Original Article

Deep learning based ultrasonic visualization of distal humeral cartilage for image-guided therapy: a pilot validation study

Wei Zhao[1#^], Xiuyun Su[2#], Yao Guo[1], Haojin Li[3], Shiva Basnet[1], Jianyu Chen[3], Zide Yang[3], Rihang Zhong[3], Jiang Liu[3], Elvis Chun-sing Chui[4], Guoxian Pei[1,2], Heng Li[3]

[1]School of Medicine, Southern University of Science and Technology, Shenzhen, China; [2]Medical Intelligence and Innovation Academy, Southern University of Science and Technology Hospital, Shenzhen, China; [3]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China; [4]Department of Orthopaedics and Traumatology, The Chinese University of Hong Kong, Hong Kong, China

Contributions: (I) Conception and design: W Zhao, G Pei, EC Chui; (II) Administrative support: X Su, H Li; (III) Provision of study materials or patients: Y Guo, H Li; (IV) Collection and assembly of data: S Basnet, J Chen; (V) Data analysis and interpretation: Z Yang, R Zhong, J Liu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Guoxian Pei, MD, PhD. Medical Intelligence and Innovation Academy, Southern University of Science and Technology Hospital, 6019 Liuxian Road, Shenzhen 518055, China; School of Medicine, Southern University of Science and Technology, 1088 Xueyuan Road, Shenzhen 518055, China. Email: nfperry@163.com; Heng Li, PhD. Department of Computer Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Road, Shenzhen 518055, China. Email: lih3@sustech.edu.cn.

Background: Ultrasound is widely used for image-guided therapy (IGT) in many surgical fields, thanks to its various advantages, such as portability, lack of radiation and real-time imaging. This article presents the first attempt to utilize multiple deep learning algorithms in distal humeral cartilage segmentation for dynamic, volumetric ultrasound images employed in minimally invasive surgery.

Methods: The dataset, consisting 5,321 ultrasound images were collected from 12 healthy volunteers. These images were randomly split into training and validation sets in an 8:2 ratio. Based on deep learning algorithms, 9 semantic segmentation networks were developed and trained using our dataset at Southern University of Science and Technology Hospital in September 2022. The performance of the networks was evaluated based on their segmenting accuracy and processing efficiency. Furthermore, these networks were implemented in an IGT system to assess their feasibility in 3-dimentional imaging precision.

Results: In 2D segmentation, Medical Transformer (MedT) showed the highest accuracy result with a Dice score of 89.4%, however, the efficiency in processing images was relatively lower at 2.6 frames per second (FPS). In 3D imaging, the average root mean square (RMS) between ultrasound (US)-generated models based on the networks and magnetic resonance imaging (MRI)-generated models was no more than 1.12 mm.

Conclusions: The findings of this study indicate the technological feasibility of a novel method for real-time visualization of distal humeral cartilage. The increased precision of ultrasound calibration and segmentation are both important approaches to improve the accuracy of 3D imaging.

Keywords: Deep learning; image-guided therapy (IGT); minimally invasive surgery; distal humeral cartilage; ultrasound visualization

Submitted Jan 03, 2023. Accepted for publication May 26, 2023. Published online Jun 25, 2023.
doi: 10.21037/qims-23-9
View this article at: https://dx.doi.org/10.21037/qims-23-9

^ ORCID: 0000-0003-3652-9283.

Introduction

Thanks to the advancements in computer science and the enhanced accuracy of tracking systems and imaging methods, surgical procedures have entered a new era of precision and minimally invasiveness (1). Novel techniques, such as computer-assisted orthopedic surgery (CAOS) and image-guided therapy (IGT), can assist surgeons in performing more complex and challenging surgeries through smaller incisions. As a routine imaging device, ultrasound (US) is widely used for IGT in many surgical fields, thanks to its various advantages, such as portability, lack of radiation and real-time imaging (2,3). In orthopedic surgery, US is utilized to visualize the contours of bones and cartilage for precise navigation of surgical instruments (4). Despite its potential, US still faces limitations in clinical applications due to constraints of traditional image segmentation algorithms.

Convolutional neural networks (CNNs) represent powerful deep learning (DL) algorithms that have displayed extraordinary advancements in several medical imaging modalities, such as X-ray, US, computed tomography (CT), magnetic resonance imaging (MRI) and endoscopy (5). Several networks such as UNet, MAnet, PSPnet, and DeepLabV3+, have been proposed specifically for performing semantic segmentation on medical images, successfully achieving impressive results (6-9). However, the current accuracy of CNNs has not yet met the IGT requirement for imaging procedures. As a fundamental component of CNNs, the convolutional kernel selectively considers a specific subset of pixels in the image during each calculation iteration. The restriction compels the network to concentrate on local patterns, thereby limiting its capacity to comprehend the broader context and long-range dependencies in the input image (10). To address this issue, transformer-based architectures have been proposed recently, which incorporate self-attention mechanisms for encoding long-range dependencies. Among these, the Medical Transformer (MedT) was specifically designed for medical image databases owing to their unique characteristics. It has been found to outperform CNNs in terms of performance (10).

In our study, we trained automatic segmentation networks, namely MedT and 9 CNNs, for US images of the distal humeral cartilage, and subsequently compared their performance at a 2D level. Furthermore, we developed IGT systems based on these networks. To evaluate the effectiveness of the algorithm and a novel US visualization method for distal humeral cartilage, we compared US-generated models with MRI-generated models in terms of 3D imaging outcomes. The purpose of this study was to establish an effective algorithmic foundation and introduce a novel visualization method using ultrasound for distal humeral cartilage. We present this article in accordance with the TRIPOD reporting checklist (11,12) (available at https://qims.amegroups.com/article/view/10.21037/qims-23-9/rc).

Methods

Experimental flowchart

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study received approval from the Ethics Committee of the Southern University of Science and Technology Hospital (No. ECSUSTH-2022-064) and was carried out at the same hospital in September 2022. All volunteers signed written informed consent prior to participating. Initially, we constructed 9 semantic segmentation networks based on CNNs (UNet, UNet++, MAnet, Linknet, FPN, PAN, PSPnet, DeepLabV3 and DeepLabV3+) along with MedT to perform cartilage segmentation on US images. We evaluated the networks' performance using segmentation accuracy and processing efficiency metrics. Additionally, we implemented the trained networks in an IGT system. Upon scanning the elbow, 3D cartilage models were automatically generated. The US-generated model was compared to the MRI-generated model on the same sample to assess the 3D imaging precision of DL-based US visualization and evaluate the technological feasibility of this technique. The study flowchart is shown in *Figure 1*.

