technologies

MDPI

Article

Automatic Robotic Ultrasound for 3D Musculoskeletal Reconstruction: A Comprehensive Framework

Dezhi Sun *,†, Alessandro Cappellari †, Bangyu Lan, Momen Abayazid, Stefano Stramigioli and Kenan Niu

Robotics and Mechatronics Group (RAM), University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands; alecappellari00@gmail.com (A.C.); b.lan@utwente.nl (B.L.); m.abayazid@utwente.nl (M.A.); s.stramigioli@utwente.nl (S.S.); k.niu@utwente.nl (K.N.)

* Correspondence: dezhi.sun@utwente.nl

† These authors contributed equally to this work.

Abstract: Musculoskeletal ultrasound (US) imaging faces challenges such as operator experience, limited spatial flexibility, and high personnel costs. This study introduces an Automated Robotic Ultrasound Scanning (ARUS) system that integrates key technological advancements to automate the ultrasound scanning procedure with the robot, including anatomical target localization, automatic trajectory generation, deep-learning-based segmentation, and 3D reconstruction of musculoskeletal structures. The ARUS system consists of a robotic arm, ultrasound imaging, and stereo vision for precise anatomical area detection. A Graphical User Interface (GUI) facilitates a flexible selection of scanning trajectories, improving user interaction and enabling customized US scans. To handle complex and dynamic curvatures on the skin, together with anatomical area detection, the system employs a hybrid position–force control strategy based on the generated trajectory, ensuring stability and accuracy. Additionally, the utilized RA-UNet model offers multi-label segmentation on the bone and muscle tissues simultaneously, which incorporates residual blocks and attention mechanisms to enhance segmentation accuracy and robustness. A custom musculoskeletal phantom was used for validation. Compared to the reference 3D reconstruction result derived from the MRI scan, ARUS achieved a 3D reconstruction root mean square error (RMSE) of 1.22 mm, with a mean error of 0.94 mm and a standard deviation of 0.77 mm. The ARUS system extends 3D musculoskeletal imaging capacity by enabling both bones and muscles to be segmented and reconstructed into 3D shapes in real time and simultaneously. These features suggest significant potential as a cost-effective and reliable option for musculoskeletal examination and diagnosis in real-time applications.

Keywords: robotic imaging; 3D reconstruction; force control; musculoskeletal reconstruction; RA-UNet

check for updates

Academic Editor: Dennis Douroumis

Received: 17 December 2024
Revised: 14 January 2025
Accepted: 7 February 2025
Published: 8 February 2025

Citation: 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. https://doi.org/10.3390/technologies13020070

Copyright: © 2025 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Ultrasound imaging (US) is a widely used noninvasive diagnostic modality that enables medical professionals to visualize and assess soft tissues within the body [1]. Compared to other imaging modalities, such as Magnetic resonance imaging (MRI) or CT, US imaging is relatively cost-effective, making it a preferred option in clinical practice, particularly for routine examinations. Recent advances in ultrasound technology have facilitated the real-time generation of three-dimensional models, thereby improving its applicability in medical diagnostics [2,3]. In sports science and orthopedics, musculoskeletal imaging is imperative to examine bone and muscular conditions. Accurate muscle diagnostics are typically performed using MRI. It is highly effective in identifying muscle injuries,

tears, and other soft tissue abnormalities, making it the gold standard for diagnosing many musculoskeletal conditions [4]. However, this technology is expensive. In contrast, US is significantly less expensive, typically costing only 10% to 20% of an MRI scan, and provides real-time imaging [5]. However, it is dependent on the operator's expertise and is limited by its 2D image output, which lacks spatial context for complex anatomical structures.

The integration of automated ultrasound scanning using robots and three-dimensional reconstruction technology effectively addresses the limitations of traditional methods. Firstly, the use of a high-precision robotic manipulator eliminates the reliance on expert operators, enhancing standardization and consistency in operation. Secondly, the proposed technology leverages advanced segmentation algorithms to achieve high-accuracy simultaneous 3D reconstruction of regions of interest, such as muscles and bones. The resulting 3D models can be further refined and provide intuitive and reliable support for clinicians in assessing musculoskeletal structures and conducting pathological analyses.

Researchers have extensively explored 3D image reconstruction from US imaging, particularly focusing on semi-automated robotic scanning systems. For example, Li et al. [6], along with other studies such as [7,8], investigated 3D spinal reconstruction using semi-automated robotic scans. In addition to focusing solely on bone reconstruction, these studies implemented assisted robot trajectory planning by manually defining the scanning paths. Most semi- or fully automated robotic US scanning systems either follow predefined trajectories without external sensing or awareness [8,9] or employ external sensors such as calibrated optical tracking systems to provide environmental awareness during scanning [10–12]. Limited studies have explored the use of camera sensors mounted directly on the robot's end effector (EE) for visual guidance. This approach is more cost-effective and requires minimal calibration compared to external optical tracking systems [13,14].

Recent advancements in 3D ultrasound imaging have leveraged Neural Radiance Fields (NeRFs) to enhance the reconstruction of bony structures, effectively suppressing soft tissue interference for improved spinal curvature estimation [15]. Similarly, deep learning approaches, such as the Deep Contextual Learning Network (DCL-Net), have been developed to reconstruct 3D ultrasound volumes without the need for tracking devices, focusing primarily on skeletal features [16]. However, these methods often overlook the importance of soft tissue components, such as muscles, which are crucial for clinical applications.

Previous studies on 3D reconstruction in ultrasound imaging have primarily focused on skeletal structures, with limited attention to muscles [17–19]. Moreover, there has been a lack of approaches capable of simultaneously imaging and segmenting both skeletal and muscular structures. Unlike rigid bones, muscles undergo shape changes during motion, presenting unique challenges in developing automated scanning systems that can adapt to these dynamic deformations [20].

This study addresses key challenges in robotic ultrasound imaging by proposing a comprehensive system that integrates advanced segmentation, control, and trajectory planning strategies. Specifically, an enhanced RA-UNet segmentation model, incorporating attention mechanisms and residual connections, is utilized to achieve simultaneous and robust segmentation of bones and muscles, even in noisy and dynamic environments. To ensure stable imaging and effective interaction with deformable tissues, a hybrid position–force control strategy dynamically regulates the probe's contact force. Additionally, an interactive trajectory generation method, combined with real-time 3D surface reconstruction of anatomical regions, enables the system to autonomously compute optimal scanning paths based on operator-defined regions of interest. These contributions collectively address critical gaps in the literature, offering an integrated solution for accurate and efficient imaging, segmentation, and reconstruction of musculoskeletal structures.

In this paper, we proposed an Automated Robotic Ultrasound Scanning (ARUS) system capable of simultaneous imaging and segmentation of both skeletal and muscular structures. By integrating real-time robotic control with advanced segmentation algorithms, the ARUS system overcomes the limitations of traditional methods and existing semi-automated solutions. This research aims to demonstrate the feasibility and advantages of this approach through experimental validation, setting the stage for broader applications in musculoskeletal diagnostics and therapeutic interventions.

2. Materials and Methods

2.1. Experimental Setup

The proposed ARUS system consists of three main building blocks: a robotic arm (Franka Emika Panda, Munich, Germany), a US imaging system (Telemed MicrUs EXT-1H, Vilnius, Lithuania), and a stereo camera (Intel Realsense d435i, Santa Clara, CA, USA). A custom 3D-printed holder was mounted onto the end effector of the robot to attach the US transducer and the stereo camera. The ultrasound probe is mounted at a 90° angle relative to the stereo camera platform. During the scanning process, a transformation adjustment is applied to align the US probe with the scanning trajectory. The US system was set to have a clear image view of the targets both during image calibration and testing.

The robotic arm includes a built-in six-degree-of-freedom (6-DoF) force/torque sensor integrated into the end effector. The sensor can measure forces along the X, Y, and Z axes, as well as torques around these axes, with a maximum force range of ±65 N and a torque range of ±2.5 Nm. The sensor resolution is 0.05 N for force and 0.02 Nm for torque, providing accurate feedback during the scanning process.

A Windows 10 PC workstation (Intel Xeon, CPU @4.00 GHz, 32 GB RAM) was used to stream the US frames via UDP to the main workstation, Ubuntu 20.04 PC (Intel i7-7700, CPU @3.60 GHz, 16 GB RAM) used for data processing and robot control. The Windows 10 machine acted as an ultrasound data acquisition device, while the Ubuntu system handled the subsequent processing tasks, including robot control, ultrasound image segmentation, 3D reconstruction, etc. The two PCs were connected using a direct Ethernet connection, and the data stream was transmitted via UDP at a rate of 2 Hz during the scan. The main components of the system are illustrated in Figure 1. The experiments were conducted on a high-performance server with the specifications listed in Table 1. The configuration includes Ubuntu 20.04.4 LTS as the operating system, an Intel Xeon Platinum 8352V processor, 251 GiB of DDR4 RAM, and 8 NVIDIA A16 GPUs, each with 15,356 MiB of memory.

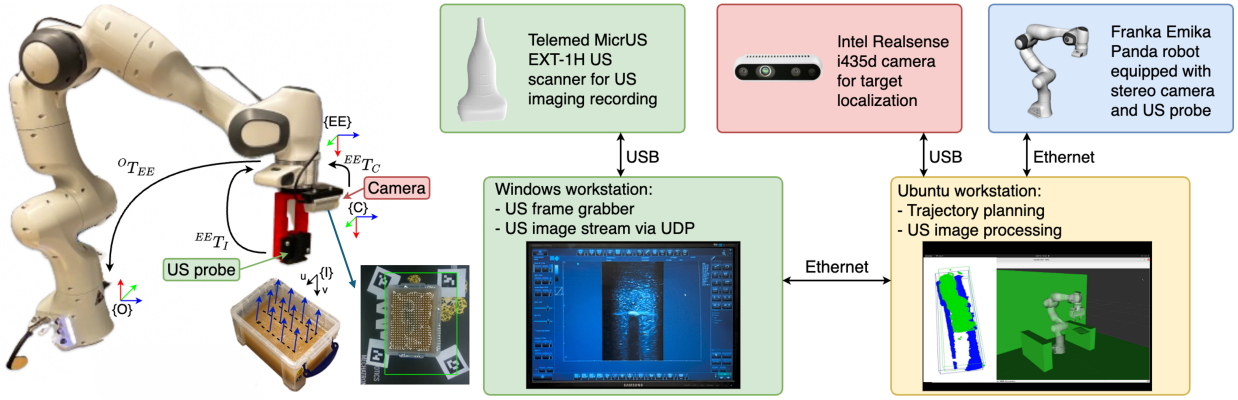


Figure 1. System overview of ARUS. On the left, the Franka Emika Panda robot equipped with the Telemed MicrUS US scanner and the Intel Realsense stereo camera, with all relevant frames and transformation matrices are highlighted. On the right are the main components and the relative connections.

