actuators

MDPI

Article

Preliminary Research on Autonomous Robotic System for DDH Ultrasound Examination

Jianwei Cui *, Yuxiang Dai, Xinyu Zhang, Yao Xiong and Wenyi Zhang

Institute of Instrument Science and Engineering, Southeast University, Nanjing 210096, China; 220243737@seu.edu.cn (Y.D.); 220253715@seu.edu.cn (X.Z.); bearshine668@gmail.com (Y.X.); 220243765@seu.edu.cn (W.Z.)

* Correspondence: cjw@seu.edu.cn

Abstract

Ultrasound examination for developmental dysplasia of the hip (DDH) in infants is highly dependent on operator experience, leading to inconsistent imaging quality and poor reproducibility between sonographers. This study proposes an autonomous robotic ultrasound system to improve the standardization and automation of hip ultrasound examinations. The system consists of a robotic arm, a six-axis force/torque sensor, an RGB-D camera and an ultrasound probe, integrating multiple functions including contact force control, visual localization, deep-learning-based segmentation and ultrasound image screening. To ensure stability and safety during scanning, an admittance-based hybrid force/position control strategy is adopted to achieve constant contact force control. For Graf standard plane acquisition, a stage-wise search strategy is designed, in which the search space is progressively narrowed through femoral head searching and multi-angle scanning. The optimal Graf standard plane is then automatically selected by combining image segmentation with a scoring mechanism. A customized hip phantom was used for validation. Experimental results show that the Dice coefficient for femoral head segmentation reaches 0.872, while the average Dice coefficient for multi-structure segmentation reaches 0.866. In 30 autonomous scanning trials, the success rate of Graf standard plane acquisition is 90.0%. Meanwhile, the system can maintain the contact force stably within the target range during scanning, validating the effectiveness of the force control strategy. These results indicate that the proposed robotic system, image recognition algorithm and visual servo control strategy exhibit favorable safety and feasibility, providing an innovative solution for automated infant hip ultrasound examination of DDH.

Keywords: robotic ultrasound system; hip examination; force control; ultrasound image segmentation

Check for updates

Academic Editor: Zhuming Bi

Received: 5 July 2026
Revised: 13 August 2026
Accepted: 15 August 2026
Published: 16 August 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

Developmental dysplasia of the hip (DDH) is one of the most common musculoskeletal disorders in infants, with an incidence of approximately 1–2‰ [1]. If not detected and treated in time, DDH may severely affect infant growth and may even lead to deformity and lifelong disability [2]. Therefore, hip examination has become an important component of infant screening. Among medical imaging, ultrasound has become the preferred method for infant hip examination due to its radiation-free, low-cost, and real-time imaging [3]. The Graf method is one of the most widely used ultrasound-based diagnostic methods for DDH. This method first requires the identification of a Graf standard plane containing

eight key imaging features, such as the lower limb of the ilium, the acetabular roof and the acetabular labrum. Diagnosis is then performed based on anatomical measurements of the α and β angles on the Graf standard plane [4]. However, in current manual scanning, the probe pose, image recognition, and feature measurement are all highly dependent on clinicians' experience, leading to large errors and low consistency in diagnostic results [5,6].

To address the limitations of manual ultrasound examination, many studies have gradually introduced robotic technology into the ultrasound imaging process to improve the accuracy, stability, and repeatability of scanning [7]. According to the involvement of the sonographer, robotic ultrasound systems are generally classified into three categories: teleoperated robots, semi-autonomous robots, and autonomous robots. In existing teleoperated robotic systems, the sonographer can use a joystick, a remote-control handle, a virtual ultrasound probe, or a haptic device to remotely control a robot carrying a real ultrasound probe to perform ultrasound scanning [8–11]. With the development of 5G and collaborative robotic technologies, various teleoperated robotic ultrasound systems have been gradually applied to telemedicine and primary care diagnostic scenarios [12,13]. In semi-autonomous robotic systems, sonographers and robots collaboratively control the probe. The robot is mainly responsible for maintaining stable probe-body contact and performing local motion control, whereas the sonographer makes high-level decisions, such as selecting the scanning region and assessing ultrasound image quality. For example, Bao et al. proposed an ultrasound robot integrating force/torque measurement and control, which can adjust the probe force according to clinical requirements, thereby improving the safety and stability of the scanning process [14]. Laurent et al. proposed the Echo-Robot system, which adjusts the probe pose based on ultrasound image quality assessment and standard-view classification, enabling semi-autonomous cardiac ultrasound scanning with visual feedback [15]. Zielke et al. used real-time semantic segmentation results to guide robotic scanning trajectory adjustment, achieving stable imaging and volume estimation of the thyroid region [16]. However, both teleoperated and semi-autonomous robotic systems still rely on sonographers' experience during key operations and decision-making processes, such as target region localization, scanning path planning, and final image interpretation. Therefore, these methods cannot fundamentally solve the problem of uneven distribution of medical resources. In contrast, autonomous robotic systems do not require continuous sonographer involvement and can independently complete ultrasound scanning and target image acquisition, representing an important development direction for robotic ultrasound systems [17].

Huang et al. proposed a robot-assisted autonomous carotid ultrasound imaging method, in which target localization and path planning were achieved through pre-scanning and scanning stages [18]. Wang et al. developed an autonomous carotid ultrasound scanning system based on visual servo navigation, improving scanning continuity through target tracking and recovery mechanisms [19]. Zakeri et al. used deep ultrasound image features for visual servo control to achieve stable tracking of standard cardiac views [20]. Su et al. proposed a fully autonomous robotic system for thyroid ultrasound and combined reinforcement learning with Bayesian optimization for probe pose adjustment [21]. Building on this work, Su et al. introduced a tissue-view map to model probe-tissue interactions and improve continuous decision-making during autonomous scanning [22]. Sun et al. developed an automatic ultrasound framework for the musculoskeletal system, integrating target localization, trajectory generation, image segmentation, and 3D reconstruction [23]. Roshan et al. combined RGB-D and thermal imaging for automatic localization and scanning path generation in the lumbar region [24]. Lee et al. combined RGB-D vision with adaptive force control to achieve automatic assessment of vascular stenosis and autonomous adjustment of the ultrasound scanning process [25]. Jiang et al. proposed a large-scale learning-based

autonomous carotid ultrasound system. By constructing a large expert-demonstration dataset and combining it with an imitation learning framework, their system achieved end-to-end autonomous scanning performance approaching that of clinical experts [26]. Overall, existing autonomous robotic ultrasound systems have mainly been developed around visual servo control, learning-driven strategies, and multimodal perception fusion. These systems have been preliminarily validated in different clinical scenarios.

Although autonomous robotic ultrasound has made progress in applications such as thyroid, carotid artery, and cardiac imaging, research on autonomous scanning for infant DDH examination remains limited. First, infant DDH scanning imposes stricter safety requirements. During scanning, the robot must not only maintain stable contact between the probe and the body surface but also avoid excessive contact force that may cause discomfort to infants. Second, the identification of the Graf standard plane depends on multiple anatomical features, including the ilium, femoral head, labrum, and chondro-osseous junction. However, existing systems mostly rely on fixed scanning paths and single features, and they lack a search strategy designed for the multi-structure characteristics of the Graf standard plane. Therefore, they are difficult to directly apply to DDH ultrasound scanning. Accordingly, developing an autonomous robotic system for DDH ultrasound examination is of great significance for reducing dependence on skilled sonographers and improving the consistency of examination.

This study proposes an autonomous robotic ultrasound system for DDH ultrasound examination. The main contributions of this work are summarized as follows:

(1) We design a quantitative scoring mechanism for Graf standard plane features. The presence and completeness of multiple anatomical structures are quantified and integrated into a unified score, enabling objective ranking of candidate images and automatic selection of the optimal Graf standard plane.
(2) We propose a standardized stage-wise scanning workflow for Graf standard plane acquisition. The workflow formalizes expert sonographers' scanning experience into a unified robotic procedure. It progressively narrows the search space, thereby improving the consistency and repeatability of Graf standard plane acquisition.
(3) We develop and validate an autonomous robotic ultrasound system for DDH examination. The system integrates the proposed scanning workflow with visual localization, compliant force control, and ultrasound image analysis. Phantom experiments verify the feasibility of the overall framework.

The remainder of this paper is organized as follows. Section 2 introduces the composition of the robotic ultrasound system, system calibration, compliant contact-force control, ultrasound image feature extraction, and the Graf standard plane search strategy. Section 3 presents the experimental results obtained on a hip phantom. Section 4 discusses the principal findings, significance, limitations, and future research directions of the study. Section 5 summarizes the main conclusions.

2. Materials and Methods

2.1. Experimental Platforms

The proposed autonomous robotic ultrasound system consists of four main modules: a six-degree-of-freedom robotic arm (RM65-B, RealMan, Beijing, China), a six-axis force/torque sensor (XJC-6F-D40-H18-A, XJCSENSOR, Shenzhen, China), an RGB-D camera (Gemini Pro, ORBBEC, Shenzhen, China), and a commercial handheld wireless ultrasound probe (D8c, Youkey Medical, Wuhan, China). The overall configuration of the system is shown in Figure 1.

