is based on an encoder–decoder architecture, retaining the dense skip connections and deep supervised paths of the original UNet++. At the same time, taking into account the characteristics of paediatric hip ultrasound images, the following two improvements have been made:

(1) Multi-scale spatial feature enhancement: An ASPP module is introduced at the end of the encoder. Through parallel dilated convolutions with varying expansion ratios, multi-scale features are extracted without compromising resolution, enabling the

network to simultaneously capture both local details and global structure, thereby accurately segmenting hip joint anatomical structures of varying scales.

(2) Fusion of temporal contextual information: A BiConvLSTM module is incorporated following the spatial encoder. This module facilitates cross-frame temporal information fusion by integrating contextual information from adjacent frames within the five-frame input sequence. This compensates for the instability of single-frame image segmentation under noise and blurring artefacts, thereby enhancing inter-frame continuity and consistency.

The overall architecture is shown in Figure 4. The input sequence of ultrasound images is first downsampled layer by layer by the encoder to extract deep semantic features; subsequently, the feature maps pass through the ASPP module to obtain multi-scale information, and resolution is gradually restored in the decoder. The BiConvLSTM then fuses features from consecutive frames to produce a multi-structure segmentation result for the hip joint. Through this improvement, the BCAS-UNet++ network is able to simultaneously address four key challenges—blurred boundaries, noise interference, significant differences in the scales of various structures, and inter-frame inconsistencies—and produce continuous and clean hip joint segmentation results.

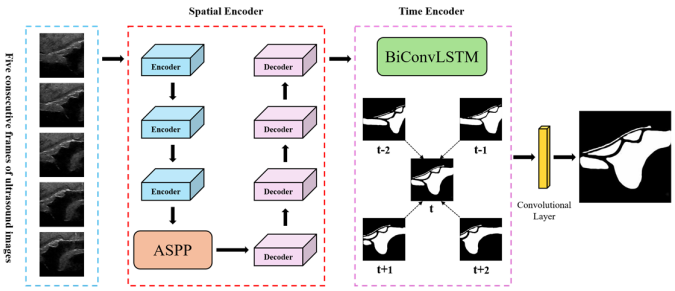


Figure 4. Diagram of the BCAS-UNet++ model architecture.

2.2.2. Spatial Encoder

This paper constructs a spatial encoder based on the UNet++ architecture, into which an ASPP module is embedded, as shown in Figure 5. When hip joint ultrasound images are fed into the encoder on the left-hand side of UNet++, the network extracts the high-level semantic information contained in the original images, which is then fed into the ASPP module to extract multi-scale features from the images. Subsequently, this high-level semantic information—which contains multi-scale features—is fed into the UNet++ decoder, where it is merged with the corresponding low-level semantic information from the left-hand encoder, ultimately producing a feature map of the same size as the original image. Furthermore, the final convolutional layer in UNet++—which uses a 1 × 1 convolutional kernel to predict segmentation results—has been removed, thereby allowing the 64-channel feature map to be fed directly into the BiConvLSTM module.

In particular, ASPP consists of a 1 × 1 convolutional layer, three 3 × 3 dilated convolutional layers with different dilation rates (6, 12 and 18, respectively) and image pooling; these branches operate in parallel, and their outputs are concatenated before being fed into a 1 × 1 convolutional layer. The convolutional layers with different expansion ratios in the ASPP can generate feature maps at different scales, thereby enabling the network to accurately segment key hip joint structures at various scales. When the expansion ratio is small, all nine weights contained in the 3 × 3 convolutional kernel act on the effective region of the feature map. However, when the expansion ratio is large, the 3 × 3 convolution kernel degenerates into a 1 × 1 convolution kernel, with only the weight at the centre of the kernel being effective. In this scenario, the convolution layer is unable to capture

semantic information from the entire image; therefore, image pooling is introduced, comprising an average pooling layer, a 1 × 1 convolution layer and an upsampling layer, thereby enabling the effective capture of semantic information from the entire image. Compared with the traditional UNet++, the incorporation of the ASPP module enhances the segmentation capability for multi-scale structures, aiding in the precise identification of anatomical structures—such as the acetabular rim and labrum—which differ significantly in scale from the femoral head.

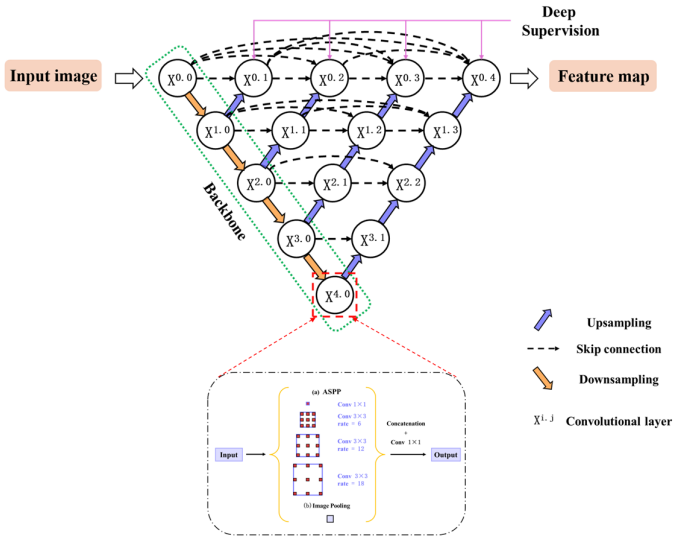


Figure 5. Schematic diagram of the spatial encoder. (**a**) Atrous Spatial Pyramid Pooling (ASPP); (**b**) Image Pooling.

2.2.3. Time Encoder

This paper introduces the BiConvLSTM module as a temporal encoder. The feature maps output by the spatial encoder are fed into the temporal encoder, where they are fused with forward and backward feature information to generate the final segmentation result. A BiConvLSTM unit consists of two ConvLSTM branches, each maintaining its own hidden state and cell state.

ConvLSTM can be used to capture unidirectional long- and short-term memories between ultrasound image sequences; by fusing spatial features, it outputs predictions containing spatio-temporal feature information. Compared to LSTM networks, ConvLSTM uses convolutional layers in place of fully connected layers, thereby preserving the spatial information in the image sequences and enabling better segmentation results by fully utilising the spatio-temporal information of the image sequences. The relevant equations for the ConvLSTM unit are shown below:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1} + b_f\right) \quad (2)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + b_o) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (4)$$

$$H_t = o_t \odot \tanh(C_t) \quad (5)$$

Here, '$*$' denotes a convolution operation, '$\odot$' denotes the Hadamard product, and 'σ' denotes the sigmoid function; W_{x*} and W_{h*} correspond to 2D convolution kernels

associated with the input states and hidden states, respectively. At each time-series step, the input $(X_1, X_2, \ldots X_t)$ is used to update the hidden states $(H_0, H_1, \ldots H_{t-1})$ and cell states $(C_1, C_2, \ldots C_t)$ via the activation gates i_t, f_t, o_t. b_i, b_f, b_o, b_c denote the bias terms.

Compared to ConvLSTM, each BiConvLSTM unit possesses two sets of hidden states and cell states: one set is used to extract forward time-series information, whilst the other is used to extract backward time-series information. This enables BiConvLSTM to integrate bidirectional contextual information within the five-frame input sequence, combining features from the two preceding and two subsequent frames to optimize the segmentation results for the current frame. Figure 6 illustrates the overall architecture of the BiConvLSTM, comprising h_f and c_f for capturing forward time-series information, h_b and c_b for capturing backward time-series information.

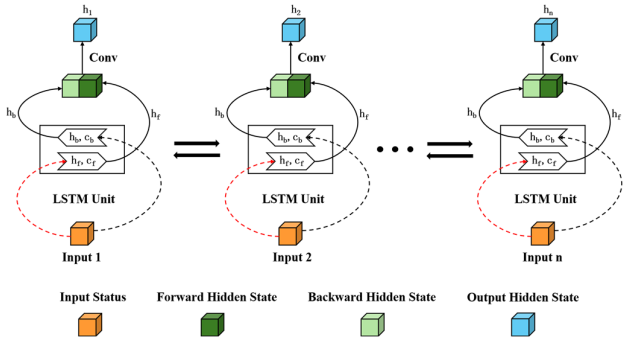


Figure 6. BiConvLSTM architecture diagram.

