



Article

Shoulder Bone Segmentation with DeepLab and U-Net

Michael Carl ¹, Kaustubh Lall ², Darren Pai ³, Eric Y. Chang ^{4,5} , Sheronda Statum ^{4,5}, Anja Brau ¹, Christine B. Chung ^{4,5}, Maggie Fung ¹ and Won C. Bae ^{4,5,*}

¹ General Electric Healthcare, Menlo Park, CA 94025, USA

² Department of Electrical and Computer Engineering, University of California-San Diego, San Diego, CA 92161, USA

³ Canyon Crest Academy, San Diego, CA 92130, USA

⁴ Department of Radiology, VA San Diego Healthcare System, San Diego, CA 92161, USA

⁵ Department of Radiology, University of California-San Diego, La Jolla, CA 92122, USA

* Correspondence: wbae@health.ucsd.edu

Abstract: Evaluation of the 3D bone morphology of the glenohumeral joint is necessary for pre-surgical planning. Zero echo time (ZTE) magnetic resonance imaging (MRI) provides excellent bone contrast and can potentially be used in the place of computed tomography. Segmentation of the shoulder anatomy, particularly the humeral head and the acetabulum, is needed for the detailed assessment of each anatomy and for pre-surgical preparation. In this study, we compared the performance of two popular deep learning models based on Google's DeepLab and U-Net to perform automated segmentation on ZTE MRI of human shoulders. Axial ZTE images of normal shoulders (n = 31) acquired at 3-Tesla were annotated for training with DeepLab and 2D U-Net, and the trained model was validated with testing data (n = 13). While both models showed visually satisfactory results for segmenting the humeral bone, U-Net slightly over-estimated while DeepLab under-estimated the segmented area compared to the ground truth. Testing accuracy quantified by Dice score was significantly higher ($p < 0.05$) for U-Net (88%) than DeepLab (81%) for the humeral segmentation. We have also implemented the U-Net model onto an MRI console for push-button DL segmentation processing. Although this is an early work with limitations, our approach has the potential to improve shoulder MR evaluation hindered by manual post-processing and may provide clinical benefit for quickly visualizing bones of the glenohumeral joint.



Citation: Carl, M.; Lall, K.; Pai, D.; Chang, E.Y.; Statum, S.; Brau, A.; Chung, C.B.; Fung, M.; Bae, W.C. Shoulder Bone Segmentation with DeepLab and U-Net. *Osteology* **2024**, *4*, 98–110. <https://doi.org/10.3390/osteology4020008>

Academic Editor: Giuseppe Banfi

Received: 29 March 2024

Revised: 27 April 2024

Accepted: 5 June 2024

Published: 11 June 2024



Copyright: © 2024 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/>).

Keywords: glenohumeral; glenoid; humeral head; image processing; ZTE; MRI; U-Net; DeepLab

1. Introduction

Evaluation of patient-specific 3D position and morphology of the glenohumeral joint bone is useful clinically for a multitude of reasons. These include a basic diagnosis of shoulder dislocation and fracture, visualization of the glenoid surface for determining glenoid bone loss and fracture [1,2], and the measurement of bone morphology for pre-surgical planning [3–6], which is needed for shoulder arthroplasty [7].

Computed tomography (CT) is the current gold standard for bone imaging in 3D [6,8–10]. While preferred for the lack of ionizing radiation, conventional magnetic resonance imaging (MRI) sequences (Figure 1A,B, provided as examples) have suboptimal contrast (i.e., bone has a similar signal to several other tissues) for distinguishing bone for visualization [9]. However, recent advances in MRI including ultrashort time-to-echo (UTE) [11,12] and zero time-to-echo (ZTE) techniques [13–16] have shown to be promising alternatives that can depict cortical bone with a uniform contrast (i.e., a low signal for the cortical bone and the air/back ground, a high signal for all other tissues such as muscle and trabecular bone and marrow; Figure 1C), making it relatively easy to isolate and visualize the bones [14,17], with minor processing. Unlike conventional magnetic resonance (MR) images that depict cortical bones and other soft tissues (such as ligaments) with low signal intensity (Figure 1A,B), ZTE

MRI provides a more uniform contrast for bone, for example with a high signal intensity in inverted ZTE images (Figure 1D). These studies have also shown that cortical bone morphology (e.g., surface contour) measured on UTE or ZTE are highly similar to that measured on a CT or a micro CT [17]. For these reasons, ZTE MRI is increasingly being prescribed when bone imaging is desired [17].