Table 1. Server Configuration and Environment.

OS	Ubuntu 20.04.4 LTS
CPU	Intel Xeon Platinum 8352V @ 2.10 GHz
GPU	8 × NVIDIA A16 (15,356 MiB per GPU)
RAM	251 GiB DDR4

Table Note: OS = Operating System; GPU = Graphics Processing Unit; CPU = Central Processing Unit; RAM = Random Access Memory.

2.2. ARUS System Calibrations

In ARUS, different devices have different coordinate systems. In order to have the conversions among these coordinates, calibration is needed. In this paper, there are two types of calibrations: (1) ultrasound image-to-robot end-effector calibration, termed robotic ultrasound image calibration; (2) stereo camera-to-robot end-effector calibration, termed hand-eye calibration.

2.2.1. Ultrasound Image Calibration

The purpose of ultrasound image calibration is to determine the transformation $\mathbf{T}_I^{EE}$, which maps the 2D ultrasound image (I) to the robot's end-effector (EE) frame, and the scale matrix $\mathbf{T}_s$, which converts pixel dimensions into meters. This calibration minimizes the error between the sphere center's position in the robot base frame and its projection from the ultrasound image frame, expressed as:

$$f = \min \sum_{i=1}^{n} \left| \mathbf{T}_{EE,i}^{O} \mathbf{T}_I^{EE} \mathbf{T}_s \mathbf{p}_i^I - \mathbf{p}^O \right|, \tag{1}$$

where:

- $\mathbf{T}_{EE,i}^{O}$: The end-effector pose in the base frame for the i-th image.
- $\mathbf{p}_i^I$: The sphere center position in the ultrasound image frame.
- $\mathbf{p}^O$: The sphere center position in the base frame, calculated using trilateration [6].

Using the estimated transformations, any pixel coordinate $\mathbf{p}^I = [u, v, 0, 1]^T$ in the ultrasound image can be mapped to a 3D point $\mathbf{p}^O$ in the robot base frame as:

$$\mathbf{p}^O = \mathbf{T}_{EE}^{O} \mathbf{T}_I^{EE} \mathbf{T}_s \mathbf{p}^I. \tag{2}$$

2.2.2. Stereo Camera-to-Robot Calibration

In stereo camera-to-robot calibration, the goal is to estimate the homogeneous transformation $\mathbf{T}_C^{EE}$, which represents the relationship between the stereo camera (C) and the robot end-effector (EE) coordinate frames. This is achieved by capturing n images of a calibration grid (e.g., a chessboard) and determining the camera's intrinsic parameters and the relative poses of the grid.

Simultaneously, the corresponding poses of the end effector in the robot base frame ($\mathbf{T}_{EE}^{O}$) are recorded. This allows estimation of the transformation $\mathbf{T}_{obj}^{C}$, which represents the relative pose of the calibration grid in the camera frame. The homogeneous transformation $\mathbf{T}_C^{EE}$ is then computed using the set of pairs $\{\mathbf{T}_{EE}^{O}, \mathbf{T}_{obj}^{C}\}_i$, where i denotes the i-th image. The ViSP library [14] is used to solve the $AX = XB$ problem, enabling computation of $\mathbf{T}_C^{EE}$.

After calibration, the position of any point in the depth camera frame can be transformed into the robot base frame using:

$$\mathbf{p}^O = \mathbf{T}_{EE}^{O} \mathbf{T}_C^{EE} \mathbf{p}^C, \tag{3}$$

where:

$$\mathbf{p}^O = [p_x^O, p_y^O, p_z^O, 1]^T, \quad \mathbf{p}^C = [p_x^C, p_y^C, p_z^C, 1]^T. \tag{4}$$

This process ensures the accurate transformation of points from the stereo camera frame to the robot base frame, enabling seamless integration of visual data into robotic operations.

2.3. Robot Scanning Trajectory

The robot trajectory defines the precise positions and orientations that the ultrasound probe must follow during US scanning. This process begins with an initial target region detection using the stereo camera, which captures both visual and depth information. The depth data generates a point cloud representing the field of view detected by the stereo camera as shown in Figure 2, while the Region of Interest (ROI) was defined by four Aruco markers, as shown in Figure 1. These markers restrict the target region and filter out noise and reflections from the background, ensuring accurate localization of the scanning target.

To optimize the point cloud for robust trajectory generation, several preprocessing steps are applied:

1. A RANSAC [21] algorithm is employed to approximate the target surface, segmenting the relevant points and reducing noise. This approach assumes a locally planar structure while accommodating minor curvatures.
2. Visual information from the depth camera is used to identify and exclude points that are inconsistent with the expected surface geometry. Points detected outside the defined Region of Interest (ROI) or those not aligned with the estimated plane are treated as outliers and removed to improve the accuracy of the point cloud.
3. The voxel grid downsampling method was applied during the post-processing stage to optimize the point cloud for efficient trajectory planning. Unlike fixed-grid acquisition methods, which may reduce scanning flexibility and accuracy, voxel grid downsampling is applied after data acquisition, ensuring that critical geometric features are preserved while reducing the number of points. This approach achieves a balance between computational efficiency and accurate representation of anatomical structures.
4. Statistical analysis filters, specifically using the Local Outlier Factor (LOF) [22] method, were applied to detect and remove points that deviate significantly from their neighbors. This process ensures a consistent and reliable dataset by eliminating isolated noise points while preserving meaningful anatomical structures in the point cloud.

The refined point cloud is subsequently transformed into the robot's base frame for trajectory planning, enabling precise spatial alignment with the scanning environment.

The trajectory generation process involves three main steps to ensure anatomically conforming scanning paths. First, specific points are manually selected on the reconstructed surface mesh via a custom Graphical User Interface (GUI). These selected points serve as key reference locations along the desired scanning path. Next, the X and Y coordinates of the trajectory points are generated through linear interpolation between consecutive user-selected points, with a fixed step distance of 0.5 mm, to ensure uniform spacing. Finally, the interpolated trajectory points are projected onto the surface mesh using a KD-tree [23] nearest-neighbor search. This projection step adjusts the z-axis positions to conform to the surface curvature and aligns the z-axis orientation with the surface normal at each point. By incorporating surface projection, the generated trajectory adapts to the natural anatomical curvature, resulting in an accurate and robust scanning path.

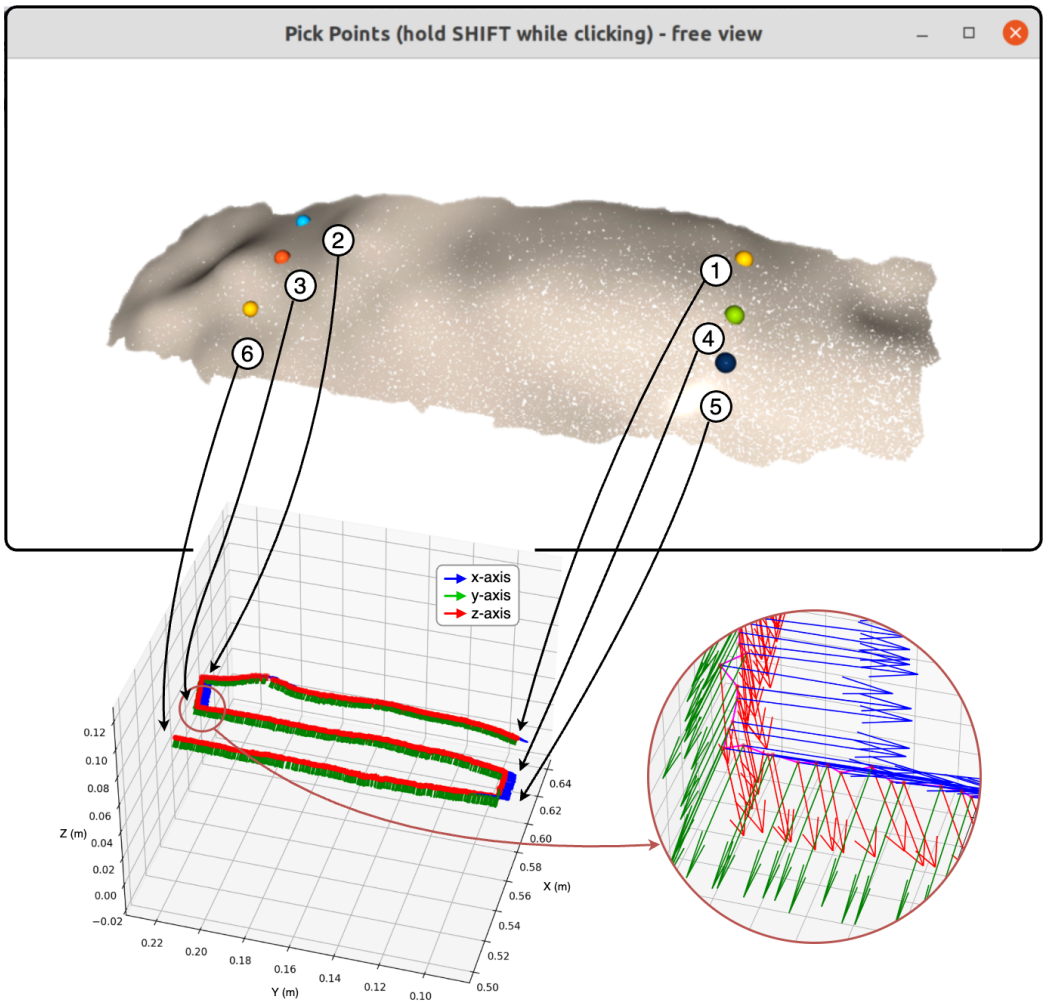


Figure 2. Mesh reconstruction generated from the point cloud. A Graphical User Interface (GUI) allows the selection of specific points (six points in this example) to define the scanning trajectory. Each trajectory point has an associated orientation: blue arrows represent the x-axis, green arrows the y-axis, and red arrows the z-axis.