Comparative experiment involving segmentation networks

CNNs

Since its proposal in 1998, the LeNet5 network has paved the way for CNNs to become comprehensive system architectures for processing computer vision (CV) tasks. CNN typically consist of convolutional layers, pooling layers, nonlinear layers and fully connected layers. In 2015, UNet was introduced for medical image segmentation, featuring a symmetrical encoder-decoder structure (6). The pyramid scene parsing network (PSPNet) was proposed by Zhu *et al.* in 2016, which a pyramid pooling module to effectively aggregate the global context information obtained from different region-based contexts for scene

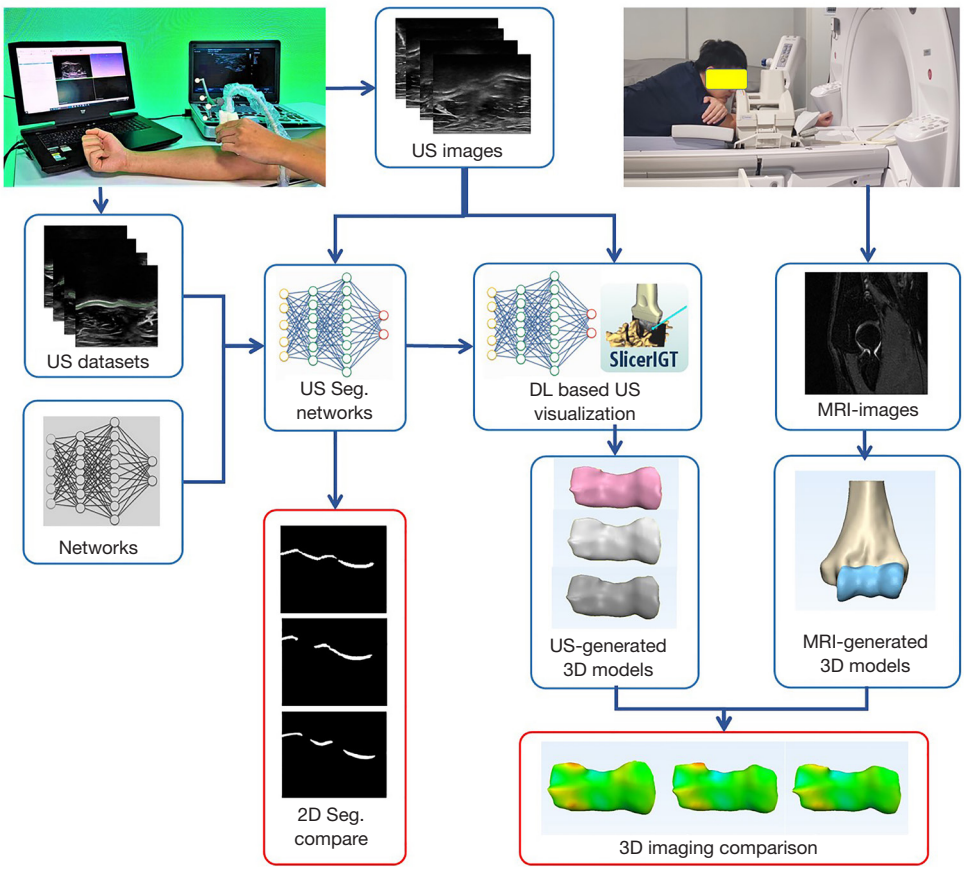

Figure 1 The flowchart of the study. US, ultrasound; DL, deep learning; MRI, magnetic resonance imaging; seg., segmentation. This image is published with the participant's consent.

parsing (13). The Google team introduced the DeepLab series, a suite of semantic segmentation algorithms, with DeepLabV3+ being the latest version proposed in 2018 (8). Compared to its predecessor, DeepLabV3+ retains the Atrous Spatial Pyramid Pooling (ASPP) structure, and introduces a reconstructed decoding structure to better capture object boundaries. The network also utilizes an improved Xception module as the backbone to reduce the number of required parameters. In 2020, Fan *et al.* proposed Manet, which introduces a self-attention mechanism allowing for the adaptive integration of local features with global dependencies in their network (9).

MedT

As a novel attention-driven building block network, the transformer was first introduced by Vaswani *et al.* for natural language processing (NLP) tasks (14). Due to its self-attention mechanism, a transformer has a strong ability to model long-range dependencies, which is why it demonstrates state-of-the-art performance on NLP tasks. For CV tasks, vision transformers (ViTs), which are built by cascading multiple transformer layers, interpret an image as a sequence of patches and process it in a way similar to NLP (15). Long-range dependencies are also significant for medical images as they can substantially enhance the efficiency of image segmentation. However, for appropriate training, many standard transformer-based networks proposed for semantic segmentation demand large-scale datasets, which are challenging to obtain in medical imaging scenarios. To solve this issue, Valanarasu *et al.* introduced MedT based on a gated axial-attention model, which adds an additional control mechanism in the self-attention architecture (10). Furthermore, a local-global training strategy was proposed to further enhance the segmentation performance of the model. By combining these two aspects of improvement, MedT achieved good performances on

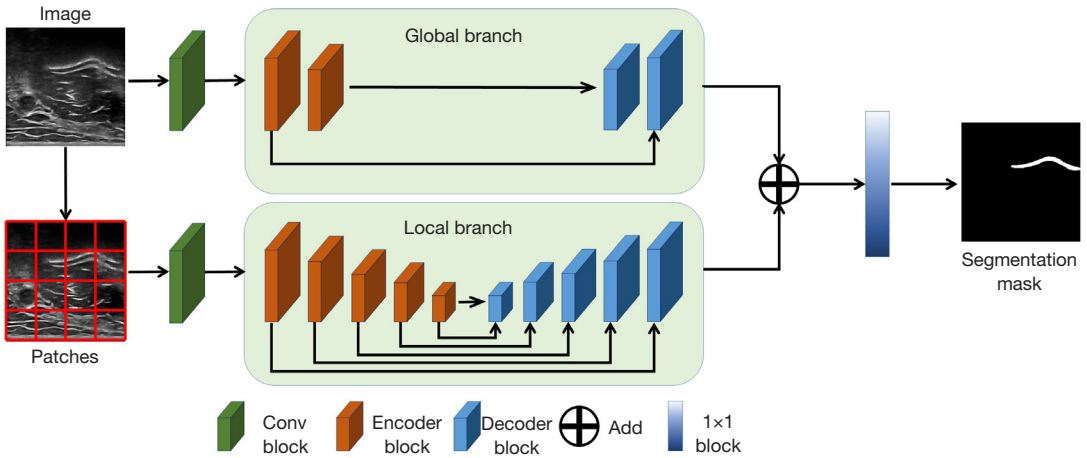

Figure 2 The network structure of Medical Transformer.

multiple medical image datasets. The network structure of MedT is displayed in *Figure 2*.

US data acquisition

There is currently no established standard for the number of samples necessary for training medical image semantic segmentation models. We took into account multiple factors when determining the sample size, such as the high distinctiveness of cartilage ultrasound images, minor variations in cartilage image features among healthy individuals, and the number of images needed for a single sample scan. A total of 5,321 2D US slices were collected from 12 volunteers (5 females, 7 males, mean age 33.3 years, range from 20 to 44 years) with healthy elbow cartilage (*Table 1*, *Figure 3*). The inclusion criteria included (I) 18–60 years old; (II) no elbow trauma and surgery history; (III) no elbow deformity and other pathological changes. Volunteers who did not meet the inclusion criteria were excluded (*Figure 3*). As the volar articular surface plays a crucial role in assessing surgical outcomes, particularly in cases of intra-articular fractures, the primary objective of this study was to use US images to visualize this surface. Therefore, all images were acquired with a US probe placed on the volar surface, vertically to the principal axis of the arm (*Figure 1*). Volunteers kept their arms straight during US scanning to maximize the exposure the cartilage of the distal humerus. All data collection was performed by one deputy chief orthopedic physician who had sufficient experience in elbow ultrasonography, using a US system (Mindray M9 ultrasound system, Mindray Bio-Medical Electronics Co., Ltd., Shenzhen, Guangdong, China) and a 2D US probe (Mindray L14-6Ns, Mindray Bio-Medical Electronics Co., Ltd.). The workstation settings were optimized by a US specialist for elbow cartilage structure visualization: a 12.6-MHz probe frequency, a 3.5-cm penetration depth, a dynamic range of 110 dB and a gain of 45 dB.

Data annotations were provided by an orthopedic surgeon who had ample experience in elbow ultrasonography. The cartilage contours were outlined on all the US images using 3D Slicer application (16). Each frame of the US images and their masks were resized to 768×768 and normalized by linearly scaling their gray level intensities to (0,1). The datasets were randomly split, allocating 80% of the data for network training and 20% used for validation.