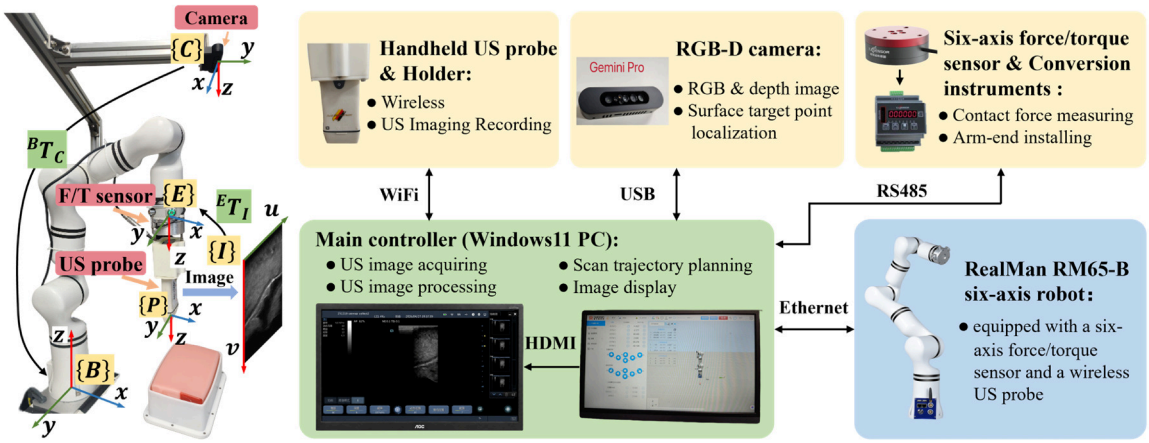


Figure 1. Structure of the autonomous robotic ultrasound system. The left shows the RM65-B robot equipped with an RGB-D camera, a force/torque sensor, and an ultrasound probe, together with all coordinate systems involved and the corresponding transformation matrices. The right shows the connections among the system modules.

The robotic arm communicates with the main control computer through a wired Ethernet connection using the UDP protocol. This communication channel is used for transmitting robot control commands and receiving robot state information. The six-axis force/torque sensor is mounted at the robot end effector through a flange and connected to an XJC-620F-6 conversion instrument. It measures the force components along the X, Y, and Z axes and the torque components around these axes. The measurement ranges are ± 20 N for force and ± 1 N·m for torque, with resolutions of 0.002 N and 1×10^{-4} N·m, respectively. The nonlinearity and hysteresis errors are both within $\pm 0.5\%$ F.S., and the repeatability error is within $\pm 0.05\%$ F.S. The conversion instrument acquires the force/torque measurements at 100 Hz and transmits them to the main control computer through an RS485 serial communication interface. The ultrasound probe is connected to the force/torque sensor through a customized probe holder. The holder is designed with a buffering mechanism to reduce instantaneous impacts that may occur during robot motion and to improve the stability of probe-skin contact. With this configuration, the force/torque sensor can directly measure the interaction force and torque between the ultrasound probe and the body surface during scanning. The ultrasound probe communicates with the main control computer via WiFi, and B-mode ultrasound images are acquired and recorded using the dedicated ultrasound application SonoiQ. The operating frequency range of the probe is 6–11 MHz, which is suitable for musculoskeletal ultrasound imaging. The RGB-D camera is installed above the robotic workspace in an eye-to-hand configuration. It is used to acquire RGB and depth images of the scanning region and to estimate the position of the body-surface target point.

The main controller of the system is a personal computer (Intel i7-13620H, CPU @ 2.40 GHz, 16 GB RAM), which is mainly responsible for robot control, data acquisition, image segmentation, and scan trajectory planning. The related functions were developed based on Visual Studio Code 1.106.2, and the software environment was managed using Anaconda to improve system stability and reproducibility.

Considering the safety and ethical limitations of directly conducting experiments on infant subjects, a customized hip gelatin phantom was first used as the experimental object in this study to construct a repeatable experimental scenario for infant hip examination. The phantom has skin-like softness in its external appearance, allowing stable contact and continuous scanning with the robotic ultrasound probe. Meanwhile, its internal structure was designed to present key anatomical structures related to the Graf standard plane in

ultrasound imaging, including the ilium, femoral, labrum, chondro-osseous junction and synovial fold, as shown in Figure 2.

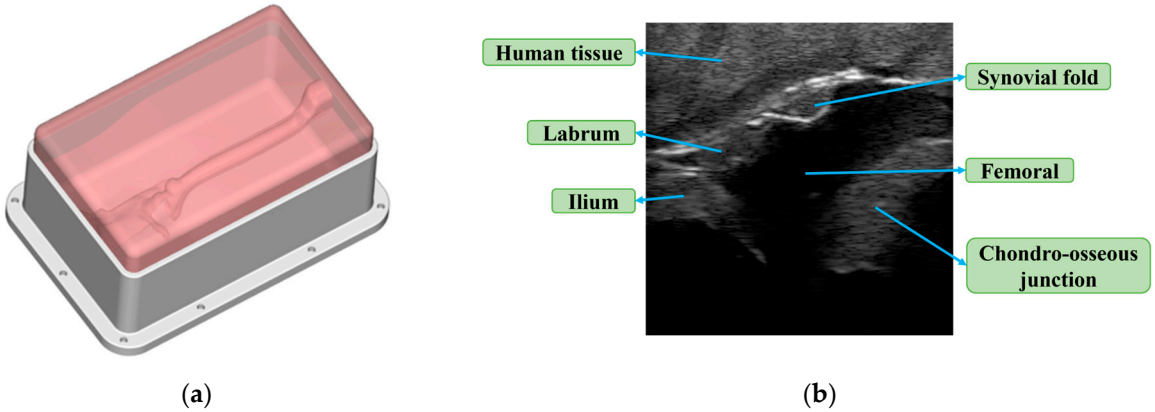


Figure 2. (**a**) Three-dimensional model of the hip gelatin phantom; (**b**) Graf standard plane ultrasound image acquired from the phantom and the corresponding key anatomical structures.

2.2. System Calibrations

In the proposed robotic system, calibration is required to unify the coordinate frames and measurement references of different modules. The calibration mainly includes coordinate calibration and gravity compensation calibration. Coordinate calibration consists of ultrasound image-to-robot calibration and RGB-D camera-to-robot calibration, while gravity compensation calibration is used to remove the influence of the end-effector weight and sensor offset on force measurement.

2.2.1. Ultrasound Image Calibration

Ultrasound image calibration is performed to determine the transformation matrix ${}^{E}T_{I}$ from the ultrasound image coordinate frame $\{I\}$ to the robot end-effector coordinate frame $\{E\}$, together with the scale matrix T_s for converting pixel dimensions into physical dimensions. Following the method in [27], these parameters are estimated by minimizing the position error of the calibration feature points:

$$f = min \sum_{i=1}^{n} \left\| {}^{B}T_{E,i} {}^{E}T_{I} T_s p_i^I - p_i^B \right\|, \tag{1}$$

where ${}^{B}T_{E,i}$ denotes the end-effector pose relative to the robot base frame when the i-th image is acquired, while p_i^I and p_i^B denote the corresponding calibration feature points in the ultrasound image and robot base coordinate frames, respectively.

After calibration, any pixel point $p^I = [u, v, 0, 1]^T$ can be mapped to the robot base coordinate frame as:

$$p^B = {}^{B}T_{E} {}^{E}T_{I} T_s p^I. \tag{2}$$

This transformation is used to convert ultrasound image features into robot coordinates for image-guided probe adjustment.

2.2.2. RGB-D Camera-to-Robot Calibration

RGB-D camera-to-robot calibration is performed to determine the transformation matrix ${}^{B}T_{C}$ from the RGB-D camera coordinate frame $\{C\}$ to the robot base coordinate frame $\{B\}$. Since the camera is fixed in an eye-to-hand configuration, ${}^{B}T_{C}$ remains constant after calibration and is estimated using the classical hand-eye calibration method [28].

After calibration, a point $p^C = \left[p_x^C, p_y^C, p_z^C, 1\right]^T$ acquired by the RGB-D camera can be transformed into the robot base coordinate frame as:

$$p^B = {}^B T_C p^C. \tag{3}$$

For a target pixel (u_c, v_c) with the corresponding depth z_c, its three-dimensional coordinate in the camera coordinate frame is calculated as:

$$p^C = \begin{bmatrix} z_c(u_c - c_x)/f_x \\ z_c(v_c - c_y)/f_y \\ z_c \\ 1 \end{bmatrix}, \tag{4}$$

where f_x and f_y are the focal lengths of the camera; c_x and c_y are the principal point coordinates of the RGB-D camera.

Equations (3) and (4) are used to convert the selected body-surface target point into the robot base coordinate frame for initial probe positioning.

2.2.3. Gravity Compensation Calibration

Gravity compensation calibration is performed to remove the effects of the end-effector tool gravity and sensor zero offset from the raw force/torque measurements. Since the robotic ultrasound scanning process is performed at a low speed, the inertial force of the end-effector tool is neglected. The force/torque sensor coordinate frame $\{S\}$ is aligned with the robot end-effector coordinate frame, and its pose can therefore be obtained directly from the robot end-effector pose.