For each target frame I_t, a five-frame sliding window is constructed, consisting of two preceding frames, the current target frame, and two subsequent frames. The window advances by one frame along the ultrasound sequence, so that each frame is successively treated as the central target frame. Let X_{t+k} denote the feature map extracted from frame I_{t+k} by the spatial encoder. Within each window, the forward branch processes the feature maps in chronological order, whereas the backward branch processes them in reverse chronological order. Consequently, at the central time step t, the forward hidden state incorporates information from the preceding frames and the current frame, while the backward hidden state incorporates information from the subsequent frames and the current frame. Their state updates are expressed as follows:

$$(\vec{H}_t, \vec{C}_t) = ConvLSTM_f(X_t, \vec{H}_{t-1}, \vec{C}_{t-1}) \tag{6}$$

$$(\overleftarrow{H}_t, \overleftarrow{C}_t) = ConvLSTM_b(X_t, \overleftarrow{H}_{t+1}, \overleftarrow{C}_{t+1}) \tag{7}$$

At the central time step t of each five-frame window, the forward and backward hidden-state feature maps are concatenated along the channel dimension and fused through a 1 × 1 convolution, which can be mathematically expressed as

$$F_t^{fuse} = Conv_{1\times1}(Concat_c[\vec{H}_t, \overleftarrow{H}_t]) \tag{8}$$

$$\hat{Y}_t = SegHead(F_t^{fuse}) \tag{9}$$

Here, $\hat{Y}_t$ represents the predicted segmentation mask of the central target frame within the current five-frame window. By shifting the window one frame at a time and repeating the same process, segmentation results are generated for all frames in the ultrasound sequence.

2.3. 3D Reconstruction Algorithms

The traditional Marching Cubes algorithm is widely used in high-quality medical imaging data such as CT scans; however, when applied directly to ultrasound images with a low signal-to-noise ratio, it gives rise to issues such as jagged surfaces, local discontinuities and loss of detail, making it difficult to meet the geometric accuracy requirements for clinical screening of DDH. To address this, this paper proposes an MB-Marching Cubes 3D reconstruction algorithm based on cubic triangular Bézier surface interpolation.

2.3.1. Isosurface Extraction

The segmentation masks generated by BCAS-UNet++ were sequentially stacked according to the acquisition order to construct a three-dimensional binary volume. In this volume, the foreground anatomical structures were assigned a voxel value of 1, while the background was assigned 0. The voxel spacing was defined as 0.05 mm × 0.05 mm × 0.2 mm, corresponding to an in-plane pixel spacing of 0.05 mm in both directions and an inter-slice interval of 0.2 mm determined by the robotic scanning protocol described in Section 2.1. To extract the boundaries between the foreground anatomical structures and the background, an isovalue of C = 0.5 was selected, as it represents the midpoint between the binary voxel values of 0 and 1. This threshold provides a consistent criterion for identifying the isosurface throughout the reconstructed volume. Based on these parameters, the Marching Cubes algorithm was applied to extract the initial triangular surface mesh.

Specifically, the three-dimensional volume was divided into a series of grid cells, each consisting of eight neighboring voxel samples located at its vertices (v_1–v_8 in Figure 7a). The algorithm traversed the volume with a step size of one grid cell along each axis. For each cell, the scalar values at the eight vertices were compared with the predefined isovalue C to determine whether the isosurface intersected the cell. Each vertex was classified as either inside or outside the isosurface according to its scalar value relative to C, resulting in $2^8 = 256$ possible vertex configurations. Based on rotational and complementary symmetries, these configurations can be grouped into 15 basic cases [24,25], as illustrated in Figure 7b. Cells whose vertices were all classified on the same side of the isovalue did not contain an isosurface intersection and were therefore excluded from surface generation. For the remaining cells, the intersection points between the isosurface and the corresponding cell edges were calculated using linear interpolation of the scalar values at the edge endpoints. The resulting intersection points were connected according to the corresponding Marching Cubes configurations to generate local triangular facets. Finally, the triangular facets extracted from adjacent grid cells were assembled to form a complete initial surface mesh, providing the geometric basis for the subsequent cubic triangular Bézier surface interpolation and connectivity optimization.

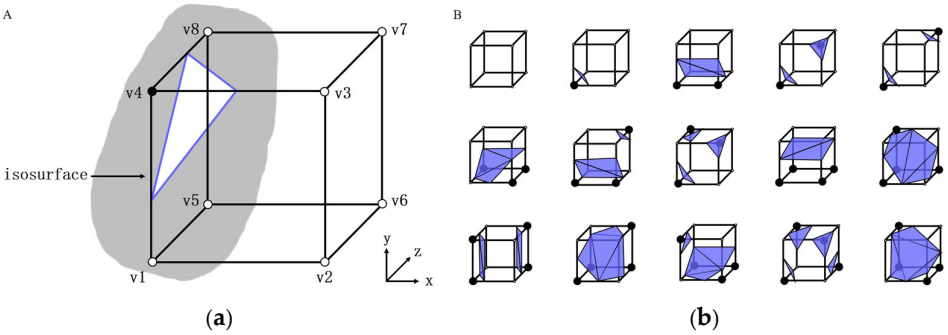


Figure 7. Schematic diagram of the Marching Cubes algorithm. (**a**) Voxel-based isosurface extraction; (**b**) 15 basic configurations.

2.3.2. Triangular Bézier Surface Interpolation

After extracting the isosurfaces using the Marching Cubes algorithm, the reconstructed model surface may exhibit jagged edges and planar facets. To improve geometric continuity and surface reconstruction accuracy, this study employs cubic triangular Bézier surface interpolation to smooth the mesh. Unlike the planar triangular facets generated by conventional Marching Cubes, Bézier interpolation incorporates local vertex positions and normal vectors to construct curved surface patches, thereby reducing surface discontinuities and better preserving anatomical curvature, particularly in regions such as the femoral head and acetabular roof. Specifically, given the positions of the triangular vertices and their corresponding normal vectors, ten Bézier control points are determined using a nine-parameter interpolation scheme to generate smooth surface patches. A cubic triangular Bézier surface patch can be expressed as

$$\begin{aligned}\mathbf{P}(u,v\,w) = &\sum_{i,jk\geq}^{i+j+k=3} \frac{3!}{i!\,j!\,k!} u^i v^j w^k \mathbf{P}_{ijk} \\ &= u^3\mathbf{P}_{300} + v^3\mathbf{P}_{030} + w^3\mathbf{P}_{003} \\ &+3u^2v\mathbf{P}_{210} + 3u^2w\mathbf{P}_{201} + 3uv^2\mathbf{P}_{120} \\ &+3v^2w\mathbf{P}_{021} + 3vw^2\mathbf{P}_{012} + 3uw^2\mathbf{P}_{102} + 6uvw\mathbf{P}_{111}\end{aligned} \tag{10}$$

Here, u, v and w denote the barycentric coordinates of a point within the triangular domain (0 ≤ u, v, w ≤ 1 and u + v+ w = 1), and $\mathbf{P}_{ijk}$ are the control points of the Bézier surface. A Bézier surface patch consists of ten control points, which form a triangular mesh known as a 'control grid'.

Various methods have been proposed for interpolating triangular meshes using cubic triangular Bézier surface patches [26–28]. In this study, the nine-parameter interpolation method—the simplest of the proposed methods—was adopted. In this approach, given the three vertices of a triangle and their corresponding normal vectors, ten control points can be determined (as shown in Figure 8a); the specific method is described below.