Experimental environment

The experiment in this study was conducted using a computer platform consisting an Intel i7-11800 CPU and an NVIDIA A100 tensor core GPU. The PyTorch library was utilized for development of the segmentation networks. To enhance the size of the datasets, the augmentation was implemented by a batch generator, which includes random rotations, random flips, Gaussian noise addition, blurring and contrast-limited adaptive histogram equalization (CLAHE). The networks were trained with the Adam optimizer (learning rate =0.001, $\beta1$=0.5, $\beta2$=0.999, learning decay =0.00003) and a weighted binary cross-entropy loss function was used.

Evaluation metrics

For the task of segmenting cartilage in the US images, the samples were divided into two categories: cartilage and

Table 1 Information of volunteers: 2D US slices were collected from 12 volunteers, 5 new samples were chosen to scan MRI and US

ID	Sex	Age (year)	Weight (kg)	Height (cm)	Side	Slices
Volunteer 1	Male	26	81	184	L, R	431
Volunteer 2	Male	25	71	172	L, R	495
Volunteer 3	Female	38	52	161	L, R	421
Volunteer 4	Male	44	75	176	L, R	487
Volunteer 5	Female	31	53	165	L, R	450
Volunteer 6	Female	32	63	168	L, R	389
Volunteer 7	Male	37	77	178	L, R	444
Volunteer 8	Male	42	86	179	L, R	427
Volunteer 9	Male	20	78	175	L, R	391
Volunteer 10	Female	25	60	164	L, R	472
Volunteer 11	Female	39	55	162	L, R	502
Volunteer 12	Male	40	74	170	L, R	412
Sample 1	Male	46	75	180	R	560
Sample 2	Female	22	67	172	L	584
Sample 3	Male	31	62	169	L	569
Sample 4	Female	31	68	174	R	512
Sample 5	Male	37	85	182	L	599

2D, 2-dimensional; US, ultrasound; MRI, magnetic resonance imaging; ID, identity; L, left; R, right.

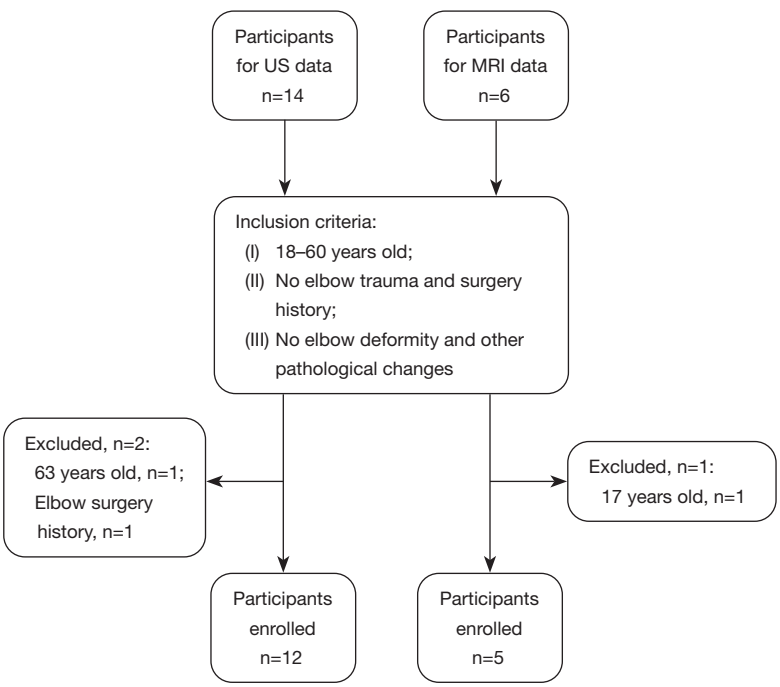

Figure 3 The flow chart showing the process of participants' selection. US, ultrasound; MRI, magnetic resonance imaging.

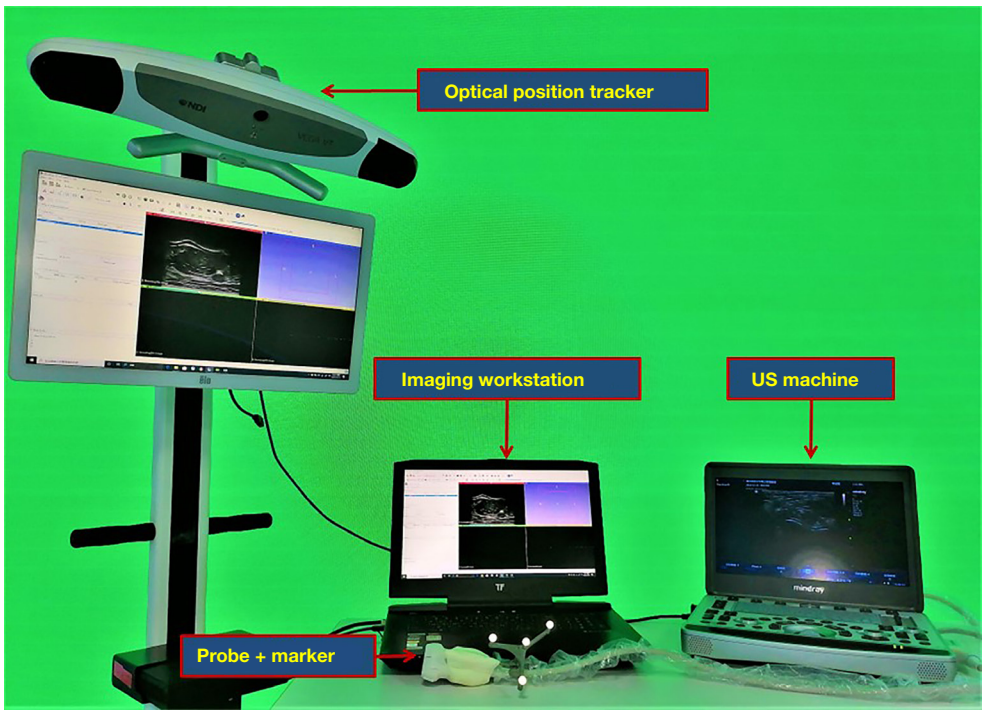

Figure 4 IGT system. US, ultrasound; IGT, image-guided therapy.

non-cartilage. By comparing the results of the segmentation networks with those of manual segmentation by an expert, four situations were identified: true positives (TPs), which are cartilage pixels that were correctly predicted; false positives (FPs), which are non-cartilage pixels that were mistakenly predicted as cartilage; true negatives (TNs), which are non-cartilage pixels that were correctly predicted; and false negatives (FNs), which are cartilage pixels that were mistakenly predicted as non-cartilage. To quantitatively assess the accuracy of the segmentation networks, two metrics were used: the intersection over union (IoU) and the Dice similarity coefficient (Dice). The calculation methods are shown in Eqs. [1] and [2], respectively.

$$IoU = \frac{TP}{TP + FP + FN} \quad [1]$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad [2]$$

The aim of this research was to develop a network that can optimize the balance between accuracy and efficiency. Efficiency was measured using three metrics: the number of parameters (Params), which gauges the spatial complexity of the network; the number of floating-point operations (FLOPs), which reflect the time complexity of the network; and the number of frames per second (FPS), which represents the inference speed. These metrics were calculated using the Python thop library.