Let the robot base coordinate frame be $\{B\}$. The compensated contact force is calculated as:

$$F_c^S = F_{raw}^S - G^S - F_0^S, \tag{5}$$

where F_{raw}^S denotes the raw force measured by the sensor, F_c^S denotes the actual contact force, G^S denotes the gravity component of the end-effector tool in the sensor coordinate frame, and F_0^S denotes the sensor zero offset.

Since gravity has a fixed direction in the robot base coordinate frame, its component in the sensor coordinate frame is calculated as:

$$G^S = {}^S R_B G^B = {}^S R_B \begin{bmatrix} 0 \\ 0 \\ -G \end{bmatrix}, \tag{6}$$

where G denotes the total weight of the ultrasound probe, probe holder, and end-effector connector, and ${}^S R_B$ is the rotation matrix from the robot base coordinate frame to the sensor coordinate frame.

During calibration, force measurements are collected at multiple end-effector poses while the ultrasound probe remains free from external contact, such that $F_c^S = 0$. The end-effector tool weight G and sensor zero offset F_0^S are then estimated using the least-squares method:

$$\min_{G, F_0^S} \sum_{i=1}^{n} \left\| F_{raw,i}^S - {}^S R_B \begin{bmatrix} 0 \\ 0 \\ -G \end{bmatrix} - F_0^S \right\|^2. \tag{7}$$

After calibration, G^S is calculated in real time according to the current robot end-effector pose, and the compensated contact force F_c^S is obtained using Equation (5) for subsequent contact force control. After compensation, the RMSE values of $F_{c,x}$, $F_{c,y}$, and

$F_{c,z}$ were 0.0834, 0.0455, and 0.0571 N, respectively. The RMSE values of $M_{c,x}$, $M_{c,y}$, and $M_{c,z}$ were 4.34×10^{-4}, 1.11×10^{-3}, and 7.61×10^{-4} N·m, respectively.

2.3. Contact Force Compliance Control

An admittance-based hybrid force/position control method was adopted to maintain stable probe-surface contact during scanning, as shown in Figure 3. The tangential motion of the probe followed the planned scanning trajectory, while its position along the surface-normal direction was adjusted according to the contact force error. The detailed control method has been presented in our previous work [29].

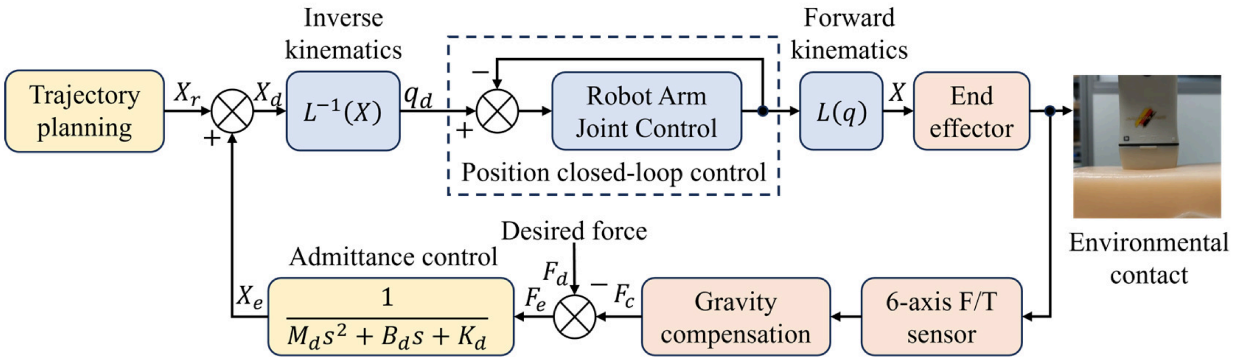


Figure 3. Hybrid force/position control framework.

The normal displacement correction was generated using a second-order admittance model:

$$M_d \ddot{x}_e + B_d \dot{x}_e + K_d x_e = F_e, \tag{8}$$

where x_e is the displacement correction generated by the admittance controller, and $\dot{x}_e$ and $\ddot{x}_e$ are its velocity and acceleration.

The admittance parameters were empirically tuned on the infant hip phantom. The virtual mass M_d was first fixed to limit rapid probe acceleration. The stiffness K_d was then gradually increased to reduce excessive displacement while maintaining sufficient compliance. Finally, the damping coefficient B_d was adjusted to suppress force oscillations without causing an overly slow response. Based on repeated contact and scanning experiments, the parameters were set to $M = 0.25$ kg, $K = 1250$ N/m, and $B = 25$ N · s/m.

A Kalman filter was applied to the measured force signal to reduce measurement noise [29]. The admittance controller was executed at 50 Hz with a control period of 0.02 s. The resulting normal displacement correction was superimposed on the planned scanning trajectory and sent to the robot controller.

To ensure safety, contact force thresholds are set. When the contact force exceeds 6 N, the robot pauses scanning and retracts along the normal direction. If the force remains excessive or exceeds 10 N, the scanning process is terminated and the robot returns to the initial safe pose.

2.4. Feature Extraction of Ultrasound Image

Ultrasound image features provide the basis for Graf standard plane searching. In this study, semantic segmentation is used to extract target structures from ultrasound images. All input images are two-dimensional B-mode ultrasound images with a size of 256×256 pixels.

SegNeXt is adopted as the basic segmentation network. Its lightweight convolutional design and multi-scale feature modeling enable efficient extraction of tissue boundaries and local structural features in ultrasound images with low computational cost [30,31]. To

meet the different requirements for inference speed and segmentation accuracy at different stages, two SegNeXt-based networks are trained. Both networks use an encoder–decoder structure and MSCAN as the feature extraction backbone, but they differ in encoder size, output classes, and application stages.

The first network is SegNeXt-T, which is used for femoral head segmentation during the searching stage. It uses the lightweight MSCAN-T backbone and outputs a binary segmentation result of the background and femoral head. Based on the segmentation mask, the pixel area and center position of the femoral head are extracted. The femoral head region is defined as:

$$R_{fh} = \{(u, v) | M_T(u, v) = 1\}, \tag{9}$$

where $M_T(u, v)$ denotes the class predicted by SegNeXt-T for pixel (u, v), and $M_T(u, v) = 1$ indicates that the pixel belongs to the femoral head region.

The pixel area of the femoral head is calculated as:

$$A_{fh} = \left| R_{fh} \right|. \tag{10}$$

The center pixel coordinates $p_c(u_c, v_c)$ of the femoral head is then calculated by averaging the coordinates of all pixels within the segmented region:

$$u_c = \frac{1}{A_{fh}} \sum_{(u,v) \in R_{fh}} u, \quad v_c = \frac{1}{A_{fh}} \sum_{(u,v) \in R_{fh}} v. \tag{11}$$

The pixel area is used to determine whether the femoral head is visible, while the center position provides visual feedback for probe adjustment, as shown in Figure 4. Since rapid image feedback is required in this stage, SegNeXt-T is selected due to its smaller model size and faster inference speed.

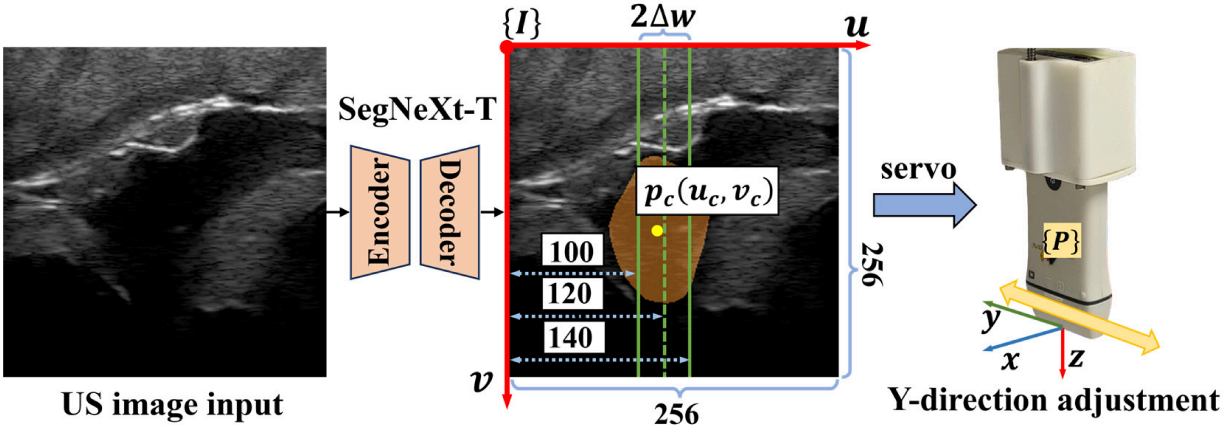


Figure 4. Femoral head feature extraction from ultrasound images for servo control.

The second network is SegNeXt-L, which is used for multi-structure segmentation of candidate Graf standard plane images. After femoral head localization, the robot performs local scanning around the selected position and acquires a series of candidate ultrasound images. These images are then input into SegNeXt-L for semantic segmentation, as shown in Figure 5. The foreground classes include the chondro-osseous junction, femoral, ilium, labrum, and synovial fold. Compared with SegNeXt-T, SegNeXt-L has a larger encoder and stronger feature representation ability, and is therefore used for fine segmentation of multiple anatomical structures in the Graf standard plane.