Using the refined point cloud, a mesh reconstruction is created to represent the target geometry, as illustrated in Figure 2. A custom Graphical User Interface (GUI) allows the operator to select trajectory points interactively. The trajectory is represented as an $N \times 7$ matrix, where N is the number of poses, and each pose is defined as:

$$\mathbf{p}_i^{US} = [p_x^{US}, p_y^{US}, p_z^{US}, q_x^{US}, q_y^{US}, q_z^{US}, q_w^{US}]^T, \tag{5}$$

where $[p_x^{US}, p_y^{US}, p_z^{US}]$ represents the position, and $[q_x^{US}, q_y^{US}, q_z^{US}, q_w^{US}]$ the orientation in quaternion format. Each trajectory point is positioned 0.5 mm apart and oriented perpendicular to the surface, maintaining a constant yaw angle relative to the end effector (EE). As depicted in Figure 2, the z-axis (red arrow) points towards the target, while the x-axis (blue arrow) maintains a consistent direction throughout the trajectory.

During scanning, trajectory poses are dynamically updated to account for the deformable nature of skin and muscles. Adjustments are applied along the EE z-axis to ensure proper contact and alignment of the US probe with the target surface.

2.4. Hybrid Position–Force Control for Robotic Ultrasound Scanning

To achieve high-quality ultrasound (US) imaging requires a precise and robust control strategy. A hybrid position–force control scheme was implemented, as illustrated in

Figure 3. This strategy dynamically activates or deactivates the force control loop based on the US probe's contact status with the target, ensuring system stability and optimal imaging quality during scanning.

The force controller maintains a constant contact force of $f_{dz} = 3\,\mathrm{N}$ between the US probe and the target. This force is critical for consistent imaging quality. Forces are measured at the end effector (EE) and assumed to be equivalent to those on the US probe, as their frames are aligned. The external wrench is calculated using joint-level torque sensors and mapped to the EE via the Jacobian matrix. The force error is defined as:

$$\mathbf{f}_{ez} = \mathbf{f}_{dz} - \mathbf{f}_{mz}, \tag{6}$$

where $\mathbf{f}_{mz}$ is the measured force at the EE.

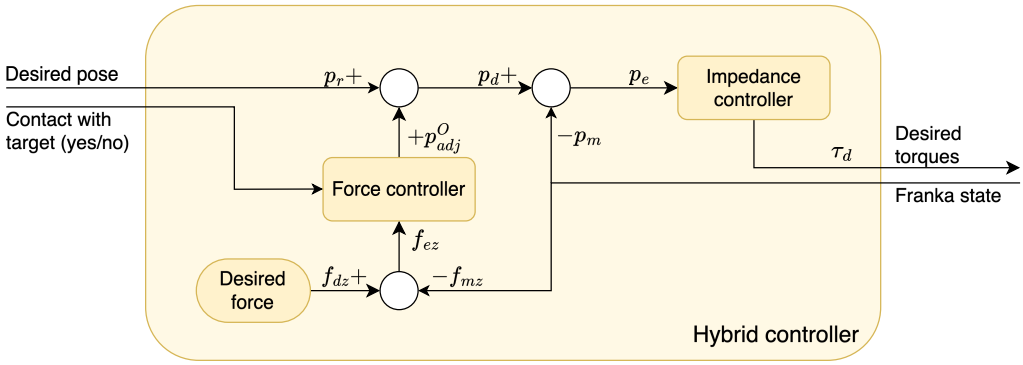


Figure 3. Overview of the hybrid control strategy. The controller enables activation or deactivation of the force control loop depending on the US probe's contact status with the target.

A proportional-integral (PI) controller computes the required position adjustment in the US frame to correct the force error:

$$\mathbf{p}^{EE}_{adj} = \begin{bmatrix} 0 \\ 0 \\ k_p \mathbf{f}_{ez} + k_i \int \mathbf{f}_{ez} dt \end{bmatrix}. \tag{7}$$

This adjustment is then transformed into the robot's base frame:

$$\mathbf{p}^{O}_{adj} = \mathbf{T}^{O}_{EE} \mathbf{p}^{EE}_{adj}. \tag{8}$$

Finally, the desired pose is updated as follows:

$$\mathbf{p}_d = [\tilde{\mathbf{p}}_r + \mathbf{p}^{O}_{adj}, \tilde{\mathbf{q}}_r], \tag{9}$$

where $\mathbf{p}_r = [\tilde{\mathbf{p}}_r, \tilde{\mathbf{q}}_r] \in \mathbb{R}^7$ represents the reference pose.

The Cartesian impedance controller calculates the joint torques necessary to minimize the pose error. Total torque is expressed as:

$$\boldsymbol{\tau}_d = \boldsymbol{\tau}_{\text{task}} + \boldsymbol{\tau}_{\text{nullspace}} + \boldsymbol{\tau}_{\text{coriolis}}. \tag{10}$$

The task-space torque is calculated as:

$$\boldsymbol{\tau}_{\text{task}} = \mathbf{J}^{\top}(-\mathbf{K}_{\text{cart}} \mathbf{p}_e - \mathbf{D}_{\text{cart}} \mathbf{J} \dot{\mathbf{q}}), \tag{11}$$

where $\mathbf{p}_e = [\tilde{\mathbf{p}}_e, \theta_e^{\sim}] \in \mathbb{R}^6$ represents the pose error. Here, $\tilde{\mathbf{p}}_e$ is the position error and $\theta_e^{\sim}$ is derived from the quaternion operation $\tilde{\mathbf{q}}_d^{-1}\tilde{\mathbf{q}}_m$. $\mathbf{K}_{\text{cart}}$ and $\mathbf{D}_{\text{cart}}$ denote the Cartesian stiffness and damping matrices, respectively.

The nullspace torque is used to manage joint motions without affecting the task space:

$$\boldsymbol{\tau}_{\text{nullspace}} = \left(\mathbf{I} - \mathbf{J}^{\top}\mathbf{J}^{+}\right)\left(\mathbf{K}_{\text{null}}(\mathbf{q}_d^{\text{null}} - \mathbf{q}) - 2\sqrt{\mathbf{K}_{\text{null}}}\dot{\mathbf{q}}\right), \tag{12}$$

where $\mathbf{q}_d^{\text{null}}$ is the desired nullspace configuration, $\mathbf{J}^{+}$ is the pseudoinverse of the Jacobian, and $\mathbf{K}_{\text{null}}$ is the nullspace stiffness matrix.

To ensure operational safety and compliance, the robot temporarily stops if the measured force f_{mz} exceeds 6 N, reducing the force before resuming scanning. If $f_{mz} > 15$ N, the scanning operation is aborted, and the robot returns to its initial pose. The control parameters were fine-tuned to minimize steady-state errors while maintaining the stability of the system during dynamic adjustments.

2.5. RA-UNet Architecture

In this study, we utilized a neural network architecture named RA-UNet, designed for the simultaneous segmentation of muscle and bone structures in ultrasound images. The network integrates the classical encoder-decoder structure of UNet [24] with residual connections and attention mechanisms to enhance feature extraction and improve the overall segmentation accuracy for both muscle and bone regions. The overall network architecture is illustrated in Figure 4.

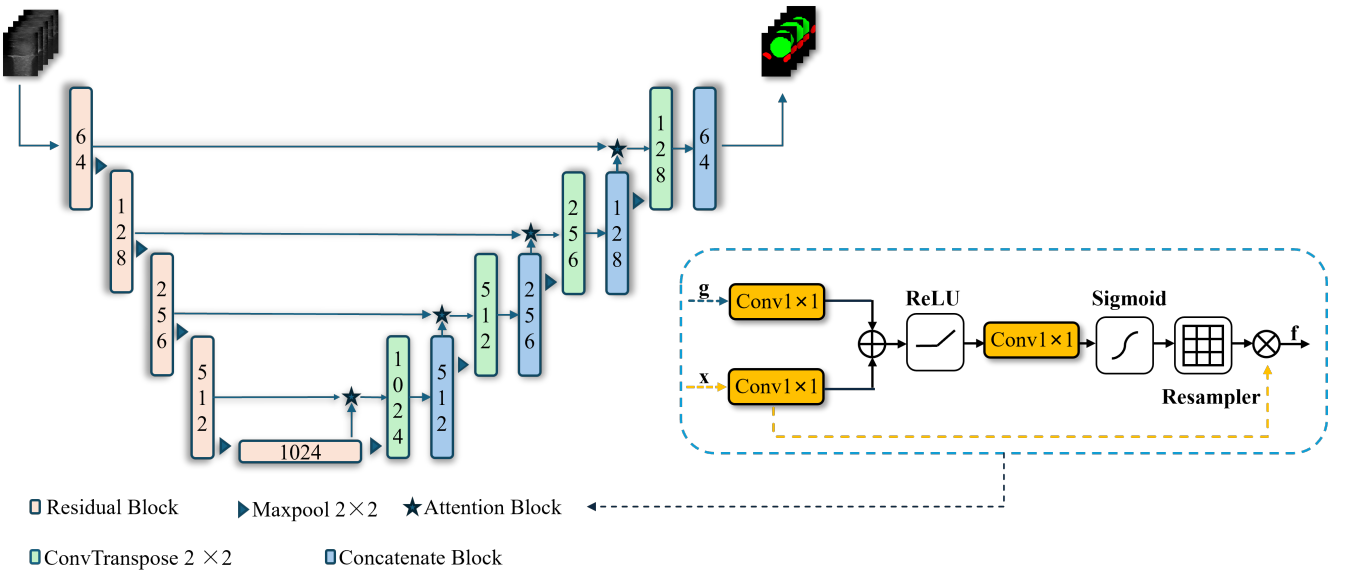


Figure 4. The architecture of the proposed RA-UNet model. The input image size is 128 × 96 pixels, with an initial feature map of 64 channels. The encoder-decoder structure includes direct skip connections (denoted by the symbol g) between corresponding layers to preserve high-resolution features. At each downsampling step, the spatial dimensions are halved while the number of feature channels doubles. In the decoder path, the spatial dimensions are restored through upsampling, and the concatenated feature maps from skip connections help retain fine details, improving segmentation accuracy. The attention mechanism, highlighted in the dashed blue box, computes attention weights based on decoder features (denoted as **g** in the formula) and encoder features (denoted as **x**). The detailed tensor sizes at each stage are shown in Table 2.