Precision experiment involving the DL-based IGT system

3D US image calibration

The IGT system utilized an US machine (Mindray M9 ultrasound system, Mindray Bio-Medical Electronics Co., Ltd., Shenzhen, Guangdong, China), an optical position tracker (NDI Polaris vega XT, Northern Digital Inc., Waterloo, Ontario, Canada), and a computer with IGT software (3D Slicer with the SlicerIGT extension) (17) (*Figure 4*). The PLUS software tool acted as an intermediary between the hard devices (18). Spatial and temporal calibration of the 2D US images and pose tracking data were performed using the free-hand calibration (fCal) application (18). The US image coordinates were referred to as "UI", while the NDI optical reference frame coordinate system attached at the end of the probe was defined as "UR". The matrix transformation from UR to UI was T_{UR}^{UI}, which

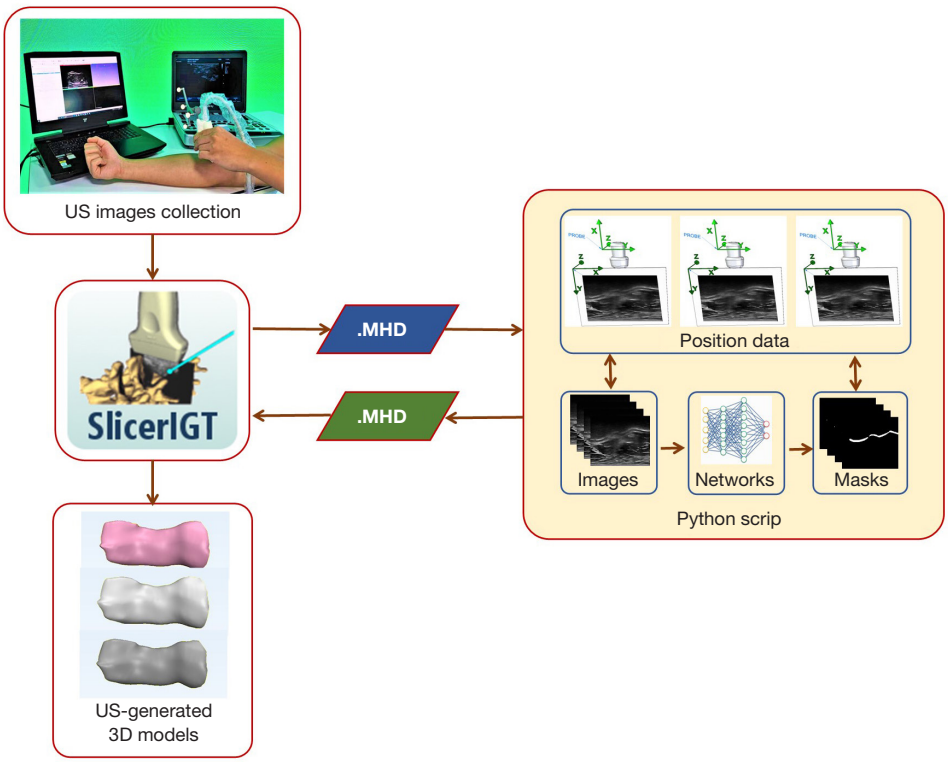

Figure 5 Deep learning based ultrasound visualization pipeline. US, ultrasound; .MHD, a .MHD file.

enabled any point P_{UI} in the US image coordinate system to be converted to P_{UR} in the UR coordinate system. The governing calculation is provided in Eq. [3]:

$$P_{UR} = T_{UR}^{UI} \times P_{UI} \quad [3]$$

After completing this step, the spatial position of the 2D ultrasound image could be determined, allowing for the creation of articular cartilage models in the subsequent stages.

Establishment of a DL-based US visualization pipeline

After a volunteer was scanned with the IGT system, the US images and the spatial position data were gathered into a .MHD file. The collected data underwent multiple processing steps, as depicted in *Figure 5*. Initially, the US image data were automatically segmented by the networks, generating 2D contours of the cartilage. Afterward, the 2D contours of the cartilage replaced the original US images and were integrated with the spatial position data to produce a new .MHD file. These two steps were programmed in a Python script utilizing the Insight Toolkit (ITK) library. Finally, the new .MHD file was imported back into the IGT system, and US-generated cartilage models were constructed.

MRI data acquisition

To assess the generalization and reliability of the networks, five new volunteers (2 females, 3 males, mean age 33.4, range from 22 to 46), with the same inclusion and exclusion criteria, were recruited for this stage of the study (*Table 1*, *Figure 3*). MRI images were captured of their elbows using a 3.0T clinical MRI scanner (GE Discovery MR750, GE Medical Systems, Milwaukee, WI, USA) with a 2D sequence, fat suppression and a knee-dedicated coil (8 channels), as referred in *Figure 1*. The sequence parameters included a repetition time of 11.89 ms, an echo time of 5.3 ms, a 1-mm section thickness, a 160- to 180-mm field of view, a base resolution of 384, and a 95% phase resolution. The images are obtained in the sagittal plane. Additionally, 3D cartilage models of each distal humerus were constructed using a 3D slicer and exported as stereolithography (STL) files. The five volunteers were also scanned by the IGT system to render US-generated

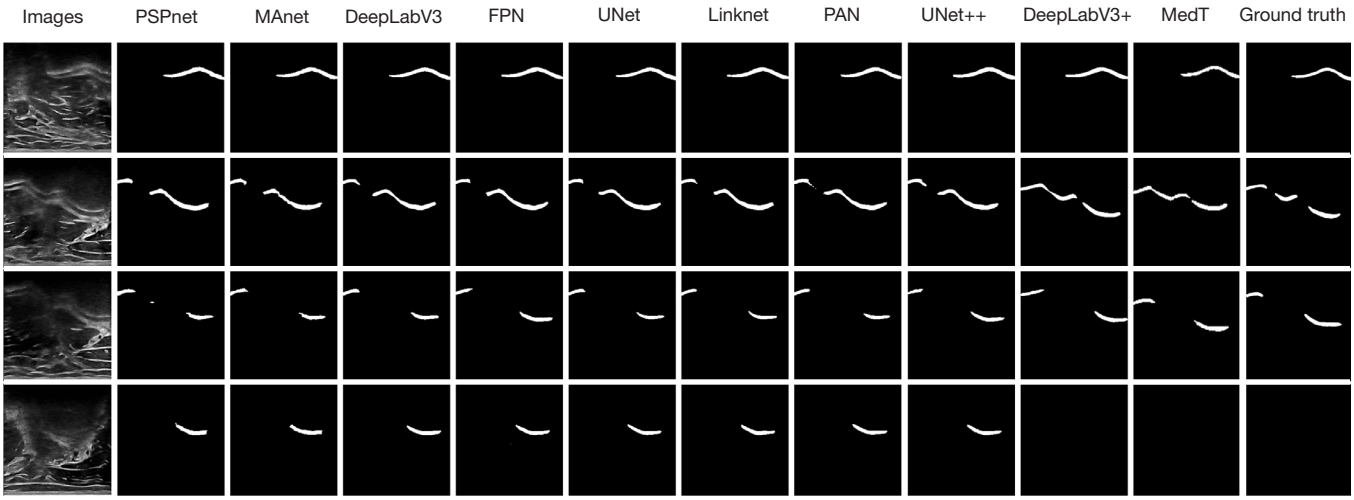

Figure 6 The results of 10 segmentation networks. PSPnet, MAnet, DeepLabV3, FPN, UNet, Linknet, PAN, UNet++, DeepLabV3+ are different neural networks based on convolutional neural network. MedT, Medical Transformer.

models. The MRI data collection process was performed by the same doctor who did US data annotations with the assistance of a radiologist.

Evaluation metrics

At present, MRI is considered the gold standard in diagnosis articular cartilage injuries and for imaging this structure. Therefore, to determine the clinical significance of the US-based visualization method, the cartilage models produced using IGT systems based on different segmentation networks were compared with MRI-generated models. The Euclidean point-to-mesh distance (EPD) was utilized to measure the US visualization error by calculating the distance between all nodes of the MRI-generated model and the surfaces of the US-generated model. Subsequently, the 3D imaging performance of the DL algorithms was evaluated by comparing the EPDs of the models rendered by IGT systems based on different segmentation networks.