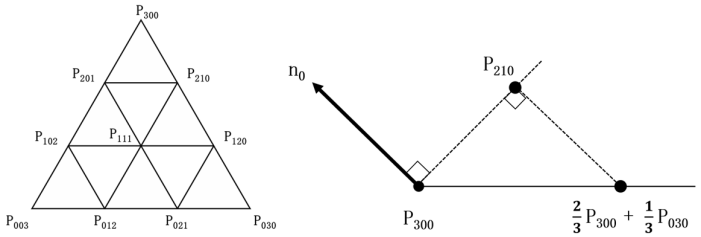


Figure 8. Nine-parameter interpolation method. (**a**) Distribution of the control points on the triangular patch; (**b**) Geometric construction of the control point P_{210} and the corresponding normal direction n_0.

Let, $\mathbf{v}_0$, $\mathbf{v}_1$ and $\mathbf{v}_2$ be the given position vectors of the vertices of a triangle in the mesh. In Figure 8b, $\mathbf{P}_{300}$, $\mathbf{P}_{030}$ and $\mathbf{P}_{003}$ correspond to the three vertices of this triangle, respectively. Therefore,

$$\mathbf{P}_{300} = \mathbf{v}_0, \mathbf{P}_{030} = \mathbf{v}_1, \mathbf{P}_{003} = \mathbf{v}_2. \tag{11}$$

Let $\mathbf{n}_0$, $\mathbf{n}_1$ and $\mathbf{n}_2$ be the unit normal vectors at each vertex of the triangle. The normal vector at each vertex is calculated by taking the average of the normal vectors of all triangles adjacent to that vertex. The internal control points on the triangle's boundary are determined as follows. Taking the boundary endpoints $\mathbf{P}_{300}$ and $\mathbf{P}_{030}$ and their

corresponding unit normal vectors $\mathbf{n}_0$ and $\mathbf{n}_1$ as examples, the control points $\mathbf{P}_{210}$ and $\mathbf{P}_{120}$ can be calculated as follows:

$$\mathbf{P}_{210} = \frac{1}{3}\left(2\mathbf{P}_{300} + \mathbf{P}_{030} - \left(\mathbf{n}_0 \cdot (\mathbf{P}_{030} - \mathbf{P}_{300})\right)\mathbf{n}_0\right), \tag{12}$$

$$\mathbf{P}_{120} = \frac{1}{3}\left(2\mathbf{P}_{030} + \mathbf{P}_{300} - \left(\mathbf{n}_1 \cdot (\mathbf{P}_{030} - \mathbf{P}_{300})\right)\mathbf{n}_1\right). \tag{13}$$

The other points, $\mathbf{P}_{021}$, $\mathbf{P}_{012}$, $\mathbf{P}_{102}$ and $\mathbf{P}_{201}$, are determined in the same way. Finally, point $\mathbf{P}_{111}$ is determined as follows:

$$\begin{aligned}\mathbf{P}_{111} = \frac{1}{4}(\mathbf{P}_{210} + \mathbf{P}_{120} + \mathbf{P}_{102} + \mathbf{P}_{201} + \mathbf{P}_{021} + \mathbf{P}_{012}) \\ -\frac{1}{6}(\mathbf{P}_{300} + \mathbf{P}_{030} + \mathbf{P}_{003}).\end{aligned} \tag{14}$$

2.3.3. Connectivity Optimisation

The reconstructed surface mesh may contain degenerate triangular facets and small disconnected fragments arising from segmentation noise and boundary artifacts. To improve mesh quality and topological consistency, a two-stage connectivity optimization procedure was adopted. First, geometrically degenerate triangles were identified and removed. Subsequently, connected-component analysis was performed to eliminate isolated mesh fragments.

Degenerate triangles are patches where three vertices are nearly collinear or where the area is extremely small. Therefore, the cross-product of the two edge vectors is calculated for each triangle to obtain its area in square units, which is then compared with a very small threshold (1×10^{-12}). Any triangle with an area below this threshold is identified as a degenerate triangle and is immediately removed. This procedure eliminates geometrically invalid elements, improves mesh quality, and reduces potential numerical instability during subsequent surface processing and quantitative evaluation.

At the same time, due to voxel noise or boundary artefacts, the Marching Cubes algorithm may sometimes generate multiple small, disjoint fragments of mesh. To address this issue, the adjacency relationships between triangles (shared edges) are organised into a graph structure, and depth-first search (DFS) is used to traverse and count all connected components, identifying the one containing the largest number of triangles—that is, the main mesh. Subsequently, only this largest connected component is retained, whilst all other small fragments are discarded. By eliminating isolated surface fragments, this procedure reduces spurious geometric elements that may otherwise affect the reconstructed surface and contribute to large local deviations during surface-distance evaluation.

3. Results

3.1. Experimental Setup

The deep learning experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU. The software environment consisted of Windows 10, Python 3.8, and PyTorch 2.3.1. The BCAS-UNet++ model was trained using the Adam optimizer, with an initial learning rate of 1×10^{-4} and a cosine annealing learning rate schedule. The batch size was set to 8, and training was configured for a maximum of 300 epochs. An early stopping strategy was employed by monitoring the validation loss after each epoch, with a patience of 40 epochs to prevent overfitting.

3.2. Results of Ultrasound Image Segmentation

3.2.1. Data Pre-Processing

To validate the performance of the BCAS-UNet++ model in the segmentation of infants' and young children's hips, a total of 900 ultrasound images were acquired from an ultrasound phantom of infants' and young children's hips through 12 independent scanning passes. The acquired images were stored sequentially in their original acquisition order within each scanning pass. The hip joint structures in each ultrasound image were manually annotated using annotation software by a clinical expert with more than 20 years of clinical experience. To ensure the reliability of the annotations, all segmentation masks were independently reviewed by two additional experts with extensive clinical experience. When discrepancies were identified, the corresponding anatomical boundaries were reassessed and discussed among the three experts until a consensus was reached. Representative annotation results are shown in Figure 9.

To ensure the reliability of model evaluation and avoid potential data leakage between adjacent frames, this study adopted a random partitioning strategy at the complete scanning-sequence level. Specifically, among the 12 ultrasound scanning sequences, 8 sequences were randomly selected for the training set, 2 for the validation set, and the remaining 2 for the test set. Since the number of images varied across scanning sequences, the resulting ratio of image counts in the training, validation, and test sets was approximately 20:5:6. The training, validation, and test sets contained 581, 145, and 174 ultrasound images, respectively, corresponding to 8, 2, and 2 independent scanning sequences. All images from the same scanning sequence were assigned exclusively to one dataset, ensuring that no frames from the same scanning sequence were shared across different datasets. To incorporate contextual information from adjacent ultrasound images and improve inter-frame segmentation consistency, the temporal sequence length of BCAS-UNet++ was set to five frames. This choice was based on the anatomical characteristics of the infant hip joint and preliminary observations of ultrasound sweep sequences. Due to the small size of the infant hip joint, anatomical structures change noticeably as scanning progresses. Extending the temporal window beyond five frames covers more pronounced anatomical variations, introduces additional information that may be less relevant to the target frame, and increases computational cost. Therefore, five consecutive frames were selected to balance anatomical continuity, contextual information, and computational efficiency. Specifically, each input sequence consisted of the current frame, two preceding frames, and two subsequent frames. These frames were jointly processed by the BiConvLSTM module to capture bidirectional temporal dependencies and enhance inter-frame segmentation consistency. For frames near the beginning or end of a scanning sequence, boundary frames were replicated when necessary to maintain a fixed sequence length of five.

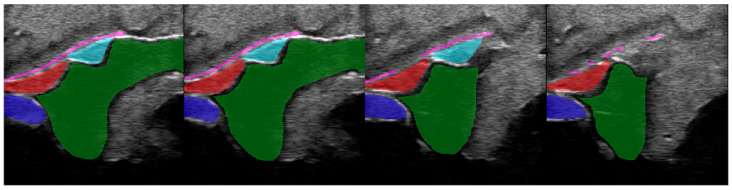


Figure 9. Results of ultrasound image labelling. Pink indicates the ilium, dark blue indicates the lower limb of the ilium, red marks the upper ilium keypoint, light blue indicates the synovial fold, and green indicates the femoral head.