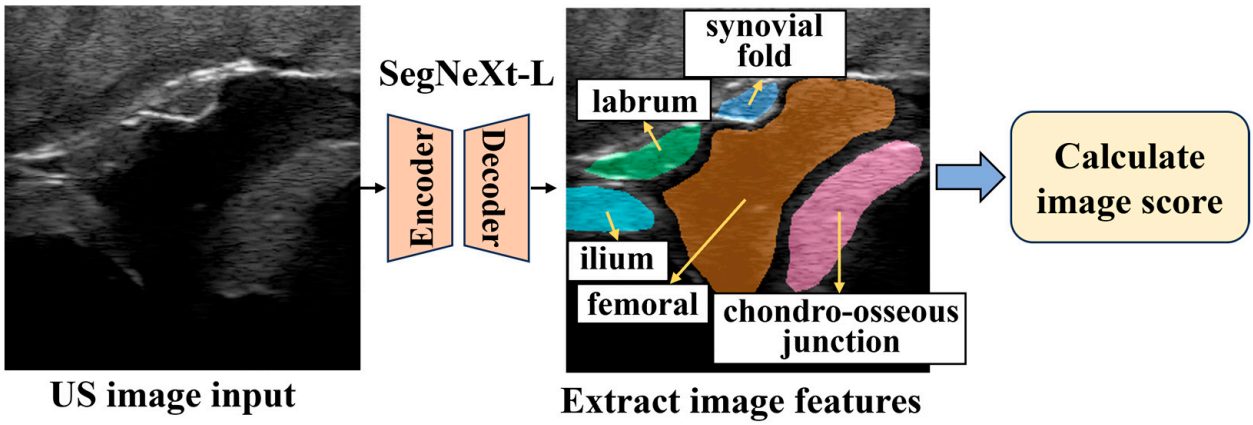


Figure 5. Feature extraction from ultrasound images for identifying high-quality Graf standard planes.

For the j-th foreground anatomical structure, its segmented region is defined as:

$$R_j = \{(u, v) \mid M_L(u, v) = j\}, j \in \{1, 2, 3, 4, 5\}, \tag{12}$$

where $M_L(u, v)$ denotes the predicted class of pixel (u, v). The pixel area of the corresponding structure is calculated from R_j and defined as:

$$A_j = |R_j|. \tag{13}$$

In this study, the pixel areas of the foreground structures are used as the main features for Graf standard plane scoring, reflecting the visibility and completeness of key anatomical structures in the image. These area features are then input into the scoring mechanism to select the image with the highest score as the optimal Graf standard plane.

In summary, SegNeXt-T and SegNeXt-L are used to construct a two-stage ultrasound image feature extraction method. SegNeXt-T focuses on rapid femoral head localization during the searching stage, whereas SegNeXt-L focuses on accurate anatomical structure extraction from candidate standard plane images. This design balances real-time performance in the searching stage and segmentation accuracy in the standard plane selection stage, providing the basis for subsequent image-guided servo control and standard plane scoring.

2.5. Dataset Construction and Training

Based on the hip phantom, two ultrasound image datasets were constructed to train the two SegNeXt-based segmentation networks described in Section 2.4. The first dataset was used for femoral head segmentation, in which the femoral head region was manually annotated at the pixel level. This dataset contained 390 ultrasound images and was divided into training and validation sets at a ratio of 8:2, with 312 images for training and 78 images for validation. The second dataset was used for multi-structure segmentation. Five foreground anatomical structures were manually annotated, including the chondro-osseous junction, femoral, ilium, labrum, and synovial fold. This dataset contained 576 ultrasound images and was also divided at a ratio of 8:2, with 461 images for training and 115 images for validation. All input images were resized to 256 × 256 pixels.

The training settings of the two datasets were kept consistent, except for the number of output classes. The models were implemented using PyTorch 2.1.0 and trained on an NVIDIA GeForce RTX 3080 GPU. During training, random scaling, random cropping, random flipping, and Gaussian noise were used for data augmentation to improve the robustness of the models to probe pose changes, noise artifacts, and local structural variations. The batch size was set to 4, and the maximum number of training

iterations was set to 10,000. Model performance was evaluated on the validation set every 1000 iterations. In each iteration, the model was updated using one batch of data. The model with the best validation performance was saved for subsequent robotic standard plane searching experiments.

The loss function was defined as a combination of cross-entropy loss and Dice loss:

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{Dice}. \tag{14}$$

Cross-entropy loss was used to improve pixel-level classification accuracy, while Dice loss was used to increase the overlap between the predicted regions and manual annotations. This combination helps reduce the influence of small foreground structures and class imbalance in ultrasound images.

The cross-entropy loss is defined as:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} log(p_{i,c}), \tag{15}$$

where N denotes the total number of pixels, C denotes the number of classes, $y_{i,c}$ denotes the ground-truth label of the i-th pixel for class c, and $p_{i,c}$ denotes the predicted probability that the i-th pixel belongs to class c.

The Dice loss is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{i=1}^{N} p_{i,c}y_{i,c} + \epsilon}{\sum_{i=1}^{N} p_{i,c}^2 + \sum_{i=1}^{N} y_{i,c}^2 + \epsilon}, \tag{16}$$

where ϵ is a smoothing term used to avoid division by zero.

AdamW was used as the optimizer, with an initial learning rate of 6×10^{-5} and a weight decay of 0.01. A learning rate scheduling strategy combining linear warm-up and polynomial decay was adopted. In the early training stage, LinearLR gradually increased the learning rate to the initial value to improve training stability. After warm-up, PolyLR gradually decayed the learning rate according to the number of iterations, as follows:

$$lr_t = lr_0\left(1\frac{t}{T}\right)^p, \tag{17}$$

where lr_t is the learning rate at the t-th iteration, lr_0 is the initial learning rate, T is the maximum number of iterations, and p is the polynomial decay coefficient. In this study, p was set to 1.0, so the learning rate approximately decayed linearly to 0 during training.

2.6. Graf Standard Plane Search Strategy

To automatically acquire the Graf standard plane of the infant hip, a three-stage search strategy is designed, as shown in Figure 6. This strategy mimics the coarse-to-fine scanning process used by sonographers, progressively narrowing the search region based on anatomical features. Since the Graf standard plane is visible only within a limited range of probe poses, exhaustive scanning is inefficient. The strategy first uses the RGB-D camera for initial surface positioning, then searches for the femoral head center based on segmentation results, and finally performs multi-angle scanning around the femoral head center. The image with the highest score given by the scoring mechanism is selected as the optimal Graf standard plane. Throughout the contact process with the phantom, the robot ensures scanning safety and stable contact force using the compliant control method described in Section 2.3.