Statistical analysis

The mean, standard deviation (SD), root mean square (RMS), Q1, median, Q3 and range of the 3D imaging error were calculated, and Tukey boxplots were employed for graphical visualization purposes. The nonparametric Friedman rank-sum test was applied to analyze the statistical significance of the results (19), While a post hoc analysis was performed with the Wilcoxon signed-rank test (20). The significance level was set as "$P<0.05$". All data analysis was done using SPSS 21.0 (Chicago, IL, USA).

Results

Segmentation task study

Figure 6 shows the segmentation results for 10 networks, showcasing the superior performance of the MedT network, which closely resembles that of the expert. The algorithmic edge segmentation by MedT is smoother and more seamless than that of other CNNs.

Comparing the performance of MedT with other CNNs, *Table 2* shows that MedT produced the best segmentation results, with an IoU score of 78.6% and a Dice score of 89.4%. However, MedT's inference speed was significantly slower than that of the CNNs, with an FPS of 2.6. *Figure* 7 illustrates the trade-off between segmentation accuracy and speed, With DeepLabV3+ yielding the best performance.

IGT imaging study

The fiducial registration error (FRE) for 3D US calibration was found to be 1.3±0.48 mm. 5 networks (PSPnet with an IoU of 74.7%, MAnet with an IoU of 75.7%, UNet with an IoU of 76.8%, DeepLabV3+ with an IoU of 77.4% and MedT with an IoU of 78.6%) were selected to build 3D models from the cartilage US images. *Figure 8*, *Figure 9* and *Table 3* show the 3D imaging errors between the cartilage US models generated by each network and the MRI-generated models. For networks with higher IoU scores, the RMS of the errors was generally smaller. The average RMS between US-generated models and MRI-models is no

Table 2 Comparison of segmentation performance of MedT and CNNs

Networks	Backbone	Image size (mm)	IoU (%)	Dice (%)	Param (M)	FLOP (G)	FPS
PSPnet	Resnet101	768*768	74.7	85.8	2.2	25.7	35.8
MAnet	Resnet101	768*768	75.7	86.6	166.4	210.8	21.1
DeepLabV3	Resnet101	768*768	75.7	86.8	58.6	543.9	14.9
FPN	Resnet101	768*768	76.3	87.4	45.1	113.6	28.9
UNet	Resnet101	768*768	76.8	87.5	51.5	139.3	28
Linknet	Resnet101	768*768	76.7	87.8	50.2	140	25.5
PAN	Resnet101	768*768	77.2	88.1	43.2	121.5	28.7
UNet++	Resnet101	768*768	77.1	88.2	68	561	15.1
DeepLabV3+	Resnet101	768*768	77.4	88.5	45.7	126	30.2
MedT	Transformer	768*768	78.6	89.4	48	186.4	2.6

MedT, Medical Transformer; CNN, convolutional neural network; IoU, intersection over union; FLOP, floating point operations per second; FPS, frames per second.

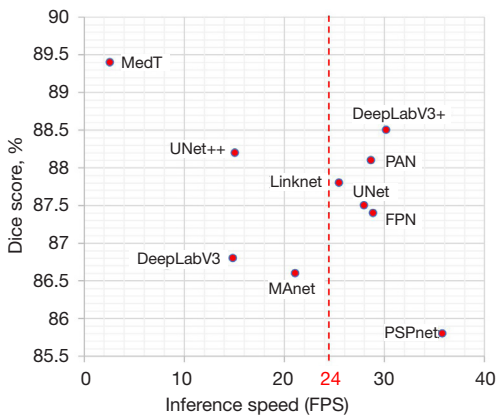

Figure 7 The trade-off between segmentation accuracy and inference speed. MedT, Medical Transformer; FPS, frames per second.

more than 1.12 mm.

The Friedman rank-sum test revealed the P values of the 5 samples' cartilage imaging errors (*Table 3*). The Wilcoxon signed-rank test demonstrated P values between two networks in the same sample, indicating that the larger the IoU gap between two networks, the less P value were obtained. For instance, the P values between the imaging errors of MedT and PSPnet (with an IoU gap of 3.9%) presented in 5 samples are "P=0.012", "P=0.012", "P=0.017", "P=0.012", "P=0.017". The P values between the imaging errors of MAnet and UNet (with an IoU gap of 1.1%) presented in each sample are "P=0.123", "P=0.574", "P=0.459", "P=0.491", "P=0.799".

Discussion

Visualizing the distal humeral articular structure in real-time, using intraoperative images, can assist doctors during minimally invasive surgical procedures, such as minimally invasive plate osteosynthesis. This article presents the first attempt to use multiple DL algorithms for distal humeral cartilage segmentation in dynamic, volumetric US images for IGT of the distal humerus.

US is an especially appealing modality for intraoperative imaging due to its ability to provide real-time images, lack of ionizing radiation, and lower cost compared to other popular modalities, like MRI and CT. Recently, DL technology has made significant advancements in US image segmentation, enabling real-time visualization of the articular surface during intraoperative US (4). In this study, we compared the segmentation performances of CNNs and MedT, as well as the 3D imaging results obtained when these algorithms were integrated with an IGT system. It was deemed feasible from a technological standpoint to use such a visualization method.

Automatic medical image segmentation is a useful tool for clinical diagnosis and treatment of diseases. In recent years, CNN-based neural networks have shown remarkable accuracy in many fields of medical imaging, comparable to that of clinical experts and medical doctors. Yang *et al.*