3.2.2. Loss Function

This paper employs BCE Loss (Binary Cross-Entropy Loss) and Dice Loss to train the BCAS-UNet++ model. The formula for BCE Loss is as follows:

$$L_{BCE} = -\sum_{i=1}^{N} \left[y_i \ln\left(\sigma(x_i)\right) + (1 - y_i) \ln\left(1 - \sigma(x_i)\right) \right] \tag{15}$$

Here, y_i is the label value of the actual sample, x_i is the model's predicted output, and σ is the sigmoid function, which maps the model's predicted output to the range 0–1. Furthermore, the formula for the Dice Loss is as follows:

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{16}$$

In this context, $|X \cap Y|$ denotes the intersection between regions X and Y, whilst $|X|$ and $|Y|$ denote the number of elements in regions X and Y respectively. Multiplying the numerator by 2 ensures that the final result of the fraction falls within the range 0–1. Consequently, the loss function ultimately adopted in this paper is the sum of the BCE Loss and the Dice Loss, as shown below:

$$L = L_{BCE} + L_{Dice} \tag{17}$$

3.2.3. Evaluation Criteria

This paper employs IoU and the Dice coefficient to evaluate the model's performance in segmenting hip ultrasound images. IoU is a metric widely used to assess segmentation performance; it calculates the ratio of the intersection to the union of the ground truth labels and the model's predictions, and is calculated as follows:

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{18}$$

Here, $A \cap B$ denotes the intersection of the predicted region and the ground-truth region, whilst $A \cup B$ denotes the union of the predicted region and the ground-truth region. The formula for calculating the Dice coefficient is as follows:

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{19}$$

Here, $|X \cap Y|$ denotes the intersection of the predicted region and the true region, whilst $|X|$ and $|Y|$ denote the number of elements in X and Y, respectively.

3.2.4. Qualitative Analysis

As shown in Figure 10, the segmentation results of the three models exhibit clear differences in anatomical boundary delineation and structural integrity. Compared with UNet and UNet++, BCAS-UNet++ more effectively preserves the contours of key anatomical structures, particularly in regions with blurred boundaries and low contrast, such as the acetabular rim and femoral head. Furthermore, the proposed model produces fewer visible artefacts and more continuous segmentation boundaries. These improvements are particularly relevant to the subsequent 3D reconstruction process, which requires accurate and spatially consistent segmentation results from consecutive ultrasound frames.

3.2.5. Quantitative Assessment

To quantitatively assess the segmentation performance of the proposed BCAS-UNet++, comparative experiments were conducted against UNet and UNet++. UNet++ was selected as the primary baseline because it constitutes the backbone of the proposed network, while UNet was included as a conventional reference model. The Dice coefficient and Intersection over Union (IoU) were used to evaluate segmentation accuracy. The comparative results are presented in Table 1.

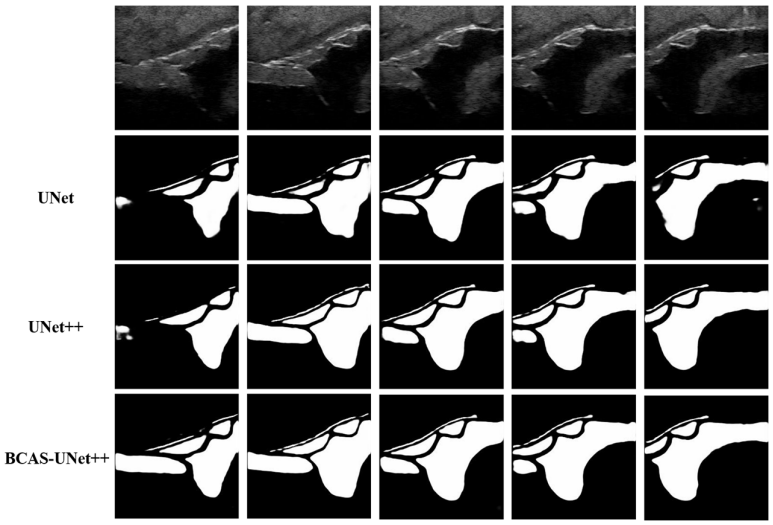


Figure 10. Comparison of segmentation results.

Table 1. Comparison of model segmentation performance.

Model	Dice	IoU
UNet	0.6205	0.5127
UNet++	0.7081	0.6681
BCAS-UNet++	0.8430	0.7852

As shown in Table 1, BCAS-UNet++ achieved the highest segmentation performance among the three evaluated models, with a Dice coefficient of 0.8430 and an IoU of 0.7852. Compared with the primary baseline UNet++, the Dice coefficient increased from 0.7081 to 0.8430, while the IoU increased from 0.6681 to 0.7852, corresponding to relative improvements of 19.05% and 17.53%, respectively. The proposed model also outperformed the conventional UNet on both metrics. These findings demonstrate the effectiveness of the proposed spatio-temporal feature fusion architecture in improving hip ultrasound image segmentation accuracy, as evidenced by the descriptive comparison of Dice and IoU values.

To further evaluate the segmentation performance of BCAS-UNet++ across different anatomical structures, class-wise Dice and IoU scores were analyzed. As shown in Table 2, BCAS-UNet++ achieved Dice scores ranging from 0.7910 to 0.9100 across five anatomical structures. The femoral head (0.9100) and lower limb of the ilium (0.8892) exhibited the highest segmentation accuracy, while the ilium (0.7910) and synovial fold (0.8002) showed relatively lower scores, reflecting the challenges associated with complex anatomical morphology and indistinct ultrasound boundaries. Overall, the class-wise results demonstrate the good segmentation performance of BCAS-UNet++ across different anatomical structures, supporting its effectiveness in infant hip ultrasound image segmentation.

Table 2. Class-wise Dice and IoU scores of BCAS-UNet++.

Anatomical Structure	Dice	IoU
Ilium	0.7910	0.7018
Lower limb of the ilium	0.8892	0.8373
Upper ilium keypoint	0.8246	0.7646
Synovial fold	0.8002	0.7285
Femoral head	0.9100	0.8939

The training and validation loss curves of the proposed BCAS-UNet++ model are presented in Figure 11. Both losses decreased rapidly during the initial training epochs and gradually stabilized as training progressed. The validation loss reached its minimum value of 0.4281 at epoch 167. Subsequently, the validation loss fluctuated without further improvement for 40 consecutive epochs, triggering early stopping at epoch 207. These results indicate that the model achieved a relatively stable validation performance, while the early stopping strategy prevented unnecessary further training.

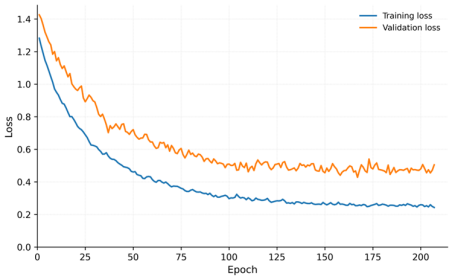


Figure 11. Training and validation loss curves of the proposed BCAS-UNet++ model.

3.2.6. Ablation Study

To investigate the contributions of ASPP and BiConvLSTM to segmentation performance, an ablation study was conducted using UNet++ as the baseline. Four model configurations were evaluated: the original UNet++, UNet++ incorporating ASPP, UNet++ incorporating BiConvLSTM, and the complete BCAS-UNet++. Each module was evaluated individually and in combination to assess its contribution to segmentation accuracy. All configurations were trained and evaluated using the same dataset partition, training settings, and evaluation metrics. The results are summarized in Table 3.

Table 3. Ablation study of the ASPP and BiConvLSTM modules (√ indicates that the corresponding module is included, while - indicates that it is not included).