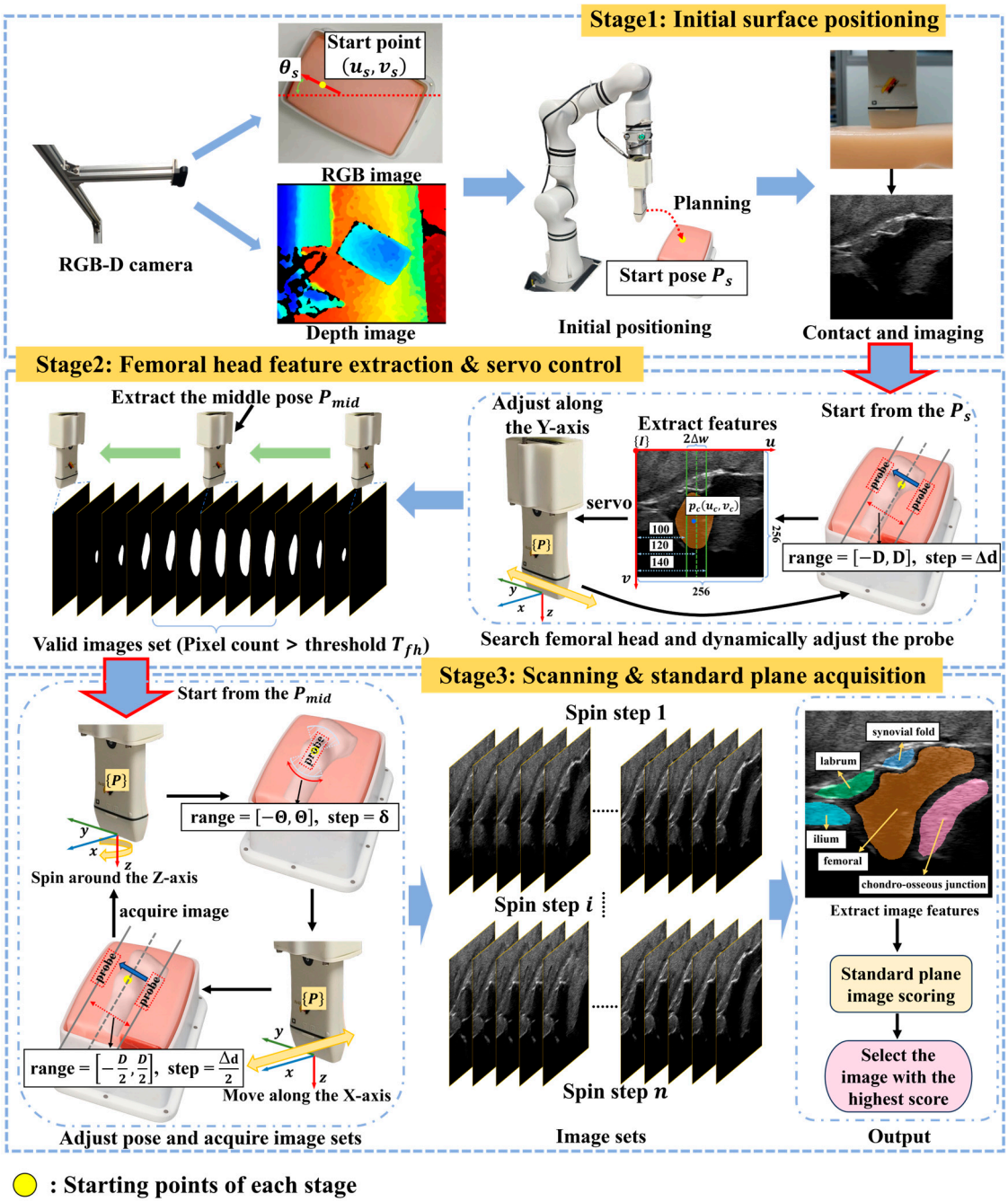


Figure 6. Complete workflow for robotic acquisition of standard plane ultrasound images. The workflow consists of three stages. In the first stage, the operator manually selects an initial positioning point on the acquired RGB image. The system then combines the depth image to calculate the real-world coordinate P_s and controls the robotic arm to contact the phantom for imaging. In the second stage, the probe plans an x-direction scanning trajectory to search for the femoral head. During this process, the femoral head features extracted by segmentation guide servo adjustment of the probe along the y-direction. After scanning, the middle pose P_{mid} among the poses where the femoral head is present is selected as the starting point for the next stage. In the third stage, the probe starts from P_{mid} and performs local multi-angle scanning. Image sequences are acquired by rotation around the z-axis and translation along the x-direction, and the optimal ultrasound image is finally selected according to the standard plane scoring result.

2.6.1. Initial Surface Positioning

The first stage is initial surface positioning. The RGB-D camera first acquires RGB and depth images of the phantom surface. Then, the operator selects an initial scanning

point near the Graf standard plane region in the RGB image and gives the initial probe direction. This step does not require precise localization or professional skills. Let the manually selected pixel point be (u_s, v_s). Its 3D coordinate P_s^C in the camera frame $\{C\}$ can be obtained from the depth image. According to the calibration relationship in Section 2.2.2, this point is transformed into the robot base frame $\{B\}$:

$$P_s^B = {}^B\boldsymbol{T}_C P_s^C. \tag{18}$$

The initial probe pose is then determined according to the selected scanning direction, and the initial end-effector pose P_s is generated. The robot moves to this pose, as shown in Stage 1 of Figure 6. After the probe contacts the body surface and the contact force becomes stable, the system enters the second stage.

2.6.2. Femoral Head Searching

The second stage aims to search for the femoral head region near the initial positioning point and determine its center along the scanning direction. Taking the contact pose obtained in the first stage as the center, the robot plans a scanning trajectory along the x-direction. The scanning range is set to $[-D, D]$, with a step size of Δd. At each sampling position, the system acquires one ultrasound image. After segmentation, the pixel area A_{fh} and center position $p_c(u_c, v_c)$ of the femoral head region are calculated. When $A_{fh} > T_{fh}$, the current image is considered to contain a valid femoral head region. At this stage, servo adjustments are performed during scanning according to the horizontal coordinate deviation between p_c and the desired position. Specifically, the desired horizontal pixel range is set to $[u_{\min}, u_{\max}] = [100, 120]$. When u_c is outside this range, the horizontal pixel deviation is calculated as $\Delta u = u_{mid} - u_c$, and is then converted into the actual physical displacement Δy. This displacement is used to guide the servo adjustment of the probe along the y-direction. Valid images and their corresponding end-effector poses are continuously recorded as long as the femoral head remains visible. When the femoral head region disappears or its pixel area falls below the threshold, the scan is terminated, as shown in Stage 2 of Figure 6. Finally, the middle pose P_{mid} is selected from the continuous valid pose sequence as the starting point for the third stage.

2.6.3. Multi-Angle Scanning and Image Screening

The third stage performs multi-angle scanning around the selected starting point P_{mid} and screens the image closest to the Graf standard plane, as shown in Stage 3 of Figure 6. The robot first rotates the probe around its local z-axis within a preset angular range $[-\Theta, \Theta]$, with an angular step of δ. At each rotation angle, the probe performs a small-range translation along its x-axis. At each sampling position, one ultrasound image is acquired. After all rotation and translation scans are completed, a candidate image set is obtained. All candidate images are then input into the SegNeXt-L network described in Section 2.4 to segment five anatomical structures, including the chondro-osseous junction, femoral, ilium, labrum, and synovial fold. The pixel area features of each structure are then extracted.

To evaluate how close each candidate image is to the Graf standard plane, the standard plane scoring mechanism is constructed based on the presence and completeness of each structure.

First, connected-component analysis is performed on the segmentation results of the five structures in each image. For the j-th structure in the i-th image, if the segmentation

result contains only one valid connected region and its pixel area is larger than the corresponding threshold T_j, the structure is considered validly present in the current image:

$$E_{i,j} = \begin{cases} 1, & C_{i,j} = 1 \text{ and } A_{i,j} > T_j \\ 0, & \text{otherwise} \end{cases}, \tag{19}$$

where $C_{i,j}$ denotes the number of valid connected regions of the j-th structure in the i-th image, $A_{i,j}$ denotes its pixel area, and T_j is the minimum area threshold of the corresponding structure.

Based on the presence of the five structures, the total presence score is defined as:

$$S_i^{exist} = \frac{\sum_{j=1}^{5} w_j E_{i,j}}{\sum_{j=1}^{5} w_j}, \tag{20}$$

where w_j is the scoring weight of the j-th structure.

Next, to evaluate the display completeness of each foreground structure in the image, a maximum reference pixel area A_j^{max} is set for each structure. The area completeness ratio of the j-th structure in the i-th image is defined as:

$$Q_{i,j} = \min\left(\frac{A_{i,j}}{A_j^{max}} 1\right). \tag{21}$$

Based on the area completeness ratios of the five structures, the total completeness score is defined as:

$$S_i^{area} = \frac{\sum_{j=1}^{5} w_j Q_{i,j}}{\sum_{j=1}^{5} w_j}. \tag{22}$$

Finally, the total score of the i-th candidate image is defined as:

$$S_i = 0.7 S_i^{exist} + 0.3 S_i^{area}. \tag{23}$$

Thus, the closer the score is to 1, the more consistent the image is with the Graf standard plane. The ultrasound image with the highest score is finally selected as the output, completing the entire search process.

3. Results

3.1. Segmentation Results and Evaluation

To evaluate the performance of the segmentation models, the commonly used Intersection over Union (IoU) and Dice coefficient were selected as evaluation metrics. For the c-th class, let P_c denote the predicted mask and G_c denote the corresponding ground-truth mask. The two metrics are defined as:

$$IoU = \frac{|P_c \cap G_c|}{|P_c \cup G_c|}, \tag{24}$$

$$Dice = \frac{2|P_c \cap G_c|}{|P_c| + |G_c|}. \tag{25}$$

Both networks were trained for 10,000 iterations, with model performance evaluated on the validation set every 1000 iterations. For both networks, the best-performing model was selected based on the mean Intersection over Union (mIoU), which represents the average IoU over all target classes.

Table 1 presents the quantitative results of the two segmentation models on the validation sets. Both models achieved good segmentation performance in their corresponding tasks. Although SegNeXt-T has a smaller model size, it still achieved high IoU and Dice scores in the binary femoral head segmentation task. In the multi-structure segmentation task, SegNeXt-L showed good segmentation ability for the five foreground anatomical structures, achieving an mIoU of 0.766 and a mean Dice coefficient of 0.866. The metrics for the labrum and synovial fold were slightly lower, which may be related to their small sizes. However, the visualization results showed that the model could still extract their main regional features, and therefore this did not affect the subsequent study.

Table 1. Experimental results of the two segmentation models.

Model	Class	IoU	Dice
SegNeXt-T	Femoral head	0.773	0.872
SegNeXt-L	Chondro-osseous junction	0.780	0.876
	Femoral	0.842	0.914
	Ilium	0.792	0.884
	Labrum	0.719	0.836
	Synovial fold	0.698	0.822

To visually demonstrate the segmentation performance of the two models, Figure 7 shows representative validation results. Figure 7a presents the femoral head segmentation results of SegNeXt-T, including cases where the femoral head is centered or shifted. The model can accurately identify the femoral head region under different positions. Figure 7b presents the multi-structure segmentation results of SegNeXt-L under different conditions. When the image is close to the Graf standard plane, the model can identify the chondro-osseous junction, femoral, ilium, labrum, and synovial fold relatively completely. When some structures are missing or incompletely displayed, the model can still identify the structures actually present in the current image.