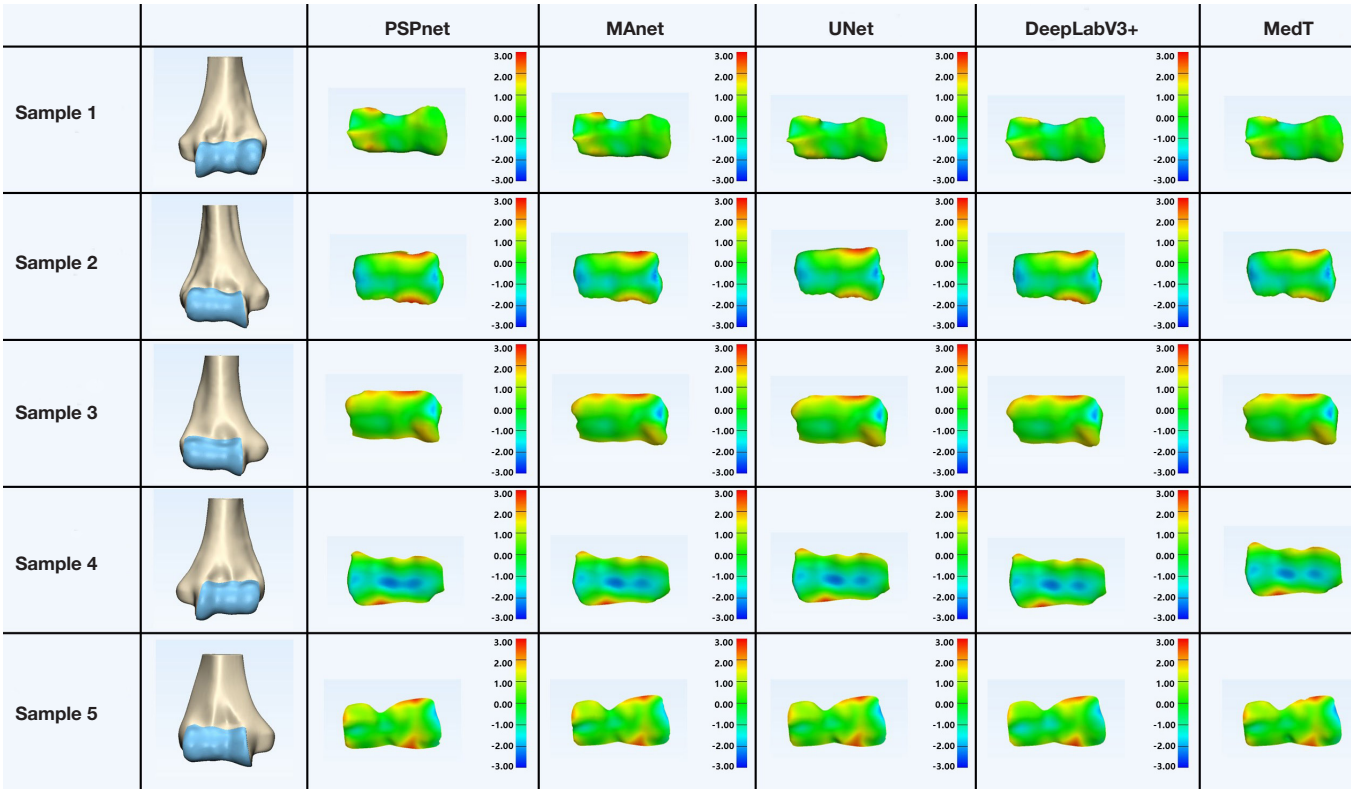

Figure 8 The heat map of 3-dimensional imaging errors. PSPnet, MAnet, UNet, DeepLabV3+ are different neural networks based on convolutional neural network. MedT, Medical Transformer.

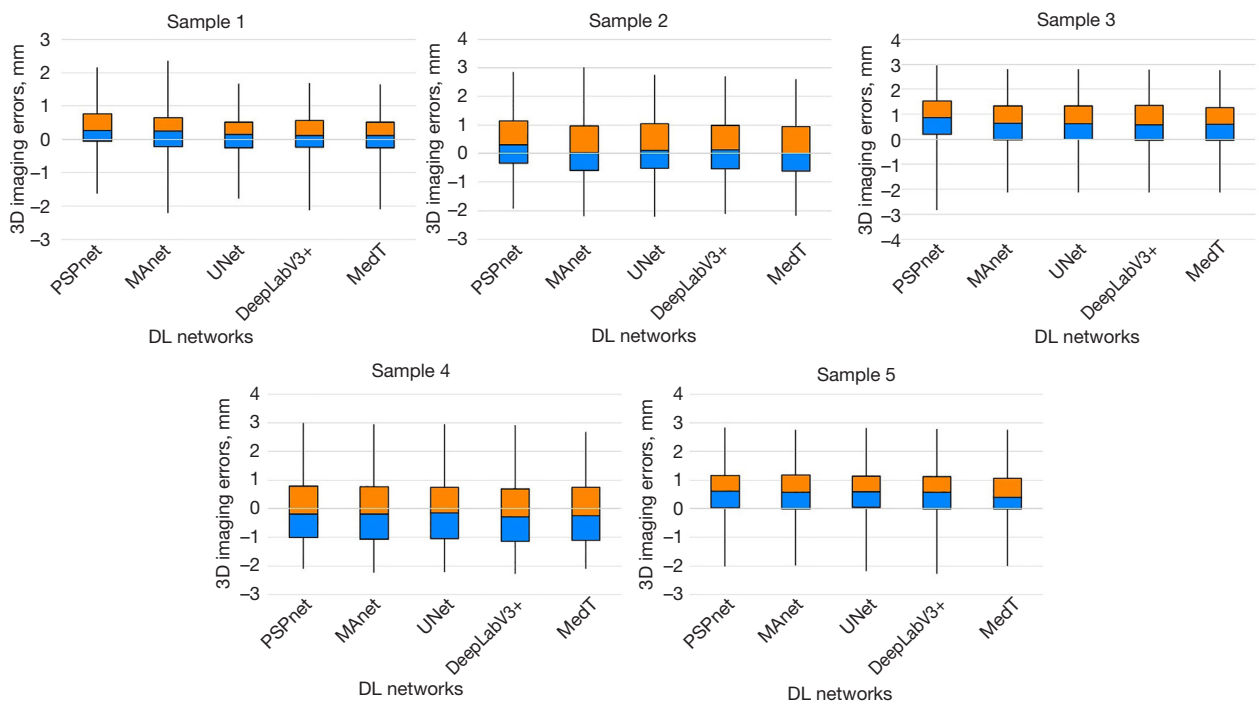

Figure 9 The box plot of 3-dimensional imaging errors. DL, deep learning; MedT, Medical Transformer.

Table 3 Comparison of 3D imaging errors between the US-generated models based on each network and the MRI-generated models

Networks	Imaging errors							Friedman test	Wilcoxon signed-rank test (P≤0.05†)			
	Range (mm)	Mean (mm)	SD (mm)	RMS (mm)	Q1 (mm)	Median (mm)	Q3 (mm)		MAnet	UNet	DeepLabV3+	MedT
Sample 1												
PSPnet	−1.62 to 2.15	0.33	0.66	0.73	−0.05	0.27	0.76	Chi-square, 21.5; P<0.001	0.262	0.012†	0.012†	0.012†
MAnet	−2.21 to 2.35	0.27	0.69	0.71	−0.22	0.25	0.64			0.123	0.068	0.05†
UNet	−1.77 to 1.67	0.12	0.65	0.66	−0.25	0.14	0.52				0.734	0.041†
DeepLabV3+	−2.12 to 1.68	0.15	0.64	0.66	−0.23	0.12	0.56					0.207
MedT	−2.09 to 1.65	0.11	0.65	0.66	−0.25	0.12	0.51					
Average	−1.96 to 1.9	0.2	0.66	0.68	−0.2	0.18	0.6					
Sample 2												
PSPnet	−1.91 to 2.85	0.4	1.04	1.11	−0.35	0.31	1.14	Chi-square, 18.5; P=0.001	0.068	0.017†	0.012†	0.012†
MAnet	−2.20 to 3.00	0.20	1.08	1.1	−0.6	0.02	0.96			0.574	0.778	0.035†
UNet	−2.21 to 2.76	0.26	1.05	1.09	−0.52	0.11	1.04				0.231	0.017†
DeepLabV3+	−2.12 to 2.69	0.21	1.02	1.04	−0.54	0.13	0.99					0.012†
MedT	−2.18 to 2.6	0.13	0.98	0.98	−0.61	0	0.94					
Average	−2.12 to 2.78	0.24	1.03	1.06	−0.52	0.11	1.01					
Sample 3												
PSPnet	−1.84 to 2.84	0.72	0.84	1.11	0.08	0.75	1.41	Chi-square, 15.1; P=0.005	0.035†	0.017†	0.03†	0.017†
MAnet	−2.12 to 2.81	0.63	0.89	1.09	−0.02	0.63	1.32			0.459	0.258	0.018†
UNet	−2.12 to 2.8	0.63	0.86	1.06	0	0.62	1.34				0.439	0.033†
DeepLabV3+	−2.13 to 2.77	0.62	0.89	1.08	−0.05	0.57	1.36					0.258
MedT	−2.12 to 2.76	0.59	0.87	1.05	−0.04	0.6	1.26					
Average	−2.01 to 2.8	0.64	0.87	1.08	−0.01	0.63	1.34					
Sample 4												
PSPnet	−2.09 to 2.98	−0.07	1.18	1.19	−1.01	−0.18	0.78	Chi-square, 18.8; P=0.001	0.018†	0.04†	0.012†	0.012†
MAnet	−2.24 to 2.96	−0.07	1.11	1.12	−1.06	−0.19	0.76			0.491	0.123	0.049†
UNet	−2.21 to 2.95	−0.08	1.11	1.11	−1.04	−0.15	0.75				0.176	0.068
DeepLabV3+	−2.28 to 2.91	−0.14	1.15	1.16	−1.15	−0.28	0.7					0.888
MedT	−2.1 to 2.67	−0.11	1.01	1.02	−1.1	−0.24	0.75					
Average	−2.18 to 2.89	−0.09	1.11	1.12	−1.07	−0.21	0.75					
Sample 5												
PSPnet	−2.02 to 2.84	0.58	0.96	1.08	0.03	0.61	1.16	Chi-square, 18.8; P=0.001	0.16	0.02†	0.011†	0.017†
MAnet	−1.97 to 2.76	0.55	0.91	1.07	−0.01	0.57	1.17			0.799	0.088	0.011†
UNet	−2.18 to 2.81	0.55	0.9	1.06	0.04	0.6	1.13				0.018†	0.123
DeepLabV3+	−2.27 to 2.77	0.49	0.92	1.04	−0.02	0.57	1.11					0.233
MedT	−2.00 to 2.75	0.41	0.90	1.02	−0.02	0.4	1.06					
Average	−2.09 to 2.79	0.52	0.92	1.05	0	0.55	1.13					