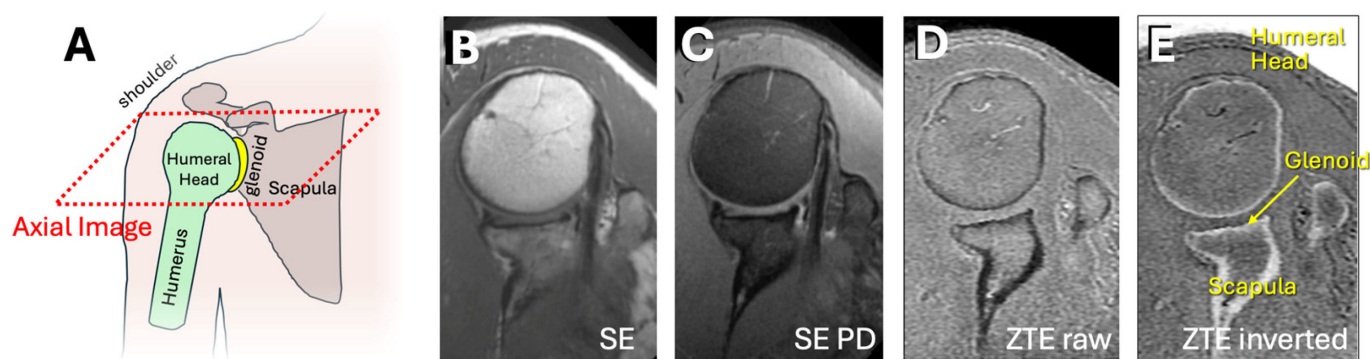


Figure 1. (A) Anatomy of the shoulder showing the major bone structures of the humerus, humeral head, glenoid, and scapula. The axial imaging plane is shown in the dotted red box. Conventional shoulder MR images were acquired using conventional (B) spin echo proton density weighted, (C) spin echo proton density weighted with fat suppression, and (D) zero echo time (ZTE) sequences. Conventional sequences do not isolate bone effectively. (E) Inverted ZTE image depicting mostly bony tissues with high signal intensity. Conventional magnetic resonance (MR) images depict non-bony tissues with similar contrast as bony tissues, making them less useful for bone-only imaging.

As mentioned above, it is useful to visualize and analyze individual bones of the glenohumeral joint (Figure 1A). This requires manual segmentation, a time-consuming process. This can be improved with traditional segmentation techniques utilizing the thresholding of pixel values [18], region growing [19], and active shape modeling [20]. Recently, deep learning techniques such as full convolutional network (FCN) [21] and U-Net [22] have been highly successful in performing image segmentation using small training data with good results. Another development includes Google's DeepLab, designed to perform image segmentation [23]. This model consists of the three main components of the feature extraction module, an atrous spatial pyramid pooling (ASPP) module, and a decoder module. It has been used widely for scene segmentation and to a limited degree for the medical imaging for tumors [24–26], but not for bone segmentation.

Specifically for the segmentation of shoulder MRI, several recent studies have used deep learning methods. One study used [27] 2D and 3D U-Net to segment the humeral head and glenoid on conventional spin echo and gradient echo MR images. Another approach [27–29] combined U-Net and AlexNet to perform rough and fine segmentation, respectively, of the humeral head and the glenoid bones. Others also used the convolutional network for shoulder bone segmentation [28–31], or bone recognition for shoulder muscle segmentation [32]. DeepLab, although readily available, has not been used previously for this application. It would be important to compare not only the performance of several models, but also the nuances of how different models behave.

Additionally, past studies focused solely on the performance of the model, not on the important aspect of how the segmentation integrates with the existing workflow. The vast majority of deep learning models require off-line processing for inference, and the raw output may not be compatible with digital imaging and communications in medicine (DICOM) viewers or the picture archiving and communication system (PACS) that maybe used by clinicians [33–35]. We aim to demonstrate an implementation that would be useful for clinical workflow.