Model	ASPP	BiConvLSTM	Dice	IoU
UNet++	-	-	0.7081	0.6681
UNet++ + ASPP	√	-	0.7652	0.7016
UNet++ + BiConvLSTM	-	√	0.8019	0.7388
BCAS-UNet++	√	√	0.8430	0.7852

As shown in Table 3, incorporating ASPP into the baseline UNet++ increased the Dice coefficient from 0.7081 to 0.7652 and IoU from 0.6681 to 0.7016, indicating that multi-scale spatial feature extraction helps improve the representation of anatomical structures at different scales. In comparison, incorporating BiConvLSTM increased Dice to 0.8019 and IoU to 0.7388, achieving larger improvements than ASPP alone. This suggests that integrating contextual information from adjacent ultrasound frames is particularly beneficial for identifying anatomical boundaries that may be unclear in individual frames.

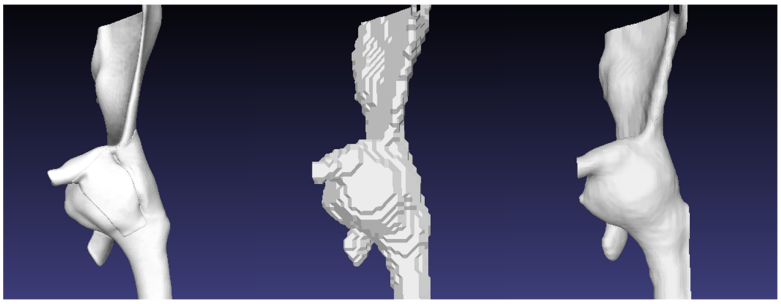
When both modules were incorporated, the complete BCAS-UNet++ achieved the highest Dice coefficient of 0.8430 and IoU of 0.7852, representing absolute improvements of 0.1349 and 0.1171 over the baseline, respectively. These results indicate that ASPP and BiConvLSTM provide complementary benefits by enhancing multi-scale spatial feature extraction and temporal information integration, thereby improving the overall segmentation performance of infant hip ultrasound images.

3.3. 3D Reconstruction Results

Firstly, initial rigid-body registration is performed by selecting anatomically significant corresponding points on both the reconstructed model and the original reference model; subsequently, a multi-scale Iterative Closest Point (ICP) algorithm is employed for fine registration, in order to minimise the spatial distance between the models and ensure that all quantitative and qualitative comparisons are carried out within a unified coordinate system.

3.3.1. Qualitative Analysis

To visually assess the visual quality and structural reconstruction capabilities of the proposed 3D reconstruction method, this study conducted a visual comparison between the original 3D model, the reconstruction results based on the traditional Marching Cubes algorithm, and those based on the MB-Marching Cubes algorithm, as shown in Figure 12.



(**a**) (**b**) (**c**)

Figure 12. Comparison of 3D reconstruction results of the hip joint. (**a**) Original phantom; (**b**) conventional Marching Cubes; (**c**) proposed MB-Marching Cubes.

As shown in Figure 12a, the original 3D model has an intact structure and a smooth surface; the acetabulum and femoral head are clearly defined and exhibit typical anatomical features. Figure 12b shows the 3D reconstruction result based on the traditional Marching Cubes algorithm. It can be seen that it has essentially restored the overall morphology of the hip joint, but there are obvious shortcomings in terms of surface detail: the model edges appear jagged, the margins of the acetabulum and the contour of the femoral head are blurred, and the reconstruction quality is poor. In contrast, the reconstruction results from the MB-Marching Cubes algorithm shown in Figure 12c are visually significantly superior to those of the traditional method. Jagged edges on the model surface have been effectively suppressed; the reconstructed surface is smoother, the anatomical boundaries are more natural, and the results more closely resemble the morphological features of the original model.

3.3.2. Quantitative Assessment

A quantitative assessment was carried out at two levels: surface geometric accuracy and voxel overlap. With regard to surface geometric accuracy, the average surface distance (ASD) and Hausdorff distance (HD) were calculated; the former reflects the overall average deviation, whilst the latter is used to measure extreme errors. Voxel overlap was assessed using the Dice coefficient and IoU to measure the spatial overlap rate of the models. The results are shown in Table 4:

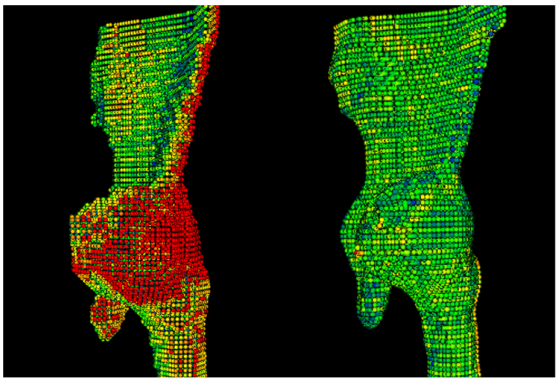
Table 4. Comparative quantitative assessment of 3D reconstruction results.

Evaluation Criteria	Marching Cubes	MB-Marching Cubes
ASD (mm)	1.5640	0.8717

HD (mm)	3.5861	2.3713
Dice	0.7379	0.8536
IoU	0.5847	0.7446

The results show that, compared with the traditional Marching Cubes algorithm, the MB-Marching Cubes algorithm proposed in this paper demonstrates significant advantages across all evaluation metrics. Its ASD (0.87 mm) is considerably lower than that of the traditional algorithm (1.56 mm), indicating that the average deviation between the reconstructed model surface and the original phantom is smaller. The HD was reduced from 3.59 mm to 2.37 mm, indicating that the improved algorithm effectively controls the maximum local error and offers greater robustness in reconstruction at complex anatomical boundaries. Meanwhile, the Dice coefficient increased from 0.74 to 0.85, and the IoU increased from 0.58 to 0.74. Bézier interpolation provides a smoother approximation of local anatomical curvature by replacing planar triangular representations with curved surface patches. Connectivity optimization removes degenerate triangles and isolated mesh fragments, reducing invalid or spurious surface elements that may contribute to reconstruction errors. These effects are consistent with the lower ASD and HD obtained using the proposed MB-Marching Cubes method.

To further visualise the spatial distribution of reconstruction errors, an error heatmap was plotted, as shown in Figure 13. In the figure, blue and green areas indicate errors of less than 1 mm, orange to yellow indicate errors of 1–1.5 mm, and red areas indicate errors greater than 1.5 mm. It can be seen that the algorithm described in this paper significantly reduces errors in high-curvature regions such as the acetabular roof and the femoral head.



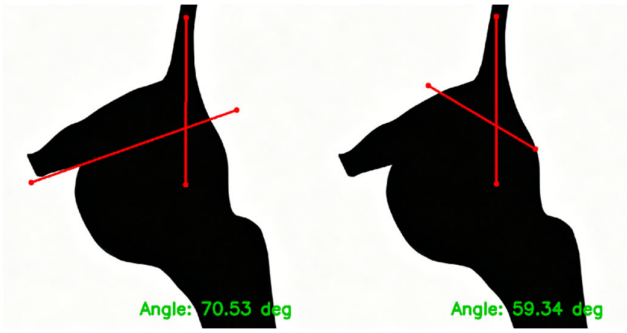
(**a**) (**b**)

Figure 13. Error heatmap. (**a**) Conventional Marching Cubes; (**b**) proposed MB-Marching Cubes. Blue and green indicate errors below 1 mm, yellow and orange indicate errors of 1–1.5 mm, and red indicates errors above 1.5 mm.

3.3.3. Measurement of Graf Parameters and Error Analysis

Using a hip joint model derived from 3D reconstruction, a cross-section was extracted using 3D Slicer software (version 5.8.1), centred on the greater trochanter and clearly displaying the inferior margin of the ilium, the acetabular roof and the acetabular labrum, to serve as the Graf standard plane. Next, within the standard plane, a baseline was drawn along the outer wall of the ilium; starting from the inferior margin of the ilium, a bony acetabular roof line was drawn along the bony acetabular roof. The angle between these two lines is the α angle, which was measured at 66.93°, as shown in Figure 14a; next, starting from the bony margin of the acetabulum and passing through the centre of the acetabular labrum, the cartilaginous acetabular roof line is drawn; the angle between this line and the baseline is the β angle, with a measured value of 70.82°, as shown in Figure 14b.