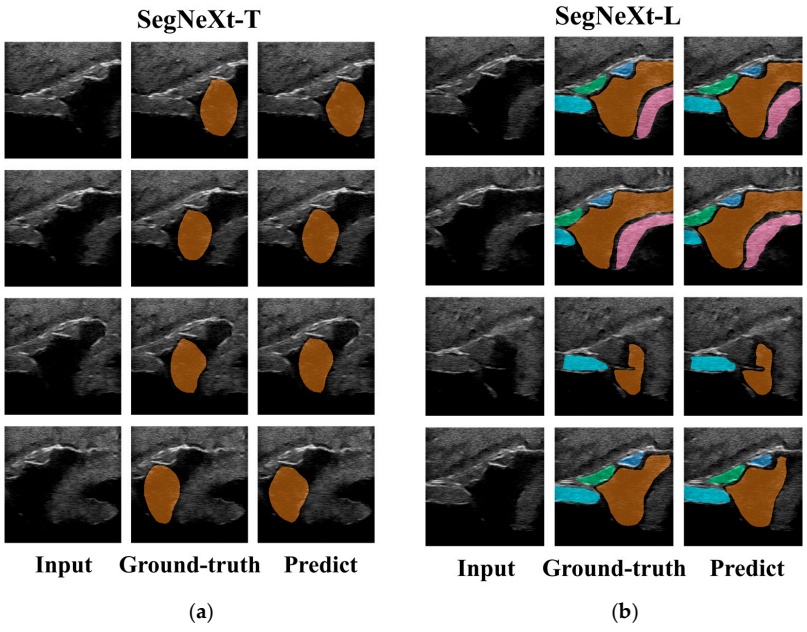


Figure 7. Visualization results of the two segmentation networks. (**a**) SegNeXt-T femoral head segmentation results. The first column shows the original ultrasound images, and the second and third

columns show the manual annotations and model predictions, respectively. (**b**) SegNeXt-L multi-structure segmentation results. The first column shows the original ultrasound images, and the second and third columns show the manual annotations and model predictions of five anatomical structures, including the chondro-osseous junction (pink), femoral (brown), ilium (cyan), labrum (green), and synovial fold (blue).

3.2. Graf Standard Plane Search Experiment

To validate the effectiveness and stability of the proposed Graf standard plane search strategy, experiments were conducted on the hip phantom. During the experiments, different conditions were simulated by changing the placement angle and orientation of the phantom, and a total of 30 trials were performed. In the initial body-surface positioning stage, the operator selected an initial scanning point near the standard plane region and an initial probe direction from the acquired RGB image. In the 30 trials, the distance error between the coarse positioning result and the target position was 1.12 ± 0.15 cm, with a maximum error of 1.47 cm and a 95% confidence interval of 1.06–1.18 cm. The angular error between the initial probe direction and the target direction was $7.36^\circ \pm 4.80^\circ$, with a maximum error of 23.10° and a 95% confidence interval of 5.57°–9.15°.

In the femoral head search stage, the robot planned an x-direction scanning trajectory centered on the initial contact pose and controlled the probe to scan along this trajectory. Based on multiple preliminary tests and actual scanning performance, the scanning range in the second stage was set to $[-3$ cm, 3 cm$]$, and the scanning step size Δd was set to 0.2 cm. According to the actual imaging results, the pixel area threshold T_{fh} for valid femoral head detection was set to 5000 pixels. When the femoral head area segmented by SegNeXt-T satisfied $A_{fh} > T_{fh}$, the system regarded the current image as containing a valid femoral head region. Servo adjustments were then performed along the y-direction according to the femoral head center position, moving the femoral head into the target pixel range [100, 140]. The femoral head search and servo adjustment process is shown in Figure 8. After initial positioning, the system was able to gradually search for the femoral head region along the scanning direction and complete position adjustment after detecting a valid femoral head region. The robot continued searching along the trajectory until no valid femoral head region was detected. In 30 trials, femoral head search succeeded in 28 cases, with a success rate of 93.3%.

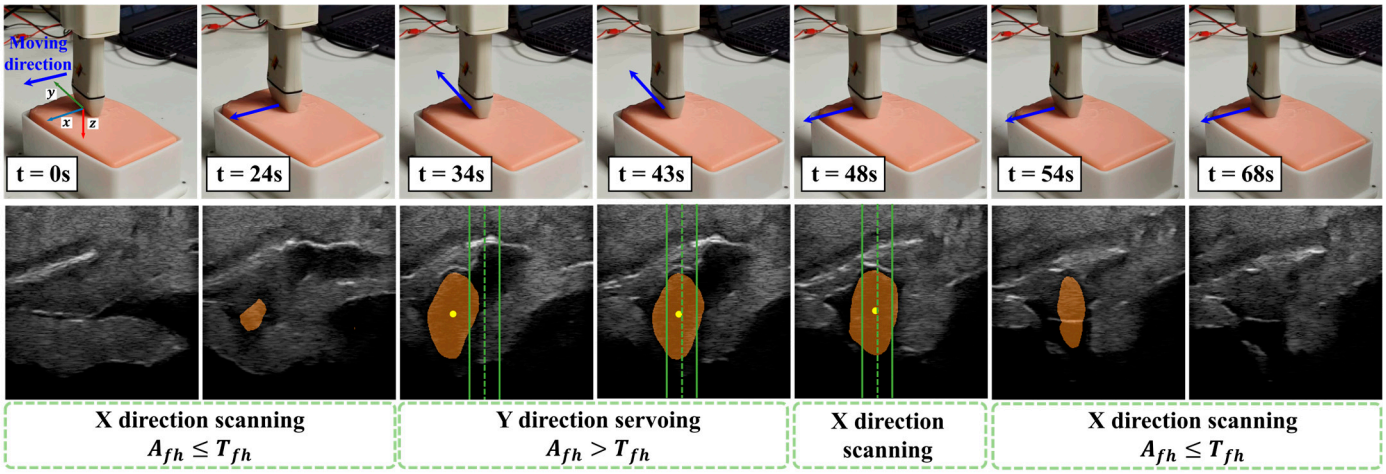


Figure 8. Femoral head search experiment. The first row shows the probe–phantom positions at different time points, and the second row shows the corresponding ultrasound images and femoral head segmentation results. Since the probe displacement was at the millimeter scale, the blue arrows indicate the main motion direction rather than the true displacement scale.

After femoral head search, the pose corresponding to the femoral head center was used as the scanning starting point, and the probe was rotated for multi-angle scanning. In this stage, the probe was first rotated around the z-axis within the range of $[-30^\circ, 30^\circ]$, with an angular step of 6°. This angular setting covered the possible region where the Graf standard plane may appear while balancing scanning accuracy and efficiency. At each rotation angle, the probe scanned over a small range along the x-axis, with a scanning range of [−1.5 cm, 1.5 cm] and a translation step of 0.1 cm. This allowed a more concentrated set of candidate ultrasound images to be acquired. All candidate images were then input into the SegNeXt-L network for segmentation of the five anatomical structures. The Graf standard plane score was calculated according to the presence and pixel-area completeness of each structure. The existence and completeness thresholds used in the scoring mechanism are listed in Table 2. The threshold parameters were determined through statistical analysis of the preliminary scanning data. The existence thresholds were calculated to distinguish reliable anatomical-structure segmentation from small false-positive regions, whereas the completeness thresholds were derived from the pixel-area distributions of clearly displayed structures. The weighting factors were selected based on their ability to distinguish candidate images of different quality levels. Within the range of imaging variations encountered in the phantom experiments, the segmentation model consistently extracted the relevant anatomical structures from images acquired at different probe poses. Therefore, the anatomy-based thresholds remained effective and demonstrated satisfactory robustness in the phantom experiments. Their generalizability to clinical images and substantially different imaging settings requires further validation.

Table 2. Key parameters in the scoring mechanism.

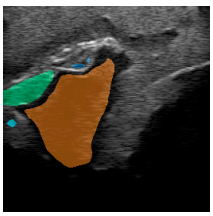
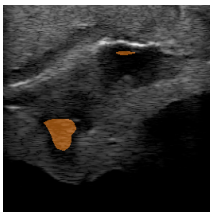
Landmark	Existence Threshold T_j (Pixel)	Completeness Threshold A_j^{max} (Pixel)
Chondro-osseous junction	2000	4700
Femoral	8000	13,500
Ilium	1000	2000
Labrum	800	1735
Synovial fold	400	860

To verify the ability of the scoring mechanism to distinguish Graf standard plane and nonstandard plane images, candidate images with different anatomical structure visibility were selected for scoring analysis, as shown in Table 3. When all five anatomical structures were visible and relatively complete, the score was usually between 80 and 100, indicating a Graf standard plane image. As anatomical structures became missing or incomplete, the score decreased accordingly. Images scoring 60–80 usually lacked one anatomical structure but were still close to the Graf standard plane. Images scoring 40–60 or 20–40 showed increasing deviation from the Graf standard plane, while images scoring below 20 contained almost no reliable anatomical structures. These results indicate that the scoring mechanism can distinguish candidate images of different quality and reflect their closeness to the Graf standard plane.