Table 2. Tensor sizes at each stage of the RA-UNet model.

Stage	Spatial Dimensions (H × W)	Channels
Input Image	128 × 96	64
After First Downsampling	64 × 48	128
After Second Downsampling	32 × 24	256
After Third Downsampling	16 × 12	512
Bottleneck	8 × 6	1024
After First Upsampling	16 × 12	512
After Second Upsampling	32 × 24	256
After Third Upsampling	64 × 48	128
Output	128 × 96	64

The encoder of RA-UNet consists of multiple residual blocks, each composed of two convolutional layers followed by batch normalization and ReLU [25] activation. These residual blocks are equipped with shortcut connections to enhance gradient flow and mitigate the vanishing gradient problem, thereby facilitating efficient training of the deep network. Each residual block is followed by a max-pooling layer for downsampling, which progressively reduces the spatial resolution while increasing the number of feature channels. The feature dimensions in the encoder path start from 64 and double at each downsampling step, reaching up to 1024, enabling the network to capture rich contextual information across multiple scales.

At the bottleneck of the network, a high-dimensional residual block is employed to extract abstract features at the lowest spatial resolution. These high-level features are critical for distinguishing muscle and bone structures and are progressively restored to high-resolution segmentation maps in the subsequent decoder path.

The decoder path employs transposed convolutions (also referred to as deconvolutions) for upsampling, which restores the spatial resolution of feature maps. To retain high-resolution information, skip connections are introduced between corresponding layers in the encoder and decoder paths. In the RA-UNet architecture, the connections shown in Figure 4 are direct skip connections between the encoder and decoder layers. These skip connections ensure that high-resolution features from the encoder are preserved and passed directly to the decoder, helping to improve segmentation accuracy. The symbol g in the figure represents these direct skip connections, which play a vital role in retaining fine-grained spatial details that would otherwise be lost during downsampling. The skip connections enable the network to better reconstruct anatomical structures, such as muscles and bones, by providing detailed context at each decoding step.

The attention mechanism computes attention coefficients α based on the features from the decoder ($\mathbf{g}$) and the encoder ($\mathbf{x}$) as follows:

$$\alpha = \sigma\left(W_g * \mathbf{g} + W_x * \mathbf{x} + \mathbf{b}_\psi\right), \tag{13}$$

where W_g and W_x are 1×1 convolutional filters, $\mathbf{b}_\psi$ is a bias term, $*$ denotes convolution, and σ is the sigmoid activation function. The attention coefficients α are used to scale the encoder features:

$$\tilde{\mathbf{x}} = \alpha \odot \mathbf{x}, \tag{14}$$

where $\odot$ represents element-wise multiplication. These refined encoder features are concatenated with the upsampled decoder features and further processed by a residual block to integrate the combined information.

Finally, a 1×1 convolutional layer is applied to the output of the decoder to produce the segmentation masks for muscle and bone. Additionally, a combined visualization is generated to facilitate the analysis of the segmentation results.

As shown in Figure 4, the RA-UNet architecture demonstrates the sequential flow of feature extraction, downsampling, upsampling, and feature integration. The attention modules, highlighted in the dashed blue boxes, play a pivotal role in dynamically modulating the features passed through the skip connections, thereby significantly enhancing the overall segmentation accuracy for muscle and bone regions.

2.6. Training and Evaluation

To effectively train the segmentation model, the dataset was manually annotated and constructed using a phantom model. As shown in Figure 5, a customized rectangular gelatin phantom was developed to replicate musculoskeletal structures. This phantom incorporated PVA (polyvinyl alcohol) components to simulate a single muscle layer and PLA (polylactic acid) components to represent two bones. This design ensures an accurate representation of anatomical features while providing a controllable and reproducible environment for system evaluation.

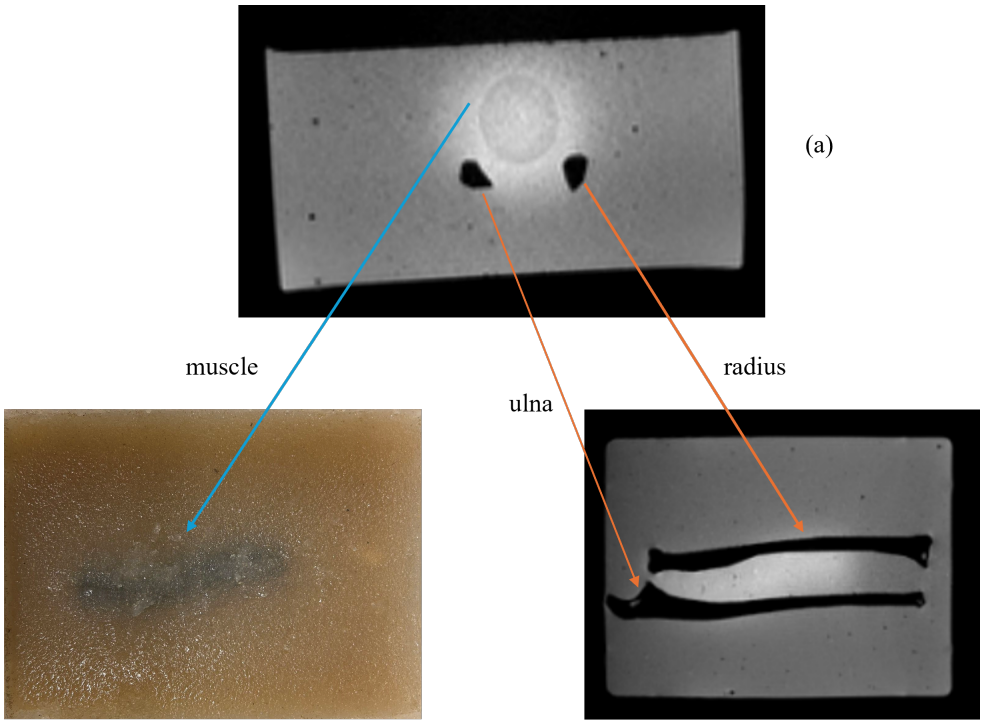


Figure 5. (**a**) Cross-sectional view of the phantom captured through MRI imaging. The circular region (highlighted with a blue arrow) corresponds to the simulated muscle layer, while the two high-contrast black regions (highlighted with orange arrows) represent the ulna and radius. (**b**) Experimental setup showing the phantom enclosed within a rectangular transparent container filled with gelatin medium. The gelatin provides a stable, solid-like environment that simulates biological tissue while preserving the structural integrity of the embedded components. (**c**) Longitudinal MRI view of the phantom, clearly illustrating the structural arrangement of the ulna and radius. Notably, the ulna is characterized by its crescent-shaped end, whereas the radius exhibits a smoother, rounded morphology.

The MRI images were acquired at the University of Twente using a Siemens Aera 1.5T MRI scanner running the Syngo MR E11 software version. The images were collected using a GRE (Gradient Echo) sequence with variable imaging parameters. The repetition time (TR) was approximately 2200 ms, and the echo time (TE) ranged from 2.0 ms to 4.0 ms, depending on the specific slice and imaging conditions. The slice thickness was 1 mm, with a field of view (FoV) of 260 mm × 260 mm and a resolution matrix of 256 × 192. The dataset consists of phantom experiments, and all images were used for muscle and bone segmentation.

The dataset initially consisted of 2D 200 ultrasound images.To improve variability and robustness, data augmentation techniques were applied, including random rotations (±15°), translations (±10 pixels), Gaussian noise addition (standard deviation: 0.01), and horizontal/vertical flips. These augmentations were carefully designed to simulate real-world challenges encountered during ultrasound scanning. For example, random rotations were used to mimic image distortions caused by probe movement and varying applied force, while Gaussian noise was added to replicate noise artifacts commonly observed during real-time ultrasound image transmission. By simulating these real-world conditions, the model's robustness and practical applicability were significantly enhanced, ensuring improved performance in diverse and dynamic scanning environments. After data augmentation, a total of 1600 images were generated. The dataset was split into training and test sets with an 8:2 ratio, resulting in 1280 images for training and 320 images for testing.

The proposed robotic ultrasound system recognizes different anatomical structures by leveraging distinct ultrasound signatures of musculoskeletal tissues. Specifically, the bone tissues are identified based on the presence of a hyperechoic line along the bone cortex, accompanied by a complete acoustic shadow. In contrast, the edges of muscle tissues are recognized due to the presence of a thin hyperechoic rim representing the fascia (epimysium) surrounding the muscle layer. These unique acoustic properties allow the robot to distinguish between muscle and bone boundaries during the segmentation process.

To train the robotic system in identifying these tissues, the segmentation model was provided with manually annotated ultrasound images that included both muscle and bone regions. The model learned to distinguish these tissues by focusing on their respective acoustic properties [26].

During training, images were normalized using the computed mean and standard deviation to ensure consistent intensity distributions across the dataset. Adam's optimizer was used with an initial learning rate of 1×10^{-5}, together with the ReduceLROnPlateau scheduler, to dynamically adjust the learning rate based on validation loss. Specifically, the learning rate is reduced by half if the validation loss does not show a significant decrease over five consecutive epochs. Furthermore, if the improvement in validation loss remains negligible for 15 consecutive epochs, an early stopping mechanism is triggered to prevent overfitting and optimize computational efficiency. The cross-entropy loss was adopted as the loss function to minimize pixel-wise discrepancies between predictions and ground truth labels. The loss is defined as:

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

where N is the total number of pixels, C is the number of classes, $y_{i,c}$ is the ground truth label for class c, and $p_{i,c}$ is the predicted probability for class c.

For validation, a dynamic threshold optimization strategy was employed to refine segmentation accuracy. Thresholds ($T \in [0.1, 1.0]$) were sampled at intervals of 0.05, and the threshold that maximized the average Dice score across all classes was selected:

$$\text{Average Dice}(T) = \frac{1}{C}\sum_{c=1}^{C}\frac{2\cdot|P_c \cap T_c|}{|P_c| + |T_c|}, \tag{16}$$

where P_c and T_c represent the predicted and ground truth masks for class c, respectively. This approach ensured optimal segmentation performance for both bone and muscle regions, leveraging the validation set for parameter tuning effectively.