To evaluate the accuracy of the measurement results, the above results were compared with the parameters of the hip joint phantom (α = 67.63°, β = 72.88°). The absolute angular difference was calculated as the absolute difference between the reconstructed measurement and the reference value, expressed in degrees (°). The relative error was calculated by dividing the absolute angular difference by the reference value and multiplying by 100%, expressed as a percentage (%). Calculations show that the absolute error for the α angle is 0.7°, with a relative error of 1.04%; the absolute error for the β angle is 2.06°, with a relative error of 2.83%. These relatively small angular differences demonstrate encouraging measurement accuracy in the infant hip phantom experiment, supporting the technical feasibility of the proposed method and providing a basis for further clinical evaluation.



(**a**) (**b**)

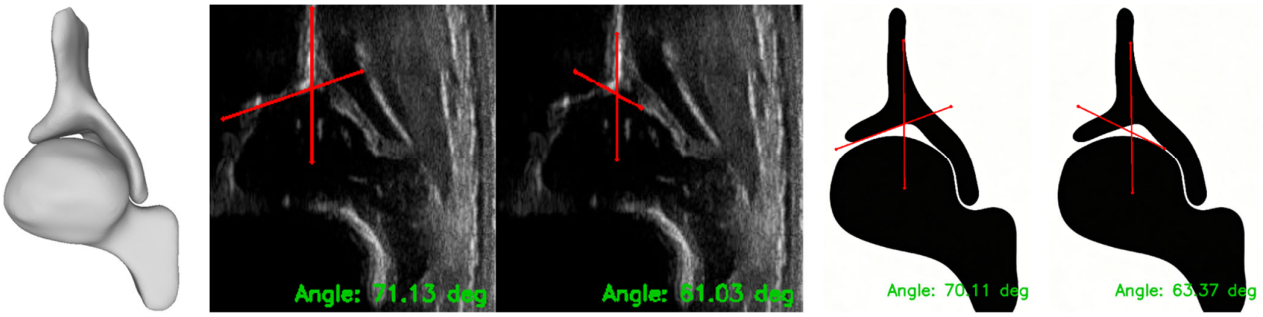
Figure 14. Graf parameter measurement. (**a**) Measured α Angle; (**b**) measured β Angle.

3.3.4. Preliminary Clinical Feasibility Study

Through the aforementioned experiments on an infant hip phantom, the feasibility of locating and extracting the Graf standard planes on a 3D reconstructed hip model for the measurement of α and β angles has been preliminarily validated. To further evaluate the applicability of the method described in this paper in real clinical settings, this section extends the experimental subjects from the infant hip phantom to actual clinical data. In compliance with ethical guidelines, the ultrasound data acquisition system developed in this paper was utilised to collect hip ultrasound images and corresponding spatial pose data from five infants aged 0–3 months at a collaborating hospital. Only the left hip of each infant was examined, resulting in a total of five evaluated hips and five paired sets of α- and β-angle measurements. To minimize spontaneous body movements and improve the stability of ultrasound acquisition, all clinical ultrasound examinations were performed while the infants were asleep. The duration of a complete hip ultrasound sweep varied with the infant's body size and the required scanning range, but typically took approximately one minute. All acquired data were stored in a standardised format (images saved as PNG files, pose data as CSV files) and archived according to case numbers.

For each subject, a clinical expert with more than 20 years of experience identified the standard plane using the conventional 2D Graf method and manually measured the α and β angles by constructing the baseline, bony roof line, and cartilaginous roof line based on the corresponding anatomical landmarks (Figure 15b,c). The anatomical landmark identification, reference line placement, and angular measurements were independently reviewed by two additional clinical experts. Any discrepancies were resolved through discussion until a consensus was reached. The final consensus measurements were used as reference values for subsequent comparison. Subsequently, 3D reconstruction of the clinical data was performed using the method described in this paper (Figure 15a), and the

Graf standard plane was extracted from the 3D reconstruction model to measure the values of angles α and β (Figure 15d,e). Taking Case 1 as an example, the α angle measured using the clinical 2D Graf method was 71.13°, and the β angle was 61.03°; the α angle of the standard plane extracted from the 3D reconstruction model in this study was 70.11°, and the β angle was 63.37°. It was therefore calculated that the absolute error for the α angle was 1.02°, with a relative error of 1.43%; the absolute error for the β angle was 2.34°, with a relative error of 3.83%. The same method was applied to validate the data collected from the five case groups. The results are shown in Table 5.



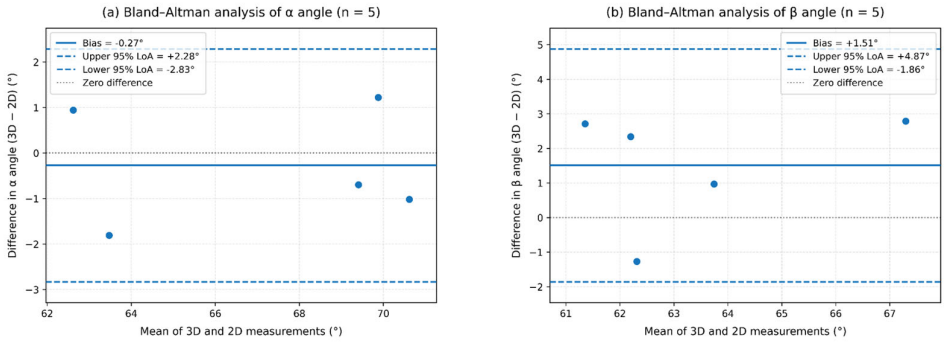
(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 15. Comparison of results from the conventional method and the 3D reconstruction method for Case 1. (**a**) 3D reconstruction results; (**b**) α angle measured by the conventional method; (**c**) β angle measured by the conventional method; (**d**) α angle measured by the proposed method; (**e**) β angle measured by the proposed method.

Table 5. Clinical data measurement results.

Case Number	2D Method α/°	3D Method α/°	Absolute Error	Relative Error	2D Method β/°	3D Method β/°	Absolute Error	Relative Error
1	71.13	70.11	1.02°	1.43%	61.03	63.37	2.34°	3.83%
2	69.76	69.06	0.70°	1.00%	60.00	62.71	2.71°	4.52%
3	62.15	63.09	0.94°	1.51%	62.95	61.68	1.27°	2.02%
4	64.38	62.57	1.81°	2.81%	63.26	64.23	0.97°	1.53%
5	69.27	70.49	1.22°	1.76%	65.90	68.69	2.79°	4.23%
Mean ± SD	/	/	1.14 ± 0.42°	1.70 ± 0.68%	/	/	2.02 ± 0.84°	3.23 ± 1.36%

For the clinical evaluation, the conventional 2D Graf measurements were used as reference values. Bland–Altman analysis was performed using the signed differences between the 3D and 2D measurements (3D − 2D). The mean difference represented the measurement bias, and the 95% limits of agreement were calculated as the mean difference ± 1.96 times the sample standard deviation of the paired differences. Across the five clinical cases, the mean absolute errors of the α and β angles were 1.14 ± 0.42° and 2.02 ± 0.84°, respectively, while the corresponding mean relative errors were 1.70 ± 0.68% and 3.23 ± 1.36% (mean ± SD). Bland–Altman analysis showed a mean bias of −0.27° for the α angle, with 95% limits of agreement ranging from −2.83° to 2.28°. For the β angle, the mean bias was +1.51°, with corresponding limits of agreement from −1.86° to 4.87°, as illustrated in Figure 16. These results provide preliminary evidence of agreement between the two measurement methods.



(**a**) (**b**)

Figure 16. Scatter plot of the Bland-Altman agreement analysis. (**a**) α angle; (**b**) β angle.

4. Discussion

4.1. Main Findings and Clinical Implications

Currently, 2D ultrasound screening remains a commonly used approach for the diagnosis of DDH. However, the acquisition and interpretation of standard planes require substantial professional expertise and extensive clinical experience. In recent years, numerous studies have focused on addressing key challenges such as anatomical structure segmentation, automated Graf-angle measurement, and standard-plane detection to assist healthcare professionals. However, most existing methods primarily analyze individual planes or static images, and only a limited number of studies have explored the 3D visualization of infant hip joint anatomy and its clinical utility.