The goal of this study was to demonstrate an end-to-end approach for creating and deploying deep learning-based models to perform shoulder MRI segmentation, comparing

two common deep learning models, to provide immediate utility without the hurdle of off-line processing. This is the first study to compare deep-learning based shoulder segmentation using U-Net and DeepLab, and we also describe an implementation for achieving this directly on the console, which may facilitate a quick clinical translation.

2. Materials and Methods

The parts of this study involving MR imaging of human subjects was approved by the institutional review board. The remaining images were obtained after de-identification.

The overview of the methodology is as follows: (1) acquire or collect ZTE MRI images of the shoulder similar to Figure 2; (2) pre-process images for uniform size and grayscale values (see Section 2.1. Pre-processing); (3) split data into training and testing sets; (4) perform annotation which is the manual segmentation of the images into three regions as shown in Figure 3; (5) perform training separately for U-Net and DeepLab models; here, pre-processed images will be the inputs, and the segmented images will be the outputs to be compared against manual segmentation images; (6) after training, perform inference on test data and determine accuracy; (7) finally, implement a deep learning model on the MRI console for immediate processing.



Figure 2. Inverted axial ZTE shoulder images used in this study were acquired with moderately varying scan parameters. (A) was acquired with TR = 88 ms, TE = 0.016 ms, FOV = 180 mm, matrix = 256×256 , and 1 mm slice thickness. (B) was acquired with TR = 458 ms, TE = 0.016 ms, FOV = 160 mm, matrix = 256×256 , and 1 mm slice thickness. (C) was similar to (B) but acquired with FOV = 180 mm, matrix = 512×512 , and 1.2 mm slice thickness. While varying in image contrast, all images shared a similar feature of depicting the bones of the shoulder with high signal intensity.

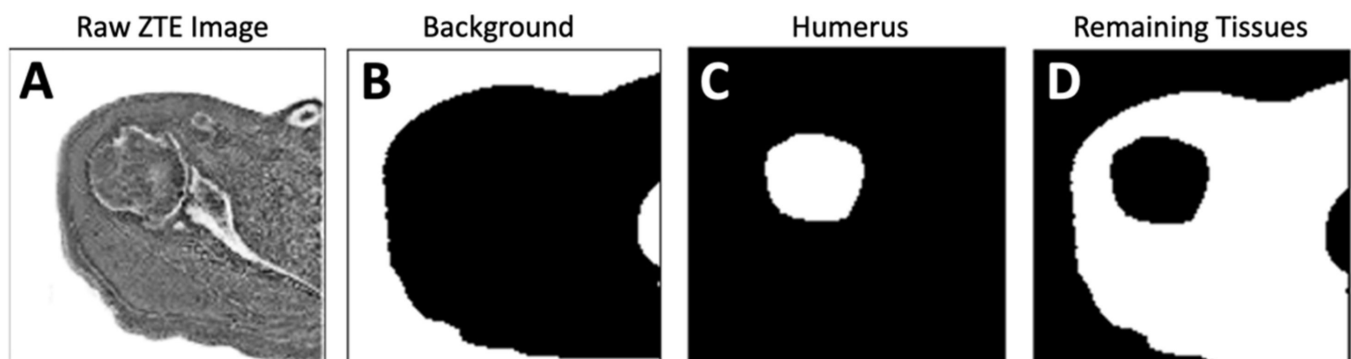


Figure 3. Manual segmentation of the MRI images. (A) Inverted ZTE MRI shoulder images acquired in the axial plane were manually annotated (segmented) into (B) background, (C) humeral head/humerus, and (D) the remaining tissues.