2.7. Point Cloud Generation from Segmented Images

After segmenting the images, all detected pixel points $\mathbf{p}^I = [u, v, 0, 1]^T$ were first converted into physical dimensions. Specifically, let $S_u = \frac{W_m}{W_p}$ and $S_v = \frac{H_m}{H_p}$ be the scaling factors, where W_m, H_m are the physical dimensions of the ultrasound image in meters, and W_p, H_p are the corresponding pixel dimensions. After applying these scaling factors, the physical coordinates $\mathbf{p}^S$ can be expressed as:

$$\mathbf{p}^S = \begin{bmatrix} S_u \cdot u \\ S_v \cdot v \\ 0 \\ 1 \end{bmatrix}. \tag{17}$$

These coordinates are then mapped into the ultrasound image coordinate system I using the transformation matrix $\mathbf{T}_S$:

$$\mathbf{p}^I = \mathbf{T}_S \cdot \mathbf{p}^S. \tag{18}$$

As established during the ultrasound image calibration step (see corresponding calibration equations in Section 2.2.1), we have the transformation $\mathbf{T}_I^{EE}$ mapping the ultrasound image frame I to the robot's end-effector frame EE:

$$\mathbf{p}^{EE} = \mathbf{T}_I^{EE} \cdot \mathbf{p}^I. \tag{19}$$

Finally, using the known end effector to robot base frame transformation $\mathbf{T}_{EE}^O$, we can map the point into the robot base frame O:

$$\mathbf{p}^O = \mathbf{T}_{EE}^O \cdot \mathbf{p}^{EE}. \tag{20}$$

To enhance the quality of the resulting point cloud, edge detection was applied to the segmented image using the Canny algorithm, isolating meaningful boundary points. Pixels near the image edges ($u < 25, u > W_p - 5, v < 10, v > H_p - 20$) were filtered out to minimize noise. The valid 3D points were then categorized by their segmentation class and saved into separate text files, each containing the 3D coordinates $[x, y, z]$ for the corresponding class. This pipeline ensures that the reconstructed point cloud is accurately aligned with the robot's coordinate frame, thereby enabling precise spatial representation for downstream applications.

In summary, the series of transformations from (17) to (20) converts 2D segmented image data into a 3D representation aligned with the robot's base frame, facilitating accurate 3D analysis and visualization in subsequent processing steps.

3. Experiment and Results

3.1. Segmentation Results and Performance of RA-UNet

The RA-UNet model was trained using a threshold adaptation mechanism to optimize the Dice score. During training, the threshold was dynamically adjusted, achieving an

optimal value of 0.85. The training process consisted of a total of 100 iterations, with the best-performing model selected based on the Dice coefficient on the validation set.

The segmentation results, as illustrated in Figure 6, highlight the effectiveness of the proposed method in segmenting musculoskeletal structures from ultrasound images. The model accurately identifies muscle and bone regions, demonstrating its robustness and capability to delineate these key anatomical features. The combined visualizations provide a comprehensive overview, reinforcing the segmentation quality and the alignment between predictions and ground truth annotations.

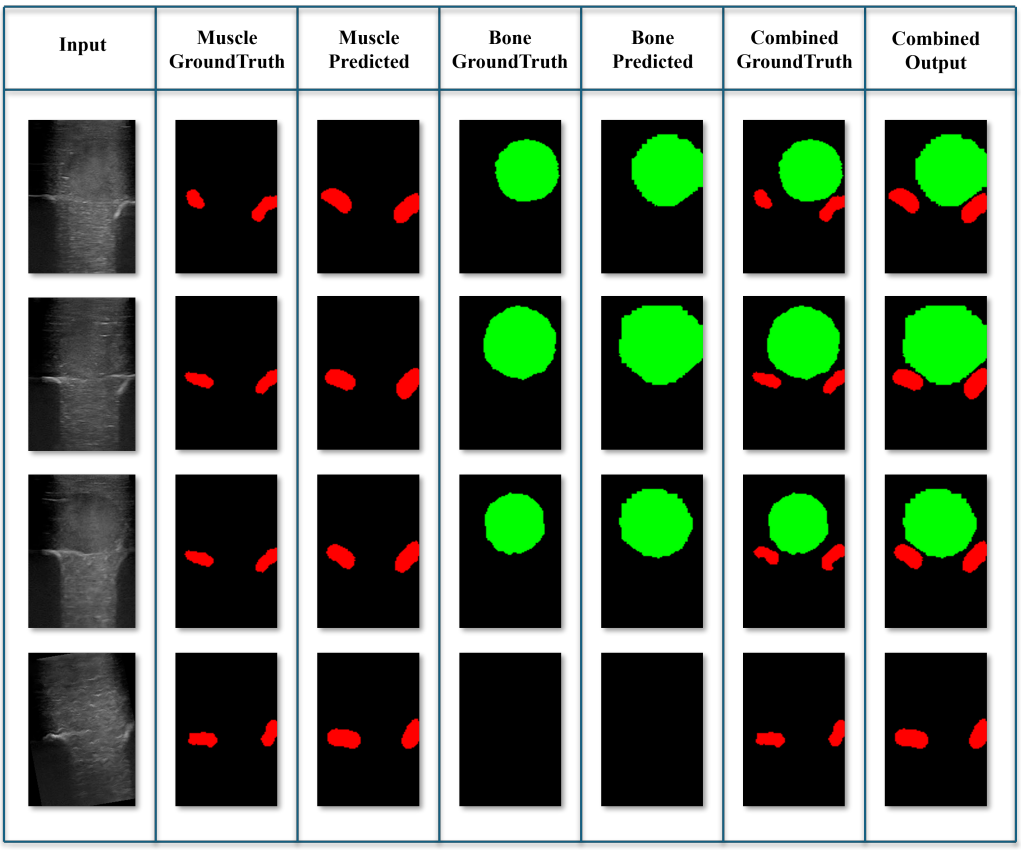


Figure 6. The figure illustrates the segmentation results of ultrasound images. The first column represents the input ultrasound images, serving as the model input. The second and third columns show the ground truth and predicted segmentation results for bone (red), respectively. Similarly, the fourth and fifth columns depict the ground truth and predicted segmentation results for muscle (green). The sixth and seventh columns present the combined ground truth and predicted segmentation results, respectively, providing an overview of all segmented structures in a single visualization.

To compare the performance of RA-UNet and the baseline UNet model, the Dice scores for both models are presented in Table 3. RA-UNet demonstrated superior segmentation performance, achieving higher Dice scores across all classes. The overall Dice score for RA-UNet was 0.83, compared to 0.785 for UNet, highlighting the effectiveness of incorporating residual and attention mechanisms in improving segmentation accuracy.

Table 3. Comparison of RA-UNet and UNet model training results for the phantom experimentation.

Class	RA-UNet Dice Score	UNet Dice Score
Muscle (Class 0)	0.82	0.78
Bone (Class 1)	0.84	0.79
Overall	0.83	0.785

The results confirm the superior performance of RA-UNet in segmenting both bone and muscle tissues. By integrating residual connections and attention mechanisms, RA-UNet effectively addressed the challenges of complex segmentation tasks, achieving more accurate and robust results compared to the baseline UNet.

3.2. Performance Evaluation of Scanning Trajectories

Several tests were conducted on the phantom, and the results are illustrated in Figure 7 and Table 4. The robotic ultrasound scanning system achieved an overall RMSE of 8.37 mm, with individual RMSE values of 6.61 mm along the x-axis, 1.79 mm along the y-axis, and 4.80 mm along the z-axis. The higher RMSE in the x-axis indicates greater deviations, likely due to challenges in maintaining consistent lateral probe contact. In contrast, the lower RMSE in the y and z axes reflects more stable performance in vertical and depth adjustments. These results highlight the need for improved lateral control and suggest that integrating real-time feedback mechanisms could further enhance accuracy. Despite these discrepancies, the system met performance requirements and demonstrated robust trajectory planning for ultrasound imaging.

Table 4. Root Mean Square Error (RMSE) values for X, Y, Z axes and overall performance

Case	RMSE (X, mm)	RMSE (Y, mm)	RMSE (Z, mm)	Overall RMSE (mm)
1	6.71	1.77	4.81	8.44
2	6.62	1.81	4.76	8.35
3	6.51	1.80	4.84	8.31
Average	6.61	1.79	4.80	8.37

Table Note: RMSE = Root Mean Square Error; mm = millimeters.

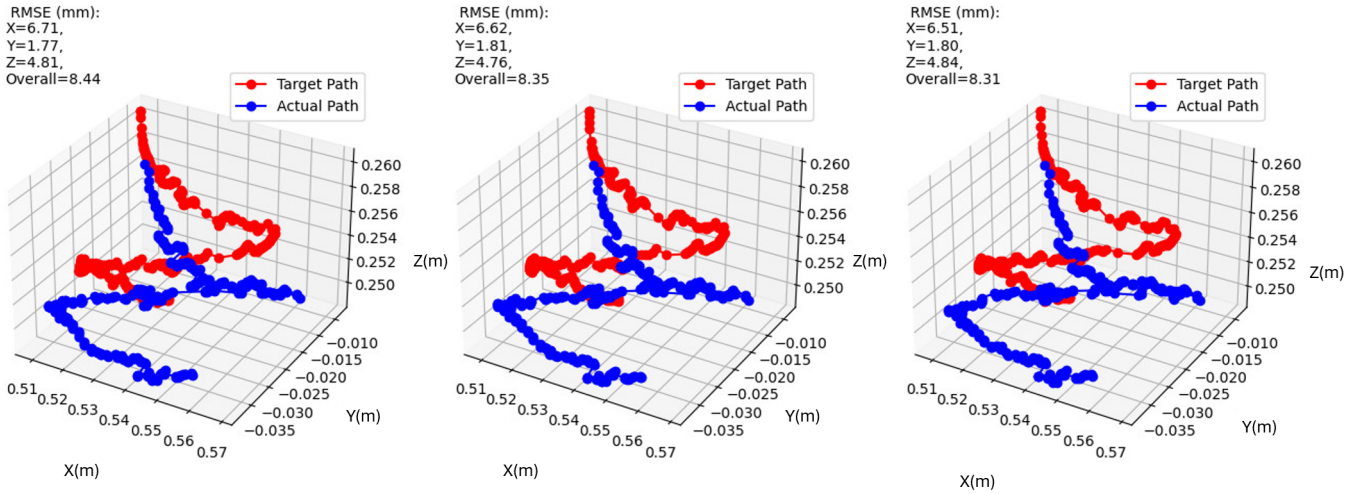


Figure 7. 3D visualization of the target path (red) and actual path (blue). RMSE values for X, Y, and Z axes (m) and overall performance are indicated, showing the deviation between the two paths in millimeters.