Based on the scoring rules, all candidate images were scored frame by frame, and the image with the highest score was selected as the final Graf standard plane output. Among the 28 trials with successful femoral head search, the system selected the Graf standard plane in 27 trials, giving a selection success rate of 96.4%. Considering all 30 repeated trials, the Graf standard plane was successfully obtained in 27 trials, with an overall success rate of 90.0%. These results indicate that the robot can autonomously complete the Graf

standard plane scanning task on the phantom and maintain stable performance under different initial conditions.

Table 3. Scoring results of images with different quality levels.

Score Range	Description	Example Image
80–100	All five anatomical structures exist and are largely complete. The image can be used as a standard-plane image.	
60–80	One anatomical structure is missing. The image is close to the Graf standard plane.	
40–60	Two anatomical structures are missing. The image shows an evident deviation from the Graf standard plane.	
20–40	Multiple anatomical structures are missing. The image shows a substantial deviation from the Graf standard plane.	
0–20	Almost no reliable anatomical structures exist. The image completely deviates from the Graf standard plane.	

3.3. Contact Force Control Performance Analysis

To evaluate the effectiveness of the proposed contact force compliance control method during ultrasound scanning, the variation in contact force during the search process was analyzed. During the experiment, the probe completed translational scanning and pose adjustment along the planned trajectory, while its position in the z-direction was corrected in real time according to force feedback. This allowed the probe to maintain stable contact during motion. To balance ultrasound image quality and patient safety, the desired contact force F_d was set to -2N, based on our previous probe pressure experiments and suggestions from sonographers [29].

Figure 9 shows the variation in the z-direction contact force during the whole scanning process. After the probe contacted the phantom, the actual contact force remained close to the desired value. During scanning, local contact conditions between the probe and

the phantom changed with probe motion, resulting in small fluctuations in the force curve. However, no obvious drift was observed. The contact force ranged from −2.492 to −1.517 N, with a mean value of −2.045 ± 0.213 N (mean ± standard deviation). The root mean square error relative to the target force was 0.229 N, indicating a small overall tracking error and stable probe–phantom contact. To further evaluate the motion stability of the robot during scanning, the Z-axis position of the end effector was recorded, as shown in Figure 10. At the beginning of the process, the end effector moved downward from the initial position to establish contact with the phantom. After stable contact was achieved, the Z-axis position remained within a narrow range and changed gradually throughout the subsequent scanning process.

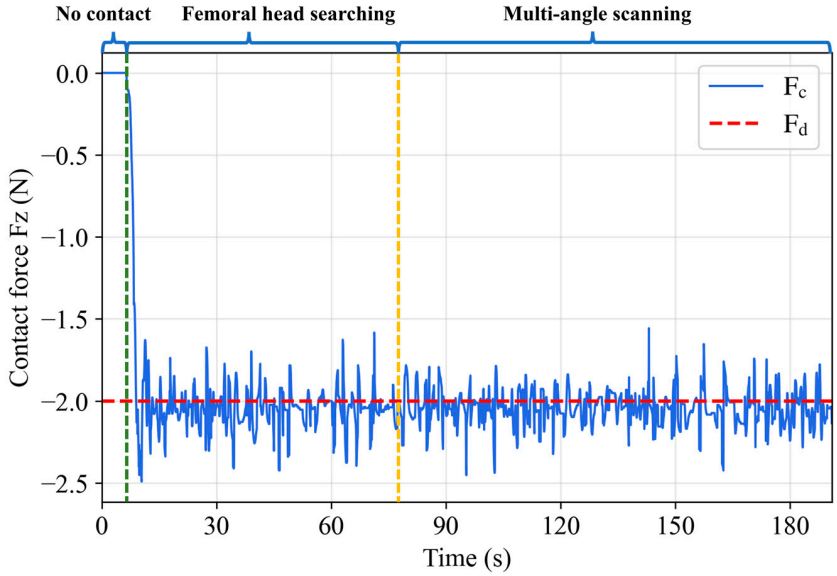


Figure 9. Variation in the *z*-direction contact force during scanning. The region before the green dashed line corresponds to the no contact stage, the region between the green and yellow dashed lines corresponds to the femoral head searching stage, and the region after the yellow dashed line corresponds to the multi-angle scanning stage. The blue solid line represents the actual contact force F_c, and the red dashed line represents the desired contact force F_d.

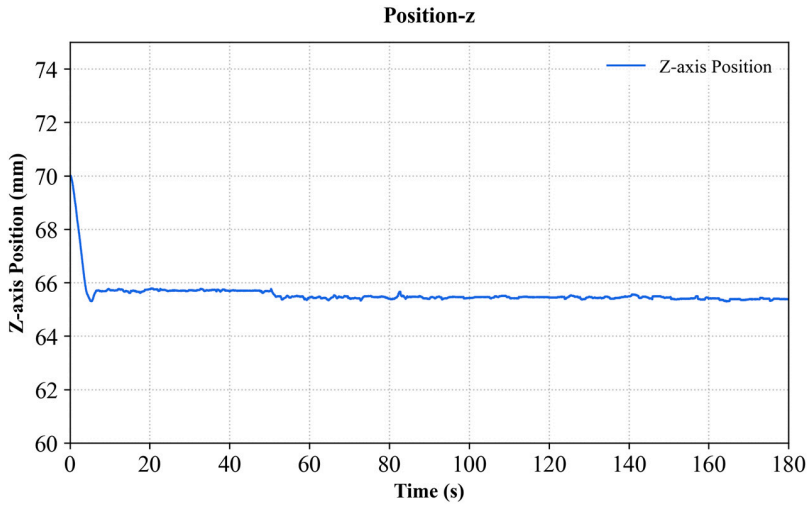


Figure 10. Variation in the Z-axis position of the robot end effector during the scanning process.

These results show that the proposed contact force compliance control method can maintain the contact force within the desired range during continuous probe motion, ensuring safe and stable image acquisition.

4. Discussion

The experimental results demonstrate the preliminary feasibility of the proposed autonomous robotic ultrasound system for Graf standard plane acquisition. By integrating stage-wise robotic searching, ultrasound image segmentation, multi-structure scoring, and compliant force control, the system achieved a Graf standard plane acquisition success rate of 90.0% while maintaining stable probe–phantom contact. These results indicate that the proposed framework can reduce dependence on operator experience and improve the consistency of standard plane acquisition.

Nevertheless, this study has several limitations. First, the experiments were mainly conducted on a stationary hip phantom. Differences between the phantom and real infants in anatomical structure, ultrasound appearance, and tissue mechanical properties may affect segmentation and force-control performance. Involuntary infant motion may also disturb probe–skin contact and cause the target anatomy to move outside the predefined search region. Second, the initial scanning point and probe direction were still manually specified, meaning that the current system has not yet achieved complete autonomy. Third, the segmentation datasets and scoring thresholds were established primarily from phantom data, and their generalizability to different subjects and clinical conditions remains to be evaluated. Finally, the present system focuses on Graf standard plane acquisition and does not yet include automatic angle measurement or DDH classification.

Future work will focus on improving system autonomy, robustness, and clinical applicability. Automated initial localization and motion compensation methods will be investigated to reduce manual intervention and improve performance under dynamic conditions. Clinical ultrasound datasets will be expanded to improve model generalization, followed by human-subject studies comparing robotic and sonographer-acquired standard planes. Automatic measurement of the α and β angles will also be integrated to establish a complete workflow from autonomous standard plane acquisition to computer-assisted DDH assessment.

5. Conclusions

This study developed an autonomous robotic ultrasound system to address the dependence on experienced operators and the limited consistency and repeatability in hip ultrasound examination. The system integrates RGB-D visual localization, contact force control, and ultrasound image segmentation. Based on system calibration, a hybrid force/position control method based on admittance control was used to maintain stable contact while the probe moved along the planned trajectory. For Graf standard plane acquisition, a three-stage search strategy was designed, including initial body surface positioning, femoral head search, and local multi angle scanning. The system uses SegNeXt-T to extract femoral head features and SegNeXt-L to segment five anatomical structures in candidate images. Combined with a multi structure scoring mechanism, the optimal Graf standard plane image can be automatically selected.

The proposed system was preliminarily validated on a customized hip phantom. The experimental results showed that SegNeXt-T achieved a Dice coefficient of 0.872 for femoral head segmentation, and SegNeXt-L achieved a mean Dice coefficient of 0.866 for the five anatomical structures. In 30 autonomous search trials under different initial conditions, the system successfully acquired the Graf standard plane in 27 trials, with an overall success rate of 90.0%. During scanning, the normal contact force ranged from -2.492 to -1.517 N, with a mean value of -2.045 N, a root mean square error of 0.229 N, and a standard deviation of 0.213 N. These results indicate that, after a coarse initial position and probe direction are manually provided, the robot can autonomously complete the subsequent search and maintain stable contact during scanning. This study provides a feasible approach for

reducing the dependence of hip ultrasound scanning on operator experience and improving the consistency and repeatability of Graf standard plane acquisition.