†, P≤0.05. PSPnet, MAnet, UNet, DeepLabV3+ are different neural networks based on convolutional neural network. 3D, 3-dimensional; US, ultrasound; MRI, magnetic resonance imaging; MedT, Medical Transformer; SD, standard deviation; RMS, root mean square.

developed a CNN network for segmenting brain low-grade glioma MRI image and found that CNN outperforms Support Vector Machine (SVM), a wildly used machine learning method, in multiple datasets (21). Wang *et al.* proposed an automatic gastric cancer segmentation model based on DeepLabV3+ (8), and compared it with other CNNs on gastric cancer pathological slice images, finding that DeepLabV3+ had better accuracy, with a Dice score of 91.66%. Similarly, Yan *et al.* and Zhu *et al.* employed PSPnet for prostate MRI and coronary angiography images (7,13), and found it to be more accurate than the baseline network (UNet). For assessing the locations and extents of liver tumors (9), Fan *et al.* proposed the use of Manet, which outperformed other state-of-the-art methods in a public dataset (MICCAI 2017 LiTS Challenge), obtaining a Dice score of 74.9%. However, in our segmentation task, PSPnet and MAnet did not perform well among all networks, with Dice scores of 85.8% and 86.6%, respectively, and lower accuracy than that of UNet. The best CNN network outcome was DeepLabV3+, second only to MedT, with a Dice score of 88.5%.

Transformers, which have shown excellent results in NLP tasks, have recently been applied to CV problems with great success. Their ability to capture the global context of an image is their most significant advantage over traditional CNNs with local receptive fields. The medical imaging field has thus witnessed a growing interest in transformers. Wang *et al.* proposed a boundary-aware transformer (BAT) to improve skin lesion segmentation tasks, achieving Dice scores of 92.1% and 91.2% on two public datasets (ISIC 2016+PH2 and ISIC 2018), respectively (22). For cardiac image segmentation tasks, Deng *et al.* proposed TransBridge, a lightweight parameter-efficient hybrid model consisting of transformers and CNN-based encoder-decoder structures for ventricle segmentation in echocardiography (23). Liang *et al.* proposed a U-shaped network, named TransConver, which combines CNN and transformer for segmenting brain tumors in MRI images. Their network achieved the highest Dice scores of 83.73% and 86.32% on BrasTS2019and BraTS2018 datasets, respectively (24). MedT, which was first proposed by Valanarasu *et al.* (10), was modified with a gated axial attention layer and a local-global training strategy for boosting the segmentation performance in different datasets, including a brain US image dataset. In our cartilage segmentation study involving US images, MedT demonstrated outstanding accuracy, although its inference speed (an FPS of 2.6) is slower than the required real-time imaging rate of at least 24 FPS (25). This low speed is mainly due to the computational complexity of the transformer network itself. Nevertheless, the sluggish inference speed of MedT should not be a major drawback of this technology. This is because image-guided fracture reductions primarily rely on the pre- and post-reduction images, and any shortcomings in speed can be overcome by upgrading the hardware of the image workstation.

CNNs are the primary means for guidance in minimally invasive knee surgery (4,26,27). Kompella *et al.* employed Mask R-CNN to segment knee cartilage in ultrasound images (26). Preprocessing the images and pretraining the network using the COCO 2016 image dataset yielded the best result, with an average Dice score of 80%. Dunnhofer *et al.* proposed Siam-U-Net, which merged UNet and the Siamese framework to improve the segmentation performance of femoral cartilage in US images (27). Compared to traditional UNet, Siam-U-Net achieved an average Dice score improvement from 64% to 70%. Antico *et al.* employed UNet to automatically segment femoral cartilage in US images (4), And proposed a novel metric named the Dice coefficient with boundary uncertainty to address intraobserver variability in manually labeled cartilage boundaries due to the inherent properties of US images. The revised Dice score of UNet was 87%, even higher than that of an expert (78%). In another study, Antico *et al.* presented the application of a Bayesian CNN based on Monte Carlo dropout to segment cartilage by contouring it on either US or MRI images, then projecting it onto the corresponding US volume (28). The authors also proposed a novel approach to evaluate model performance involving probabilistic ground-truth annotations generated from registered US and MRI volumes. With these two modifications, the authors obtained better outcomes than traditional UNet, with a Dice score increase from 6% to 8%.

With the aid of navigation tools and computer-assisted surgery devices, the IGT technique has been developed to enhance the targeting and localization of diseased tissue, thus improving minimally invasive surgery. IGT highly relies on image segmentation, especially for IGT based on intraoperative US imaging (29). Combining IGT with a DL segmentation algorithm has the potential to broaden its application scope. Hu *et al.* developed a navigation approach for breast-conserving surgery, using a real-time automatic UNet-based tumor contouring process for intraoperative guidance (30). The UNet achieved an average Dice score of 78%, significantly improving the efficiency of breast-conserving surgery navigation systems. Ungi *et al.*

proposed an automatic US segmentation method for 3D spine visualization and scoliosis measurement, to address the difficulties in US usage for spine imaging (31). Their method constructed 3D volumes with a maximum error of 2.2° compared to X-ray results. In our study, we constructed 3D cartilage models by automatically segmenting US contours using IGT technology. The error of the 3D models was the metric used to evaluate the DL algorithm's performance. Through statistical analysis, we found that the larger the IoU difference between two networks was, the more significant their imaging errors. Although MedT had better accuracy on 2D images, the difference between the 3D imaging errors of MedT and DeepLabV3+ did not demonstrate statistical significance in most samples. The primary reason for this situation was likely that 3D imaging errors consist of 2D segmentation errors and the FRE of US calibration. The average FRE in this study was 1.3 mm, which could account for some relatively low segmentation errors. Therefore, decreasing the FRE should also be an essential approach to improve the accuracy of 3D imaging.

In this research, we utilized a combination of the DL algorithm and IGT technique to develop a novel visualization method for distal humerus cartilage. We compared MedT and several CNN-based segmentation networks to assess their imaging accuracy at the 2D and 3D levels. However, this study had several limitations. Firstly, the datasets used for training and testing the networks consisted only of healthy samples while ignoring fracture patients. It is worth noting that distal humeral cartilage segmentation might become significantly more challenging when a fracture is present, given the existence of fracture fragments and lipohemarthrosis. Therefore, testing these segmentation networks on patients with distal humerus fractures will be the focus of the next logical study. Secondly, the sample size used for testing was relatively small; hence, a larger sample is required to fully verify the performance of this visualization method. Lastly, due to a large FRE of US calibration, it had a considerably adverse impact on the 3D imaging results.