3.3. Performance Evaluation of Reconstruction

To evaluate the accuracy of the system's 3D reconstruction, we quantitatively compared the system's output with the MRI ground truth (see Figure 8). The comparison resulted in an RMSE of 1.22 mm, a mean error of 0.94 mm, and a standard deviation of 0.77 mm, demonstrating the system's high accuracy in 3D reconstruction and its ability to effectively replicate the three-dimensional structure of the target.

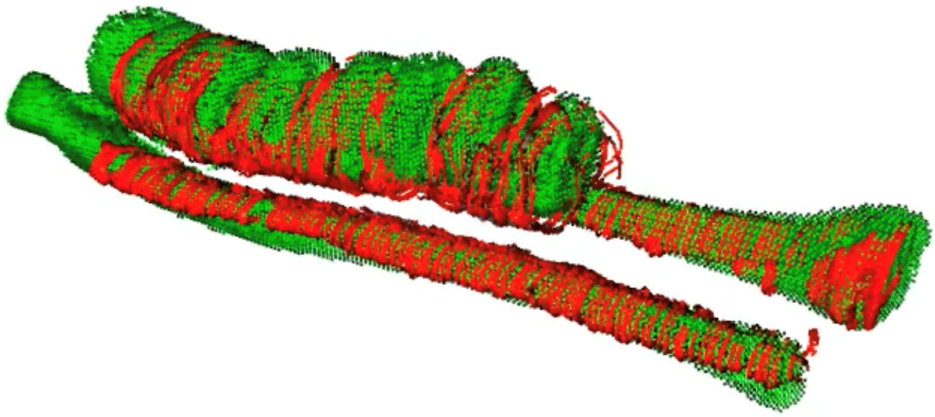


Figure 8. The comparison between MRI (green) and US reconstruction (red) demonstrates good alignment and overlap between the two visual representations. Quantitatively, the RMSE is 1.22 mm, with a mean error of 0.94 mm and a standard deviation of 0.77 mm.

3.4. Force Control Performance and Stability Analysis

Table 5 summarizes the results of force control performance during robotic ultrasound scanning across three trials. The root mean square error (RMSE) values for Trials A, B, and C were 0.3181 N, 0.2723 N, and 0.2519 N, respectively, with an average RMSE of 0.2808 N. The mean applied force across all trials remained consistent, averaging 2.9420 N, while the standard deviation averaged 0.2745 N, indicating stable force application throughout the experiments.

Table 5. Force Data Statistics for Three Trials (A, B, C) and Average.

Trial	RMSE (N)	Mean (N)	Standard Deviation (N)
A	0.3181	2.9445	0.3132
B	0.2723	2.9323	0.2637
C	0.2519	2.9493	0.2467
Average	0.2808	2.9420	0.2745

Table Note: RMSE = Root Mean Square Error; N = Newton.

Figure 9 illustrates additional details of the z-axis force control. The left panel shows the force curve over time for one of the trials, demonstrating minimal fluctuation with an RMSE of 0.2485 N and a mean force of 2.9582 N. The standard deviation of 0.2450 N confirms the stability of the applied force during the scanning process.

The right panel presents a heat map of the force distribution along the scanning trajectory. Regions with consistent colors reflect uniform force application and stable probe-tissue contact, while areas with notable color variation indicate minor fluctuations caused by adjustments to the probe angle or contact alignment.

We carefully assessed the clarity of the images and the visibility of anatomical structures, such as bones and muscles. It is important to note that minor fluctuations do not affect the contact between the ultrasound probe and the surface of the phantom, which is a necessary condition for maintaining high-quality ultrasound imaging. Therefore, although minor fluctuations in probe pressure were observed, we determined that these fluctuations did not significantly impact segmentation accuracy.

Overall, these results validate the robustness and reliability of the hybrid position–force control system in maintaining steady probe-tissue interaction. This stability is critical for achieving high-quality ultrasound imaging and reducing potential artifacts, highlighting the system's capability to ensure consistent scanning performance across multiple trials.

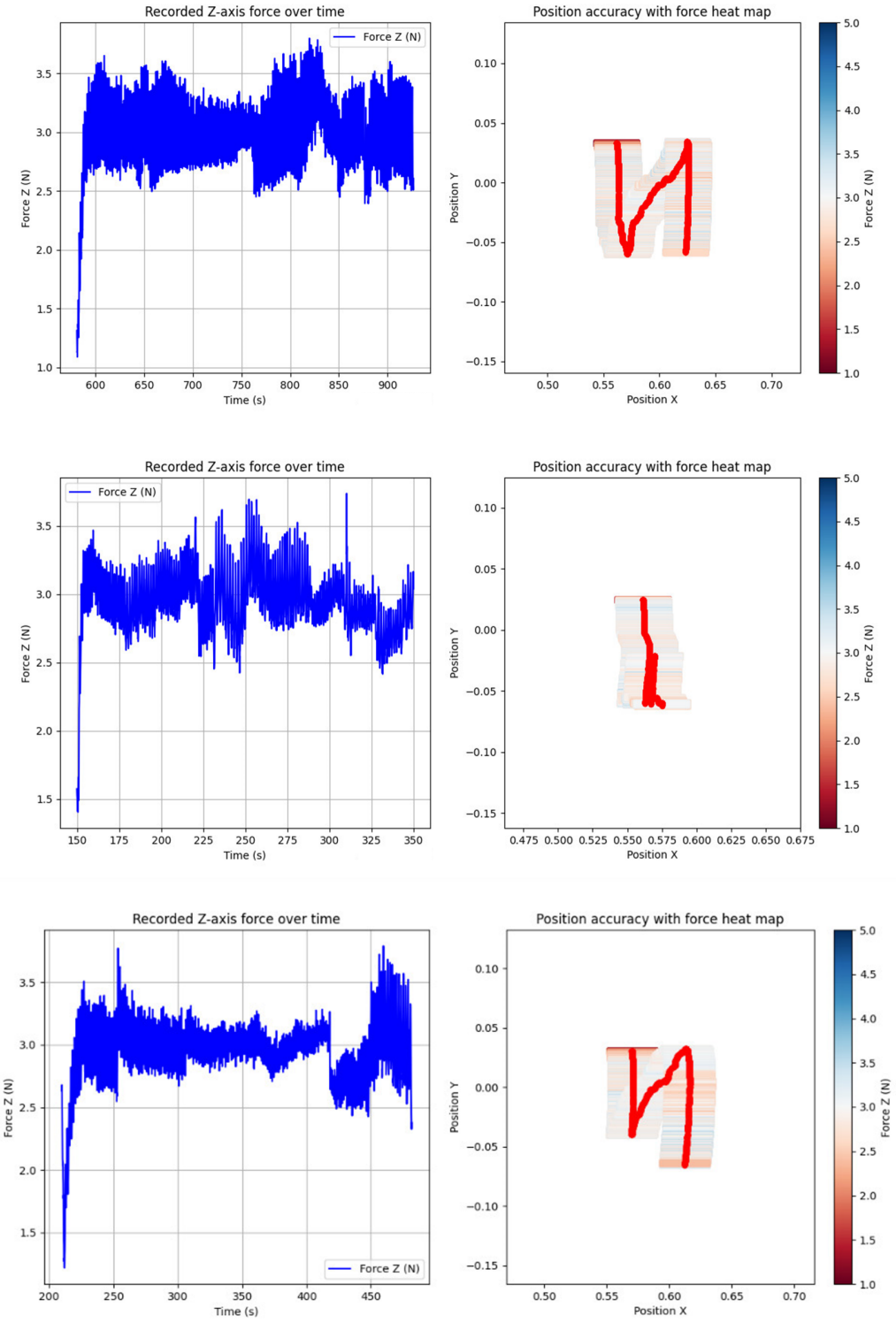


Figure 9. (**Left**) The Z-axis force plot demonstrates consistent application with minimal variation. (**Right**) The heat map illustrates the spatial distribution of applied force, highlighting stable regions and minor fluctuations along the scanning trajectory.

4. Discussion

This study presents an ARUS system that integrates a hybrid position–force control strategy and an enhanced RA-UNet model, addressing key challenges in musculoskeletal imaging. The system demonstrated significant advancements in segmentation accuracy, 3D reconstruction precision, and operational safety. Below, we discuss the broader implications, limitations, and potential future directions of this work.

4.1. Segmentation Performance

The ARUS system achieved robust segmentation results, with Dice coefficients of 0.82 and 0.84 for muscle and bone structures, respectively. These outcomes highlight the effectiveness of the RA-UNet model in addressing challenges inherent in ultrasound imaging, such as noise and tissue variability. The incorporation of residual connections [27] and attention mechanisms enhanced feature extraction and spatial focus, while skip connections improved the integration of multi-scale information [24,28,29].

In the study by Marzola et al. [30], an ensemble model was proposed that combined the predictions of four deep networks (Unet [24], Unet++ [31], FPN [32], and Attention Unet [33]). The final segmentation results achieved a Dice coefficient of 0.9. However, their work focused solely on muscle segmentation, neglecting the segmentation of bone structures. This limitation reduces the applicability of their model when dealing with more complex anatomical structures, such as the simultaneous segmentation of muscles and bones.

The experimental results of the phantom data demonstrate that the model performs exceptionally well in the segmentation of bones and muscles, highlighting its robustness in handling complex anatomical structures, noise, and variability. These findings showed the reliability of our approach in phantom-based experiments, laying a foundation for its future application in human studies. However, the limitation of phantom data lies in its lack of diversity, which cannot fully represent the complexity of patient-specific anatomical structures in real clinical environments.