The infant hip joint exhibits complex 3D anatomical structures; therefore, understanding its spatial relationships solely from 2D ultrasound images remains challenging. In this study, we implemented and evaluated a 3D visualization method for the infant hip joint based on 2D ultrasound image sequences. This approach can assist healthcare professionals in obtaining a more intuitive understanding of the overall hip joint anatomy and its spatial relationships.

This method utilises the BCAS-UNet++ ultrasound image segmentation network, which fuses spatio-temporal features to produce continuous and clean hip joint segmentation results; it also employs an MB-Marching Cubes 3D reconstruction algorithm based on cubic triangular Bézier surface interpolation, incorporating a nine-parameter interpolation method and connectivity optimisation to generate continuous and smooth 3D models of the infant's hip joint; Subsequently, the standard plane was localised based on the 3D reconstruction model and the values of the α and β angles were measured, which were then validated against the true values from a phantom; finally, clinical data validation was conducted.

In the preliminary clinical evaluation, the mean relative errors for the α and β angles were 1.70% and 3.23%, respectively. These relatively small measurement discrepancies are encouraging and suggest the potential of the proposed method for assisting Graf-angle assessment, although further validation in larger, independent clinical cohorts is needed to confirm its clinical reliability. Furthermore, this method enables the acquisition of a complete 3D model of the infant's hip joint whilst remaining compatible with existing clinical equipment. It also decouples the ultrasound scanning process from the standard plane localisation procedure, meaning that the localisation of the standard plane no longer relies entirely on real-time adjustments to the probe's position and orientation, but is instead performed through positioning and cropping within the 3D visualisation model. Furthermore, the 3D reconstruction provides a comprehensive spatial representation of the hip joint, supporting post-examination plane re-slicing and repeated measurements,

thereby facilitating retrospective analysis and potentially improving the consistency of standard-plane assessment.

4.2. Study Limitations and Future Directions

However, this study also has some limitations. Firstly, the diagnosis of DDH for the 3D reconstructed model still relies on a 2D method involving the identification of standard planes; further research should therefore focus on developing 3D quantitative diagnostic protocols better suited to 3D models. Furthermore, the clinical evaluation included only five left hips from five infants aged 0–3 months at a single clinical center. The small sample size, unilateral assessment, and limited age range restrict the generalizability of the findings. Unlike the phantom experiments, which used the original 3D model as a geometric reference, the clinical cases lacked corresponding in vivo volumetric ground truth. Therefore, clinical evaluation focused on comparing the reconstructed α and β angles with conventional 2D Graf measurements rather than directly assessing volumetric overlap or surface distances. The clinical results should be interpreted as preliminary evidence of technical feasibility. Future studies involving larger and more diverse clinical cohorts are needed to further validate the accuracy and reliability of the proposed method.

Infant motion during ultrasound scanning is another potential limitation of the proposed framework. In our previous studies, robotic contact force control methods were developed to maintain stable probe–skin interaction during DDH ultrasound scanning [29,30]. In the present clinical evaluation, ultrasound data were acquired while the infants were asleep to minimize spontaneous movements. These measures help improve acquisition stability; however, they do not directly compensate for anatomical displacement caused by involuntary infant motion. As the current reconstruction process does not incorporate explicit inter-frame motion correction, substantial motion may still affect the spatial consistency of ultrasound sequences and the accuracy of 3D reconstruction. Future work will investigate motion detection and registration methods to further improve robustness under clinical conditions.

Furthermore, although the hydrogel phantom was designed to simulate the anatomical structures and acoustic properties of the infant hip joint, differences from living neonatal tissues remain. These differences are mainly reflected in ultrasound image clarity and tissue-boundary contrast and may be associated with variations in acoustic impedance and other acoustic properties. Despite these differences, the phantom provided a controlled experimental environment for evaluating the proposed method, while the preliminary clinical results further supported its feasibility under real examination conditions.

5. Conclusions

In this paper, we developed and evaluated a 3D visualization method based on 2D ultrasound image sequences. By acquiring sequential 2D ultrasound images of the infant hip joint, the proposed method enables 3D reconstruction of the hip joint, visual localization of the Graf standard plane, and accurate measurement of the α and β angles. This approach allows the ultrasound scanning process to be separated from the diagnostic assessment. By reconstructing a 3D visualization model of the infant hip joint, the method provides an additional anatomical perspective for DDH assessment, facilitates Graf standard-plane localization, and may help reduce reliance on real-time probe adjustments. In the phantom experiments, BCAS-UNet++ achieved a segmentation Dice coefficient of 0.8430 and an IoU of 0.7852. Compared with conventional Marching Cubes, the proposed MB-Marching Cubes reduced the ASD from 1.5640 mm to 0.8717 mm and the HD from 3.5861 mm to 2.3713 mm, with relative errors of 1.04% and 2.83% for the α and β angles, respectively. In the preliminary clinical evaluation involving five infants, the mean absolute errors of the α and β angles were 1.14 ± 0.42° and 2.02 ± 0.84°, respectively, with

corresponding mean relative errors of 1.70% and 3.23%. Overall, this proof-of-concept study provides preliminary evidence of the technical feasibility of the proposed robot-assisted 3D visualization framework for infant hip assessment. In future work, its reliability will be further validated using larger-scale clinical datasets to facilitate its progression toward clinical trials.

Author Contributions: J.C. and J.W. managed the project. W.Z. designed the program and drafted the first version of the manuscript. Y.D. designed the project. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Foundation of Yangzhou Key Laboratory of Robotic System for Hip Joint Ultrasound Screening (No. YZ2024255), the Jiangsu Provincial Frontier Technology Research and Development Program (No. SBF20260303245) and Yangzhou Municipal Natural Science Foundation (No. YZ2025159).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Yangzhou Maternal and Child Health Care Hospital [approval no. (2024) Ethics Approval No. 021; approval date: 24 May 2024].

Informed Consent Statement: Written informed consent was obtained from the parents or legal guardians of all infants included in the study for the research use and academic publication of their deidentified clinical data, ultrasound images, and videos.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to approval by the relevant ethics committee and data-owning institution. Owing to privacy, ethical, and institutional restrictions associated with clinical ultrasound examinations involving infants, the raw data are not publicly available.

Acknowledgments: The authors would like to express their sincere gratitude to Jiakuan Wang and the clinical team at Yangzhou Maternal and Child Health Care Hospital for their valuable assistance with ultrasound data collection, standard plane assessment, image annotation, manual image segmentation, and Graf-angle measurement. Their clinical expertise and dedicated support contributed substantially to the construction and validation of the infant hip ultrasound dataset used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DDH	developmental dysplasia of the hip
2D	two-dimensional
3D	three-dimensional
BCAS-UNet++	BiConvLSTM ASPP UNet++
ASPP	atrous spatial pyramid pooling
BiConvLSTM	bidirectional convolutional long short-term memory
ASD	average surface distance
HD	Hausdorff distance

References

1. Riboni, G.; Bellini, A.; Serantoni, S.; Rognoni, E.; Bisanti, L. Ultrasound screening for developmental dysplasia of the hip. *Pediatr. Radiol*. **2003**, *33*, 475–481. https://doi.org/10.1007/s00247-003-0940-7.
2. Kang, Y.R.; Koo, J. Ultrasonography of the pediatric hip and spine. *Ultrasonography* **2017**, *36*, 239–251. https://doi.org/10.14366/usg.16051.
3. Heneghan, M. Developmental dysplasia of the hip. *JAAPA* **2021**, *34*, 48–49. https://doi.org/10.1097/01.JAA.0000758256.52498.da.