Conclusions

The aim of this study was to develop a novel and practical method for visualizing distal humeral cartilage using intraoperative US imaging. A total of 10 networks, including MedT and 9 CNNs, were evaluated and compared for both 2D segmentation and 3D imaging tasks. In terms of 2D segmentation, MedT demonstrated greater accuracy, but it required a more powerful GPU to improve inference speed. Among the 10 networks, DeepLabV3+ achieved the best trade-off between accuracy and inference speed. Regarding 3D imaging, the average RMS between US-generated models based on the networks and MRI models is no greater than 1.12 mm. However, due to the influence of the US calibration error, networks with small differences in 2D segmentation accuracy did not show much distinction. These findings suggest that this method is technologically feasible for visualizing distal humeral cartilage in real time. Nevertheless, additional experiments are needed to further assess its clinical feasibility.

Acknowledgments

Funding: This study was supported by Shenzhen Science and Technology Program (CN) (No. SGDX20211123114204007); Basic and Applied Fundamental Research Foundation of Guangdong Province (No. 2020A1515110286); China Postdoctoral Science Foundation, No. 72 General Fund (No. 2022M721474); Research Startup Fund of the Southern University of Science and Technology (No. Y01416214).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-9/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-9/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study received approval from the Ethics Committee of the Southern University of Science and Technology Hospital (No. ECSUSTH-2022-064). All volunteers signed written informed consent prior to participating.

References

1. Murphy SB, Ecker TM, Tannast M. THA performed using conventional and navigated tissue-preserving techniques. Clin Orthop Relat Res 2006;453:160-7.
2. Welch JN, Johnson JA, Bax MR, Badr R, Shahidi R. A real-time freehand 3D ultrasound system for image-guided surgery. In 2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium 2000;2:1601-4.
3. Abu Anas EM, Seitel A, Rasoulian A, St John P, Pichora D, Darras K, Wilson D, Lessoway VA, Hacihaliloglu I, Mousavi P, Rohling R, Abolmaesumi P. Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. Int J Comput Assist Radiol Surg 2015;10:959-69.
4. Antico M, Sasazawa F, Dunnhofer M, Camps SM, Jaiprakash AT, Pandey AK, Crawford R, Carneiro G, Fontanarosa D. Deep Learning-Based Femoral Cartilage Automatic Segmentation in Ultrasound Imaging for Guidance in Robotic Knee Arthroscopy. Ultrasound Med Biol 2020;46:422-35.
5. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
6. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Proceedings, Part III 18:234-41
7. Yan L, Liu D, Xiang Q, Luo Y, Wang T, Wu D, Chen H, Zhang Y, Li Q. PSP net-based automatic segmentation network model for prostate magnetic resonance imaging. Comput Methods Programs Biomed 2021;207:106211.
8. Wang J, Liu X. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. Comput Methods Programs Biomed 2021;207:106210.
9. Fan T, Wang G, Li Y, Wang H. Ma-net: A multi-scale attention network for liver and tumor segmentation. IEEE Access 2020;8:179656-65.
10. Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM. Medical transformer: Gated axial-attention for medical image segmentation. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2021: 24th International Conference, Proceedings, Part I 24:36-46.
11. Collins GS, Reitsma JB, Altman DG, Moons, KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Annals of Internal Medicine 2015;162:55-63.
12. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.
13. Zhu X, Cheng Z, Wang S, Chen X, Lu G. Coronary angiography image segmentation based on PSPNet. Comput Methods Programs Biomed 2021;200:105897.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in Neural Information Processing Systems 2017;30.
15. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010 2020;11929.
16. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323-41.
17. Ungi T, Lasso A, Fichtinger G. Open-source platforms for navigated image-guided interventions. Med Image Anal 2016;33:181-6.
18. Lasso A, Heffter T, Rankin A, Pinter C, Ungi T, Fichtinger G. PLUS: open-source toolkit for ultrasound-guided intervention systems. IEEE Trans Biomed Eng 2014;61:2527-37.
19. Eisinga R, Heskes T, Pelzer B, Te Grotenhuis M. Exact p-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers. BMC Bioinformatics 2017;18:68.
20. Woolson RF. Wilcoxon signed-rank test. Wiley Encyclopedia of Clinical Trials 2007:1-3.
21. Yang Q, Zhang H, Xia J, Zhang X. Evaluation of magnetic

resonance image segmentation in brain low-grade gliomas using support vector machine and convolutional neural network. Quant Imaging Med Surg 2021;11:300-16.
22. Wang J, Wei L, Wang L, Zhou Q, Zhu L, Qin J. Boundary-aware transformers for skin lesion segmentation. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2021: 24th International Conference, Proceedings, Part I 24:206-216.
23. Deng K, Meng Y, Gao D, Bridge J, Shen Y, Lip G, Zhao Y, Zheng Y. Transbridge: A lightweight transformer for left ventricle segmentation in echocardiography.In Simplifying Medical Ultrasound: Second International Workshop, ASMUS 2021, Held in Conjunction with MICCAI 2021, Proceedings 2:63-72.
24. Liang J, Yang C, Zeng M, Wang X. TransConver: transformer and convolution parallel network for developing automatic brain tumor segmentation in MRI images. Quant Imaging Med Surg 2022;12:2397-415.
25. Yu C, Gao C, Wang J, Yu G, Shen C, Sang N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Compputer Vision 2021;129:3051-3068.
26. Kompella G, Antico M, Sasazawa F, Jeevakala S, Ram K, Fontanarosa D, Pandey AK, Sivaprakasam M. Segmentation of Femoral Cartilage from Knee Ultrasound Images Using Mask R-CNN. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:966-9.
27. Dunnhofer M, Antico M, Sasazawa F, Takeda Y, Camps S, Martinel N, Micheloni C, Carneiro G, Fontanarosa D. Siam-U-Net: encoder-decoder siamese network for knee cartilage tracking in ultrasound images. Med Image Anal 2020;60:101631.
28. Antico M, Sasazawa F, Takeda Y, Jaiprakash AT, Wille M-L, Pandey AK, Crawford R, Carneiro G, Fontanarosa D. Bayesian CNN for segmentation uncertainty inference on 4D ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy. IEEE Access 2020;8:223961-223975.
29. Kapur T, Egger J, Jayender J, Toews M, Wells WM. Registration and segmentation for image-guided therapy. Intraoperative Imaging and Image-Guided Therapy 2014:79-91.
30. Hu Z, Nasute Fauerbach PV, Yeung C, Ungi T, Rudan J, Engel CJ, Mousavi P, Fichtinger G, Jabs D. Real-time automatic tumor segmentation for ultrasound-guided breast-conserving surgery navigation. Int J Comput Assist Radiol Surg 2022;17:1663-72.
31. Ungi T, Greer H, Sunderland KR, Wu V, Baum ZMC, Schlenger C, Oetgen M, Cleary K, Aylward SR, Fichtinger G. Automatic Spine Ultrasound Segmentation for Scoliosis Visualization and Measurement. IEEE Trans Biomed Eng 2020;67:3234-41.

Cite this article as: Zhao W, Su X, Guo Y, Li H, Basnet S, Chen J, Yang Z, Zhong R, Liu J, Chui EC, Pei G, Li H. Deep learning based ultrasonic visualization of distal humeral cartilage for image-guided therapy: a pilot validation study. Quant Imaging Med Surg 2023;13(8):5306-5320. doi: 10.21037/qims-23-9