Future studies will focus on validating the model performance using human data by incorporating more complex anatomical structures and diverse patient datasets to comprehensively evaluate its applicability in real-world clinical scenarios. Additionally, we will explore multi-class segmentation frameworks to integrate additional anatomical features, such as vascular structures, further enhancing the system's robustness and generalizability. These efforts will provide critical support for advancing the model's deployment in clinical practice.

4.2. 3D Reconstruction Quality

Several studies have focused on isolated aspects of 3D anatomical reconstruction. For instance, bone reconstruction has been the primary focus in the works of Kim et al. [34] and Roth et al. [35], utilizing X-ray images and biplanar radiographs to derive skeletal structures for clinical applications. On the other hand, muscle reconstruction has been addressed in the studies by McColl et al. [36] and Guenanten et al. [37], which employed MRI and 3D ultrasound to evaluate soft tissue properties such as muscle-tendon unit lengths and muscle pathology. Additionally, this study compared the reconstructed muscle volumes obtained through ultrasound with MRI benchmarks, focusing on the overall muscle tissue. This comparison provides a comprehensive assessment of the reconstructed musculoskeletal structures. Additionally, systems like ARUS have demonstrated exceptional accuracy in 3D reconstruction, achieving an average error of 1.22 mm compared to MRI benchmarks and offering real-time results for clinical applications.

In contrast to these approaches, our proposed method offers a significant advantage by enabling simultaneous 3D reconstruction of both bone and muscle tissues. This integration provides clinicians with a holistic anatomical view, which is critical for understanding the complex interactions between the skeletal and muscular systems. Furthermore, while systems like ARUS emphasize real-time performance and precision in stable probe-tissue interactions, our method builds upon these strengths by incorporating multi-tissue reconstruction capabilities. This not only enhances diagnostic accuracy but also broadens the potential applications, making it particularly suitable for cases requiring detailed evalu-

ation of both bone and muscle structures, such as in sports medicine, rehabilitation, and complex surgical planning.

However, the reliance on phantom validation introduces limitations. Real-world applications will likely involve more complex anatomical structures and dynamic factors, such as tissue elasticity and patient-specific variability. These factors could affect reconstruction accuracy and require further investigation.

To address these challenges, future studies should validate the system using live patient data under controlled conditions. Additionally, integrating complementary imaging modalities, such as elastography, could enhance the diagnostic value by providing biomechanical information, such as tissue stiffness or elasticity.

4.3. Operational Safety

Safety is a critical consideration for the ARUS system, and the hybrid position–force control strategy has demonstrated reliable and consistent performance. Across multiple trials, the system achieved an average applied force of 2.9420 N with RMSE of 0.2808 N, ensuring stable and safe probe-tissue interaction. The uniform force distribution observed in the heat map analysis underscores the system's robustness in maintaining consistent contact, minimizing imaging artifacts, and achieving high-quality ultrasound imaging.

Despite these promising results, the current reliance on joint torque sensors for force estimation presents certain limitations. These sensors, while effective, are susceptible to computational delays and calibration sensitivity, particularly under high-motion conditions or in extreme joint configurations. Such challenges may compromise the accuracy of force estimation and impact the system's reliability in dynamic or demanding operational scenarios.

To address these limitations, future work will integrate real-time visual feedback by combining depth cameras and ultrasound imaging. Depth cameras will measure the degree of surface deformation on the muscle, while ultrasound images will track changes in muscle thickness. This dual-modality feedback will enable dynamic adjustments to the applied force, enhancing the system's adaptability to varying anatomical and operational conditions. These advancements are expected to further increase the robustness of the system and broaden its applicability to complex clinical scenarios, such as dynamic tissue interactions and patient-induced motion during scanning.

4.4. Path Error Analysis

In this study, larger errors were observed in the X direction (forward direction) compared to the Z direction (depth direction), primarily due to the mechanical properties of the phantom model used. The phantom model is softer and more deformable than human skin and soft tissues, resulting in greater deformation in the X direction during scanning. In contrast, the Z direction, representing the depth of the probe's contact with the surface, was better controlled by maintaining a consistent contact force throughout the scanning process, thereby reducing errors in this direction. Additionally, task torque primarily affects joint-level control and has minimal impact on the positional accuracy of the end effector in the X and Y directions. Therefore, the deformation and contact dynamics of the phantom model are considered the main contributors to the observed errors in the X direction, and future work will explore real-time deformation correction algorithms to further improve path accuracy.

4.5. Clinical Implications

The proposed ARUS system offers promising potential in clinical musculoskeletal imaging, especially in preoperative planning, rehabilitation monitoring, and sports medicine. By providing real-time 3D reconstruction of bones and muscles, the system enhances diagnostic precision and offers a comprehensive anatomical perspective. The

reduction in operator dependency achieved through hybrid position–force control ensures consistent image quality across diverse scenarios. Furthermore, the ability to visualize both soft and hard tissues simultaneously can be beneficial for clinicians in evaluating injuries or conditions involving complex interactions between muscles and bones, thus improving the effectiveness of therapeutic interventions. Compared to conventional imaging modalities such as MRI, the ARUS system provides a cost-effective solution, making it more accessible for clinical and research applications in resource-limited settings. This affordability makes it a practical alternative for institutions seeking advanced imaging capabilities without incurring substantial costs.

4.6. Limitations and Future Work

While the ARUS system shows promising potential for musculoskeletal diagnostics, several limitations should be acknowledged. First, the system has primarily been tested on phantom models, which differ from real human tissue in terms of mechanical properties and deformation characteristics. This discrepancy may affect the accuracy of the scanning path and image quality when applied to real patients. To ensure clinical applicability, future efforts will focus on validating the ARUS system with real patient datasets to assess its robustness across diverse anatomical and pathological conditions. By bridging gaps in current musculoskeletal imaging, ARUS has the potential to become a practical, reliable tool for diagnostics and advanced therapeutic planning.

Second, the current system does not incorporate real-time deformation correction algorithms. Although the system maintains a consistent contact force during scanning, tissue deformation may still cause deviations in the scanning path. Existing studies have explored real-time deformation tracking and correction methods to improve the accuracy of robotic ultrasound systems. For instance, Virga et al. [38] proposed a force feedback-based correction algorithm, while Chen et al. [39] and Jiang et al. [40] investigated deformation prediction and path-planning strategies. These studies highlight the importance of real-time deformation correction in improving system stability and imaging accuracy.

Third, the dataset used for training and validating the segmentation algorithm is relatively small, primarily consisting of phantom model images. This limitation may hinder the generalizability of the segmentation results across diverse anatomical and pathological conditions. Future work will focus on expanding the dataset by incorporating real patient images and conducting extensive validation to ensure the system's reliability across various clinical scenarios.

Furthermore, improving the network architecture and segmentation algorithm is critical to enhancing the performance of the ARUS system. Recent studies have demonstrated significant improvements in medical image segmentation tasks using advanced architectures such as Transformer-based models, attention mechanisms, and hybrid frameworks. For instance, Zhang et al. [41] proposed TransUNet, a hybrid Transformer-CNN framework that effectively captures both global and local features, achieving superior performance in complex medical image segmentation tasks. Chen et al. [42] introduced AAU-Net, an adaptive attention UNet designed for breast lesion segmentation, while Ling et al. [43] proposed MTANet, a multi-task attention network that integrates segmentation and classification tasks. These advancements are expected to enhance the robustness and clinical applicability of the ARUS system across diverse clinical scenarios.

Lastly, the system currently requires manual selection of specific points for trajectory generation, which is time-consuming and subject to operator variability. Future research will focus on automating this process using advanced image processing and machine learning techniques. Reinforcement learning (RL) has shown significant potential in automating robotic ultrasound scanning by learning optimal trajectories through interaction with the

environment. Studies have proposed RL-based frameworks for autonomous ultrasound imaging, including force-guided control methods and path-planning strategies to address tissue deformation and acoustic shadowing [44–46]. These approaches highlight the potential of RL to enable fully automated and adaptive trajectory generation, improving both the robustness and clinical applicability of the ARUS system.

5. Conclusions

This paper presents an ARUS system that integrates a hybrid position–force control strategy and an enhanced RA-UNet model for real-time segmentation of muscle and bone structures as well as precise 3D reconstruction. Experimental validation on a customized musculoskeletal phantom demonstrated that the system achieved average Dice coefficients of 0.82 and 0.84 for muscle and bone segmentation, respectively, with a 3D reconstruction root mean square error (RMSE) of 1.22 mm, a mean error of 0.94 mm, and a standard deviation of 0.77 mm compared to the MRI benchmark. These results showed relatively high accuracy and robustness. This study highlights the potential of the ARUS system in the field of musculoskeletal imaging, providing a foundation for its application in real-time image analysis, musculoskeletal biomechanical modeling, and related downstream tasks. Future work will focus on extending the system's applicability, thereby further advancing its value in medical diagnosis and treatment.

Author Contributions: Conceptualization, D.S. and A.C.; Formal analysis, D.S.; Investigation, A.C.; Methodology, D.S. and A.C.; Project administration, S.S. and K.N.; Resources, K.N.; Software, D.S. and B.L.; Supervision, S.S. and K.N.; Validation, D.S. and A.C.; Visualization, D.S.; Writing—original draft, D.S. and A.C.; Writing—review and editing, D.S., A.C., B.L., M.A., S.S. and K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

US	Ultrasound
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
RA-UNet	Residual Attention UNet
ROI	Region of Interest
EE	End Effector
TGC	Time Gain Compensation
PC	Personal Computer
UDP	User Datagram Protocol
RMSE	Root Mean Square Error
MRE	Mean Relative Error
PI Controller	Proportional-Integral Controller
CE Loss	Cross-Entropy Loss
GPU	Graphics Processing Unit
CPU	Central Processing Unit
RAM	Random Access Memory