Author Contributions: J.C. managed the project. Y.D. designed the program and drafted the first version of the manuscript. Y.X. led the force control research. Y.D., X.Z., Y.X. and W.Z. designed the project and revised the manuscript. 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) and Yangzhou Municipal Natural Science Foundation (No. YZ2025159).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank Jiakuan Wang from Yangzhou Maternal and Child Health Care Hospital for his professional guidance during this study, especially for his important suggestions on infant hip ultrasound scanning procedures, Graf standard plane image recognition, and phantom design.

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
RGB-D	Red, Green, Blue, and Depth
UDP	User Datagram Protocol
CPU	Central Processing Unit
GPU	Graphics Processing Unit
AdamW	Adam with Decoupled Weight Decay
LinearLR	Linear Learning Rate
PolyLR	Polynomial Learning Rate
IoU	Intersection over Union

References

1. Heneghan, M. Developmental dysplasia of the hip. *JAAPA* **2021**, *34*, 48–49. [CrossRef] [PubMed]
2. Nicholson, A.; Dunne, K.; Taaffe, S.; Sheikh, Y.; Murphy, J. Developmental dysplasia of the hip in infants and children. *BMJ* **2023**, *383*, e074507. [CrossRef] [PubMed]
3. Riboni, G.; Bellini, A.; Serantoni, S.; Rognoni, E.; Bisanti, L. Ultrasound screening for developmental dysplasia of the hip. *Pediatr. Radiol.* **2003**, *33*, 475–481. [CrossRef] [PubMed]
4. Graf, R. *Hip Sonography: Diagnosis and Management of Infant Hip Dysplasia*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 40–421.
5. 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. [CrossRef] [PubMed]
6. Sioutis, S.; Kolovos, S.; Papakonstantinou, M.-E.; Altsitzioglou, P.; Polyzou, M.; Chlapoutakis, K.; Karampikas, V.; Gavriil, P.; Mitsiokapa, E.; Koulalis, D. Impact of probe tilt on Graf ultrasonography accuracy for neonatal hip dysplasia screening. *SICOT-J.* **2025**, *11*, 22. [CrossRef] [PubMed]
7. Priester, A.M.; Natarajan, S.; Culjat, M.O. Robotic ultrasound systems in medicine. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2013**, *60*, 507–523. [CrossRef] [PubMed]
8. Jiang, Z.; Salcudean, S.E.; Navab, N. Robotic ultrasound imaging: State-of-the-art and future perspectives. *Med. Image Anal.* **2023**, *89*, 102878. [CrossRef] [PubMed]
9. Hidalgo, E.M.; Wright, L.; Isaksson, M.; Lambert, G.; Marwick, T.H. Current applications of robot-assisted ultrasound examination. *Cardiovasc. Imaging* **2023**, *16*, 239–247. [CrossRef] [PubMed]

10. Black, D.; Oloumi Yazdi, Y.; Hadi Hosseinabadi, A.H.; Salcudean, S. Human teleoperation-a haptically enabled mixed reality system for teleultrasound. *Hum. –Comput. Interact.* **2024**, *39*, 529–552. [CrossRef]
11. Najdovski, Z.; Pedrammehr, S.; Chalak Qazani, M.R.; Abdi, H.; Deshpande, S.; Liu, T.; Mullins, J.; Fielding, M.; Hilton, S.; Asadi, H. Haptiscan: A haptically-enabled robotic ultrasound system for remote medical diagnostics. *Robotics* **2024**, *13*, 164. [CrossRef]
12. He, T.; Pu, Y.-Y.; Zhang, Y.-Q.; Qian, Z.-B.; Guo, L.-H.; Sun, L.-P.; Zhao, C.-K.; Xu, H.-X. 5G-based telerobotic ultrasound system improves access to breast examination in rural and remote areas: A prospective and two-scenario study. *Diagnostics* **2023**, *13*, 362. [CrossRef] [PubMed]
13. Ren, J.-Y.; Lei, Y.-M.; Lei, B.-S.; Peng, Y.-X.; Pan, X.-F.; Ye, H.-R.; Cui, X.-W. The feasibility and satisfaction study of 5G-based robotic teleultrasound diagnostic system in health check-ups. *Front. Public Health* **2023**, *11*, 1149964. [CrossRef] [PubMed]
14. Bao, X.; Wang, S.; Zheng, L.; Housden, R.J.; Hajnal, J.V.; Rhode, K. A novel ultrasound robot with force/torque measurement and control for safe and efficient scanning. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [CrossRef] [PubMed]
15. Laurent, E.; Soemantoro, R.; Jenner, K.; Kardos, A.; Tang, G.; Zhao, Y. Echo-Robot: Semi-autonomous cardiac ultrasound image acquisition using AI and robotics. *IEEE Trans. Med. Robot. Bionics* **2025**, *7*, 1307–1316. [CrossRef]
16. Zielke, J.; Eilers, C.; Busam, B.; Weber, W.; Navab, N.; Wendler, T. RSV: Robotic sonography for thyroid volumetry. *IEEE Robot. Autom. Lett.* **2022**, *7*, 3342–3348. [CrossRef]
17. Huang, Q.; Zhou, J.; Li, Z. Review of robot-assisted medical ultrasound imaging systems: Technology and clinical applications. *Neurocomputing* **2023**, *559*, 126790. [CrossRef]
18. Huang, Q.; Gao, B.; Wang, M. Robot-Assisted Autonomous Ultrasound Imaging for Carotid Artery. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–9. [CrossRef]
19. Wang, Z.; Han, Y.; Zhao, B.; Xie, H.; Yao, L.; Li, B.; Meng, M.Q.H.; Hu, Y. Autonomous Robotic System for Carotid Artery Ultrasound Scanning With Visual Servo Navigation. *IEEE Trans. Med. Robot. Bionics* **2024**, *6*, 1436–1447. [CrossRef]
20. Zakeri, E.; Spilkin, A.; Elmekki, H.; Zanuttini, A.; Kadem, L.; Bentahar, J.; Xie, W.-F.; Pibarot, P. Robust Deep Feature Ultrasound Image-Based Visual Servoing: Focus on Cardiac Examination. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 6407–6419. [CrossRef]
21. Su, K.; Liu, J.; Ren, X.; Huo, Y.; Du, G.; Zhao, W.; Wang, X.; Liang, B.; Li, D.; Liu, P.X. A fully autonomous robotic ultrasound system for thyroid scanning. *Nat. Commun.* **2024**, *15*, 4004. [CrossRef] [PubMed]
22. Su, K.; Du, G.; Wang, X.; Guan, Q. Tissue-View Map for Robotic Carotid Artery Ultrasound Scanning Using Reinforcement Learning. *IEEE Robot. Autom. Lett.* **2025**, *10*, 5178–5185. [CrossRef]
23. Sun, D.; Cappellari, A.; Lan, B.; Abayazid, M.; Stramigioli, S.; Niu, K. Automatic Robotic Ultrasound for 3D Musculoskeletal Reconstruction: A Comprehensive Framework. *Technologies* **2025**, *13*, 70. [CrossRef]
24. Roshan, M.C.; Isaksson, M.; Pranata, A.; Hidalgo, E.M. A sensor fusion approach to autonomous ultrasound imaging of the lumbar region. *Biomed. Signal Process. Control* **2025**, *108*, 106818. [CrossRef]
25. Lee, C.-Y.; Gao, Y.-C.; Ciou, W.-S.; Wu, M.-J.; Du, Y.-C. Combining RGB-D Sensing With Adaptive Force Control in a Robotic Ultrasound System for Automated Real-Time Fistula Stenosis Evaluation. *IEEE Sens. J.* **2025**, *25*, 5446–5456. [CrossRef]
26. Jiang, H.; Zhao, A.; Yang, Q.; Yan, X.; Wang, T.; Wang, Y.; Jia, N.; Wang, J.; Wu, G.; Yue, Y.; et al. Towards expert-level autonomous carotid ultrasonography with large-scale learning-based robotic system. *Nat. Commun.* **2025**, *16*, 7893. [CrossRef] [PubMed]
27. Li, R.; Niu, K.; Vander Poorten, E. A framework for fast automatic robot ultrasound calibration. In Proceedings of the 2021 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, 17–19 November 2021; pp. 1–7.
28. Tsai, R.Y.; Lenz, R.K. A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration. *IEEE Trans. Robot. Autom.* **1989**, *5*, 345–358. [CrossRef]
29. Cui, J.; Zhang, X.; Dai, Y.; Zhang, W. Research on Robotic Force Control for Infant Hip Ultrasound. *Actuators* **2026**, *15*, 333. [CrossRef]
30. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. Segnext: Rethinking convolutional attention design for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1140–1156. [CrossRef]
31. Pulik, Ł.; Czech, P.; Kaliszewska, J.; Mulewicz, B.; Pykosz, M.; Wiszniewska, J.; Łęgosz, P. Artificial Intelligence Algorithm Supporting the Diagnosis of Developmental Dysplasia of the Hip: Automated Ultrasound Image Segmentation. *J. Clin. Med.* **2025**, *14*, 6332. [CrossRef] [PubMed]

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.