4. Swarup, I.; Penny, C.L.; Dodwell, E.R. Developmental dysplasia of the hip: An update on diagnosis and management from birth to 6 months. *Curr. Opin. Pediatr.* **2018**, *30*, 84–92. https://doi.org/10.1097/MOP.0000000000000574.
5. Engesæter, I.Ø.; Lie, S.A.; Lehmann, T.G.; Furnes, O.; Vollset, S.E.; Engesæter, L.B. Neonatal hip instability and risk of total hip replacement in young adulthood. *Acta Orthop.* **2008**, *79*, 321–326. https://doi.org/10.1080/17453670710015201.
6. Graf, R. The diagnosis of congenital hip-joint dislocation by the ultrasonic Combound treatment. *Arch. Orthop. Trauma Surg.* **1980**, *97*, 117–133. https://doi.org/10.1007/BF00450934.
7. Graf, R.; Lercher, K.; Scott, S.; Spieß, T. *Essentials of Infant Hip Sonography According to Graf*; Edition Stolzalpe Sonocenter: Stolzalpe, Austria, 2017.
8. Graf, R. *Hip Sonography: Diagnosis and Management of Infant Hip Dysplasia*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006. https://doi.org/10.1007/3-540-30958-6.
9. Geng, C.; Xu, H.; Zhan, X.; Li, L.; Song, Q.; Zhang, L.; Ge, L. Using 3-dimensional ultrasound iSlice technology for the diagnosis of developmental dysplasia of the hip. *J. Ultrasound Med*. **2020**, *39*, 1117–1123. https://doi.org/10.1002/jum.15193.
10. Kolb, A.; Benca, E.; Willegger, M.; Puchner, S.E.; Windhager, R.; Chiari, C. Measurement considerations on examiner-dependent factors in the ultrasound assessment of developmental dysplasia of the hip. *Int. Orthop*. **2017**, *41*, 1245–1250. https://doi.org/10.1007/s00264-017-3455-9.
11. He, J.; Cui, L.; Chen, T.; Lyu, X.; Yu, J.; Guo, W.; Wang, D.; Qin, X.; Zhao, Y.; Zhang, S. Study on multiplanar measurements of infant hips with three-dimensional ultrasonography. *J. Clin. Ultrasound* **2022**, *50*, 639–645. https://doi.org/10.1002/jcu.23185.
12. Mostofi, E.; Chahal, B.; Zonoobi, D.; Hareendranathan, A.; Roshandeh, K.P.; Dulai, S.K.; Jaremko, J.L. Reliability of 2D and 3D ultrasound for infant hip dysplasia in the hands of novice users. *Eur. Radiol*. **2019**, *29*, 1489–1495. https://doi.org/10.1007/s00330-018-5699-1.
13. El-Hariri, H.; Mulpuri, K.; Hodgson, A.; Garbi, R. Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*; Springer: Cham, Switzerland, 2019; pp. 12–20. https://doi.org/10.1007/978-3-030-32245-8_2.
14. Zhang, Z.; Tang, M.; Cobzas, D.; Zonoobi, D.; Jagersand, M.; Jaremko, J.L. End-to-end detection-segmentation network with ROI convolution. In *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018*; IEEE: New York, NY, USA, 2018; pp. 1509–1512. https://doi.org/10.1109/ISBI.2018.8363859.
15. Jaremko, J.L.; Hareendranathan, A.; Bolouri, S.E.S.; Frey, R.F.; Dulai, S.; Bailey, A.L. AI aided workflow for hip dysplasia screening using ultrasound in primary care clinics. *Sci. Rep*. **2023**, *13*, 9224. https://doi.org/10.1038/s41598-023-35603-9.
16. Zhang, X.; Cui, J.; Dai, Y.; Zhang, W. DDH-Net: A Graf-compliant method for computer-aided assessment using infant hip ultrasound. *Bioengineering* **2026**, *13*, 1079. https://doi.org/10.3390/bioengineering13091079.
17. Zonoobi, D.; Hareendranathan, A.; Mostofi, E.; Mabee, M.; Pasha, S.; Cobzas, D.; Rao, P.; Dulai, S.K.; Kapur, J.; Jaremko, J.L. Developmental hip dysplasia diagnosis at three-dimensional US: A multicenter study. *Radiology* **2018**, *287*, 1003–1015. https://doi.org/10.1148/radiol.2018172592.
18. Zheng, G.; von Recum, J.; Nolte, L.-P.; Grützner, P.A.; Steppacher, S.D.; Franke, J. Validation of a statistical shape model-based 2D/3D reconstruction method for determination of cup orientation after THA. *Int. J. Comput. Assist. Radiol. Surg.* **2012**, *7*, 225–231. https://doi.org/10.1007/s11548-011-0644-7.
19. Schumann, S.; Liu, L.; Tannast, M.; Bergmann, M.; Nolte, L.-P.; Zheng, G. An integrated system for 3D hip joint reconstruction from 2D X-rays: A preliminary validation study. *Ann. Biomed. Eng*. **2013**, *41*, 2077–2087. https://doi.org/10.1007/s10439-013-0822-6.
20. Wan, D.; Wang, D.; Ma, A.; Dai, K.; Ai, S.; Wang, L. Automatic outer surface extraction of femoral head in CT images. *J. Shanghai Jiaotong Univ. (Sci.)* **2017**, *22*, 377–384. https://doi.org/10.1007/s12204-017-1847-z.
21. Zhang, S.; Wang, J.; Wang, X.; Konan, S.; Stoyanov, D.; Mazomenos, E.B. 3D acetabular surface reconstruction from 2D pre-operative X-ray images using SRVF elastic registration and deformation graph. In *Medical Image Computing and Computer Assisted Intervention—MICCAI* 2025; Springer: Cham, Switzerland, 2025; Volume 15975, pp. 3–12. https://doi.org/10.1007/978-3-032-05325-1_1.
22. Pham, T.-T.; Le, L.H.; Andersen, J.; Lou, E.H. Optimal configurations of an electromagnetic tracking system for 3D ultrasound imaging of pediatric hips—A phantom study. *Med. Eng. Phys*. **2024**, *131*, 104221. https://doi.org/10.1016/j.medengphy.2024.104221.
23. Hoang, T.H.; Pham, T.-T.; Lou, E.H.; La, T.-G.; Wang, J.; Le, L.H. 3D ultrasonic visualization of hip displacement: A phantom study. *Ultrasonics* **2026**, *158*, 107808. https://doi.org/10.1016/j.ultras.2025.107808.

24. Visutsak, P. Marching cubes and histogram pyramids for 3D medical visualization. *J. Imaging* **2020**, *6*, 88. https://doi.org/10.3390/jimaging6090088.
25. Powell, M.J.D.; Sabin, M.A. Piecewise quadratic approximations on triangles. *ACM Trans. Math. Softw*. **1977**, *3*, 316–325. https://doi.org/10.1145/355759.355761.
26. Barnhill, R.; Farin, G. C^1 quintic interpolation over triangles: Two explicit representations. *Int. J. Numer. Methods Eng.* **1981**, *17*, 1763–1778. https://doi.org/10.1002/nme.1620171203.
27. Mann, S. Cubic precision Clough–Tocher interpolation. *Comput. Aided Geom. Des*. **1999**, *16*, 85–88. https://doi.org/10.1016/S0167-8396(98)00038-7.
28. Vlachos, A.; Peters, J.; Boyd, C.; Mitchell, J.L. Curved PN triangles. In *Proceedings of the 2001 Symposium on Interactive 3D Graphics, Research Triangle Park, NC, USA, 19–21 March 2001*; ACM: New York, NY, USA, 2001; pp. 159–166. https://doi.org/10.1145/364338.364387.
29. Cui, J.; Zhang, X.; Dai, Y.; Zhang, W. Research on Robotic Force Control for Infant Hip Ultrasound. *Actuators* **2026**, *15*, 333. https://doi.org/10.3390/act15060333.
30. Cui, J.; Dai, Y.; Zhang, X.; Xiong, Y.; Zhang, W. Preliminary Research on Autonomous Robotic System for DDH Ultrasound Examination. *Actuators* **2026**, *15*, 446. https://doi.org/10.3390/act15080446.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.