References

1. Miller, D.L.; Smith, N.B.; Bailey, M.R.; Czarnota, G.J.; Hynynen, K.; Makin, I.R. Overview of therapeutic ultrasound applications and safety considerations. *J. Ultrasound Med.* **2012**, *31*, 623–634. [CrossRef]
2. Lyu, Y.; Shen, Y.; Zhang, M.; Wang, J. Real-Time 3D Ultrasound Imaging System Based on a Hybrid Reconstruction Algorithm. *Chin. J. Electron.* **2024**, *33*, 245–255. [CrossRef]
3. Lan, B.; Abayazid, M.; Verdonschot, N.; Stramigioli, S.; Niu, K. Deep Learning based acoustic measurement approach for robotic applications on orthopedics. *arXiv* **2024**, arXiv:2403.05879.
4. Erlandson, M.; Lorbergs, A.; Mathur, S.; Cheung, A. Muscle analysis using pQCT, DXA and MRI. *Eur. J. Radiol.* **2016**, *85*, 1505–1511. [CrossRef] [PubMed]
5. Jacobson, J.A. Musculoskeletal Ultrasound: Focused Impact on MRI. *Am. J. Roentgenol.* **2009**, *193*, 619–627. [CrossRef]
6. Li, R.; Davoodi, A.; Niu, K.; Cai, Y.; Vander Poorten, E. Automatic Robotic Scanning for real-time 3D Ultrasound Reconstruction in Spine Surgery. In Proceedings of the 11th Conference on New Technologies for Computer and Robot Assisted Surgery, Napoli, Italy, 25–27 April 2022.
7. Victorova, M.; Navarro-Alarcon, D.; Zheng, Y.P. 3D Ultrasound Imaging of Scoliosis with Force-Sensitive Robotic Scanning. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 262–265. [CrossRef]
8. Li, R.; Davoodi, A.; Cai, Y.; Niu, K.; Borghesan, G.; Cavalcanti, N.; Massalimova, A.; Carrillo, F.; Laux, C.J.; Farshad, M.; et al. Robot-assisted ultrasound reconstruction for spine surgery: From bench-top to pre-clinical study. *Int. J. Comput. Assist. Radiol. Surg.* **2023**, *18*, 1613–1623. [CrossRef]
9. Davoodi, A.; Li, R.; Cai, Y.; Niu, K.; Borghesan, G.; Poorten, E.V. A Comparative Study for Control of Semi-Automatic Robotic-assisted Ultrasound System in Spine Surgery. In Proceedings of the 21st International Conference on Advanced Robotics (ICAR), Abu Dhabi, United Arab Emirates, 5–8 December 2023.
10. Li, R.; Niu, K.; Poorten, E.V. 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. [CrossRef]
11. Li, R.; Cai, Y.; Niu, K.; Poorten, E.V. Comparative Quantitative Analysis of Robotic Ultrasound Image Calibration Methods. In Proceedings of the 2021 20th International Conference on Advanced Robotics (ICAR), Ljubljana, Slovenia, 6–10 December 2021; pp. 511–516. [CrossRef]
12. Cai, Y.; Davoodi, A.; Li, R.; Ourak, M.; Niu, K.; Deprest, J.; Poorten, E.V. Development of Robot-assisted Ultrasound System for Fetoscopic Tracking in Twin to Twin Transfusion Syndrome Surgery. In Proceedings of the 2023 International Symposium on Medical Robotics (ISMR), Atlanta, GA, USA, 19–21 April 2023; pp. 1–7. [CrossRef]
13. Huang, D.; Yang, C.; Zhou, M.; Karlas, A.; Navab, N.; Jiang, Z. Robot-Assisted Deep Venous Thrombosis Ultrasound Examination Using Virtual Fixture. *IEEE Trans. Autom. Sci. Eng.* **2024**, *22*, 381–392. [CrossRef]
14. Enebuse, I.; Foo, M.; Ibrahim, B.S.K.K.; Ahmed, H.; Supmak, F.; Eyobu, O.S. A Comparative Review of Hand-Eye Calibration Techniques for Vision Guided Robots. *IEEE Access* **2021**, *9*, 113143–113155. [CrossRef]
15. Li, H.; Chen, H.; Jing, W.; Li, Y.; Zheng, R. 3D Ultrasound Spine Imaging with Application of Neural Radiance Field Method. In Proceedings of the 2021 IEEE International Ultrasonics Symposium (IUS), Xi'an, China, 11–16 September 2021; pp. 1–4. [CrossRef]
16. Guo, H.; Xu, S.; Wood, B.; Yan, P. Sensorless Freehand 3D Ultrasound Reconstruction via Deep Contextual Learning. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2020*; Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L., Eds.; Springer: Cham, Switzerland, 2020; pp. 463–472.
17. Shiode, R.; Kabashima, M.; Hiasa, Y.; Oka, K.; Murase, T.; Sato, Y.; Otake, Y. 2D–3D reconstruction of distal forearm bone from actual X-ray images of the wrist using convolutional neural networks. *Sci. Rep.* **2021**, *11*, 15249. [CrossRef]
18. Liu, T.; Lu, Y.; Xu, J.; Yang, H.; Hu, J. 3D reconstruction of bone CT scan images based on deformable convex hull. *Med. Biol. Eng. Comput.* **2024**, *62*, 551–561. [CrossRef] [PubMed]
19. Shrestha, P.; Xie, C.; Shishido, H.; Yoshii, Y.; Kitahara, I. 3D Reconstruction of Bone from Multi-View X-Ray Images Using Planar Markers. In Proceedings of the 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), Osaka, Japan, 7–9 March 2022; pp. 189–193. [CrossRef]
20. Lan, B.; Niu, K. Predicting Muscle Thickness Deformation from Muscle Activation Patterns: A Dual-Attention Framework. *arXiv* **2024**, arXiv:2409.18266.
21. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]
22. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying Density-Based Local Outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD '00), Dallas, TX, USA, 15–18 May 2000; ACM: New York, NY, USA, 2000; pp. 93–104. [CrossRef]
23. Bentley, J.L. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM* **1975**, *18*, 509–517. [CrossRef]

24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]
25. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010; pp. 807–814.
26. Doe, J.; Smith, A.; Brown, B. Training principles for musculoskeletal tissue differentiation in ultrasound imaging. *Am. J. Phys. Med. Rehabil.* **2022**, *101*, e189–e193. [CrossRef]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Lan, B.; Abayazid, M.; Verdonschot, N.; Stramigioli, S.; Niu, K. SIRC-UNet: Improving Bone Tracking Precision of A-Mode Ultrasound by Decoding Hierarchical Resolution Features. *IEEE Sens. J.* **2024**, *24*, 38174–38184. [CrossRef]
29. Lan, B.; Stramigioli, S.; Niu, K. Anatomical Region Recognition and Real-time Bone Tracking Methods by Dynamically Decoding A-Mode Ultrasound Signals. *arXiv* **2024**, arXiv:2405.19542.
30. Marzola, F.; van Alfen, N.; Doorduin, J.; Meiburger, K.M. Deep learning segmentation of transverse musculoskeletal ultrasound images for neuromuscular disease assessment. *Comput. Biol. Med.* **2021**, *135*, 104623. [CrossRef] [PubMed]
31. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; pp. 3–11.
32. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
33. Oktay, O.; Schlemper, J.; Le Folgoc, L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
34. Kim, H.; Lee, K.; Lee, D.; Baek, N. 3D Reconstruction of Leg Bones from X-Ray Images using CNN-based Feature Analysis. In Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 16–18 October 2019; pp. 669–672. [CrossRef]
35. Roth, T.A.; Jokeit, M.; Sutter, R.; Vlachopoulos, L.; Fucentese, S.F.; Carrillo, F.; Snedeker, J.G.; Esfandiari, H.; Fürnstahl, P. Deep-learning based 3D reconstruction of lower limb bones from biplanar radiographs for preoperative osteotomy planning. *Int. J. Comput. Assist. Radiol. Surg.* **2024**, *19*, 1843–1853. [CrossRef]
36. McColl, R.W.; Fleckenstein, J.L.; Bowers, J.; Theriault, G.; Peshock, R.M. Three-dimensional reconstruction of skeletal muscle from MRI. *Comput. Med. Imaging Graph.* **1992**, *16*, 363–371. [CrossRef]
37. Guenanten, H.; Retailleau, M.; Dorel, S.; Sarcher, A.; Colloud, F.; Nordez, A. Muscle-Tendon Unit Length Measurement Using 3D Ultrasound in Passive Conditions: OpenSim Validation and Development of Personalized Models. *Ann. Biomed. Eng.* **2024**, *52*, 997–1008. [CrossRef]
38. Virga, S.; Göbl, R.; Baust, M.; Navab, N. Use the force: Deformation correction in robotic 3D ultrasound. *Int. J. Comput. Assist. Radiol. Surg.* **2018**, *13*, 619–627. [CrossRef] [PubMed]
39. Chen, T.; Zhang, J.; Li, X.; Wang, Y. Ultrasound-Guided Robotic Autonomous Operation Based on Real-Time Deformation Tracking and Prediction. *IEEE Trans. Ind. Inform.* **2024**, *21*, 1369–1378. [CrossRef]
40. Jiang, Z.; Zhou, Y.; Bi, Y.; Zhou, M.; Wendler, T.; Navab, N. Deformation-Aware Robotic 3D Ultrasound. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7675–7682. [CrossRef]
41. Zhang, Y.; Liu, H.; Hu, Q.; Cao, Z.; Guo, D.; Hu, H. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Strasbourg, France, 27 September–1 October 2021; pp. 91–101. [CrossRef]
42. Chen, G.; Li, L.; Dai, Y.; Zhang, J.; Yap, M.H. AAU-Net: An Adaptive Attention U-Net for Breast Lesions Segmentation in Ultrasound Images. *IEEE Trans. Med. Imaging* **2023**, *42*, 1289–1300. [CrossRef] [PubMed]
43. Ling, Y.; Wang, Y.; Dai, W.; Yu, J.; Liang, P.; Kong, D. MTANet: Multi-Task Attention Network for Automatic Medical Image Segmentation and Classification. *IEEE Trans. Med. Imaging* **2024**, *43*, 674–685. [CrossRef]
44. Ning, G.; Zhang, X.; Liao, H. Autonomic Robotic Ultrasound Imaging System Based on Reinforcement Learning. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 2787–2797. [CrossRef]
45. Ning, G.; Chen, J.; Zhang, X.; Liao, H. Force-Guided Autonomous Robotic Ultrasound Scanning Control Method for Soft Uncertain Environment. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 2189–2199. [CrossRef] [PubMed]
46. Bi, Y.; Qian, C.; Zhang, Z.; Navab, N.; Jiang, Z. Autonomous Path Planning for Intercostal Robotic Ultrasound Imaging Using Reinforcement Learning. *arXiv* **2024**, arXiv:2404.09927.

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.