2.1. Zero Echo Time (ZTE) MRI Data

Training Data: For deep learning (DL) training data, we obtained a de-identified MRI shoulder dataset through an existing study where 7935 images from $n = 31$ normal asymptomatic unilateral (left or right) shoulders, acquired on General Electric 3-Tesla scanners, were available. The images were acquired in the axial plane using the ZTE sequence with mixed scan parameters: Time-to-Repetition (TR) = 100 to 600 ms, Time-to-Echo (TE) = 0.016 to 0.028 ms, field of view (FOV) = 160 to 240 mm, image matrix = 256×256 to 512×512 , slice thickness = 0.7 to 1.2 mm, number of slices = 90 to 200. Despite some variations in the scan parameters, the images had generally similar appearances (Figure 2A–C), depicting bone and air with low signal intensity, and most of the other soft tissues with a moderate to high signal intensity. As a part of the standard processing on the MRI console, raw ZTE images were intensity-corrected, and then inverted (i.e., from Figure 1C,D) to depict bone with high signal intensity (Figure 2). Examples of inverted axial ZTE shoulder images are shown in Figures 1D and 2. Unfortunately, demographical information (age, sex, etc.) was not available.

Testing Data: For testing data, additional axial ZTE shoulder data (1860 images) were obtained from $n = 13$ subjects that included existing data from nine volunteers (demographic information was not available) and four newly acquired datasets from subjects with recent shoulder pain (3 males, 1 female; age range 40 to 55 years old). These were all acquired with similar scan parameters to the training data.

Pre-Processing: Training and testing data were pre-processed prior to input into the deep learning models by normalizing the voxel values in each 3D stack (separately for each shoulder) between 0 and 1, and then conversion to an 8-bit image with voxel values between 0 and 255. The images were then resized to 256×256 in-plane using bilinear interpolation.

2.2. Opportunistic CT Data for Comparison

We obtained opportunistic data where both a clinical CT scan and a ZTE MRI of the same shoulder were available. ZTE scan parameters were similar to the training data. CT scanning parameters were: voltage = 140 kVp, current = 120 mA, reconstruction diameter = 275 mm, image matrix = 512×512 , and slice thickness = 1.25 mm. The image data were first registered using MATLAB, then the registered ZTE data were segmented using DL, while the CT data were segmented manually. The Dice score for humerus segmentation was determined, and the respective segmented images were fused and 3D rendered for visual comparison.

2.3. Annotation/Manual Segmentation

All images were annotated using ImageJ [36] (v2.1.0). Inverted ZTE MR images (Figure 3A) were loaded as a three-dimensional (3D) stack and annotated using the Segmentation Editor plugin. On every 2 to 3 slices, the boundaries of the humerus were drawn using a polygon selection tool, interpolated between slices, then filled to create a binary 3D image for the humerus (Figure 3C). For the background segmentation (Figure 3B), we performed thresholding and additional manual clean-up. Finally, segmentation of the remaining tissues was created through inversion of the background image and subtraction of the humerus image. This yielded three separate binary 3D images, representing (i) the background/air (Figure 3B) including the bulk of the lung, (ii) the humeral head and humerus bone (Figure 3C), and (iii) all other soft and bony tissues (i.e., glenoid, acromion, etc., Figure 3D). Note that humerus segmentation was performed loosely around the structure to avoid accidental cropping of the humeral head. While we used a single observer (a non-physician with 10 years of experience in imaging research, trained in the musculoskeletal section), we felt that the bone segmentation task did not require extensive training due to a high contrast between bone and soft tissues, and that our approach of using loosely fitting annotation allowed for rapid annotation that still included the structure of interest without error.

2.4. Deep Learning Segmentation Models

We have implemented two-dimensional (2D) U-Net [22] and DeepLab v3 [23] deep learning (DL) models in MATLAB with the Deep Learning Toolbox (R2021b) to perform the segmentation of shoulder ZTE MR images. The DL models first take in inverted ZTE shoulder images that have been resized to 256×256 voxels.

The U-Net (Figure 4A) applies 64 convolution filters with a 3×3 kernel size, and two convolution operations at each step, with an encoder depth of 5 (or 9 layers, 4 down-sampling followed by 4 up-sampling). The DeepLab v3 (Figure 4B) uses spatial pyramid pooling module for encoding, which captures contextual information at multiple scales using 3×3 convolution filters with different dilation rates of 6, 12, and 18. This is then pooled and concatenated in the decoder to produce the final segmentation.

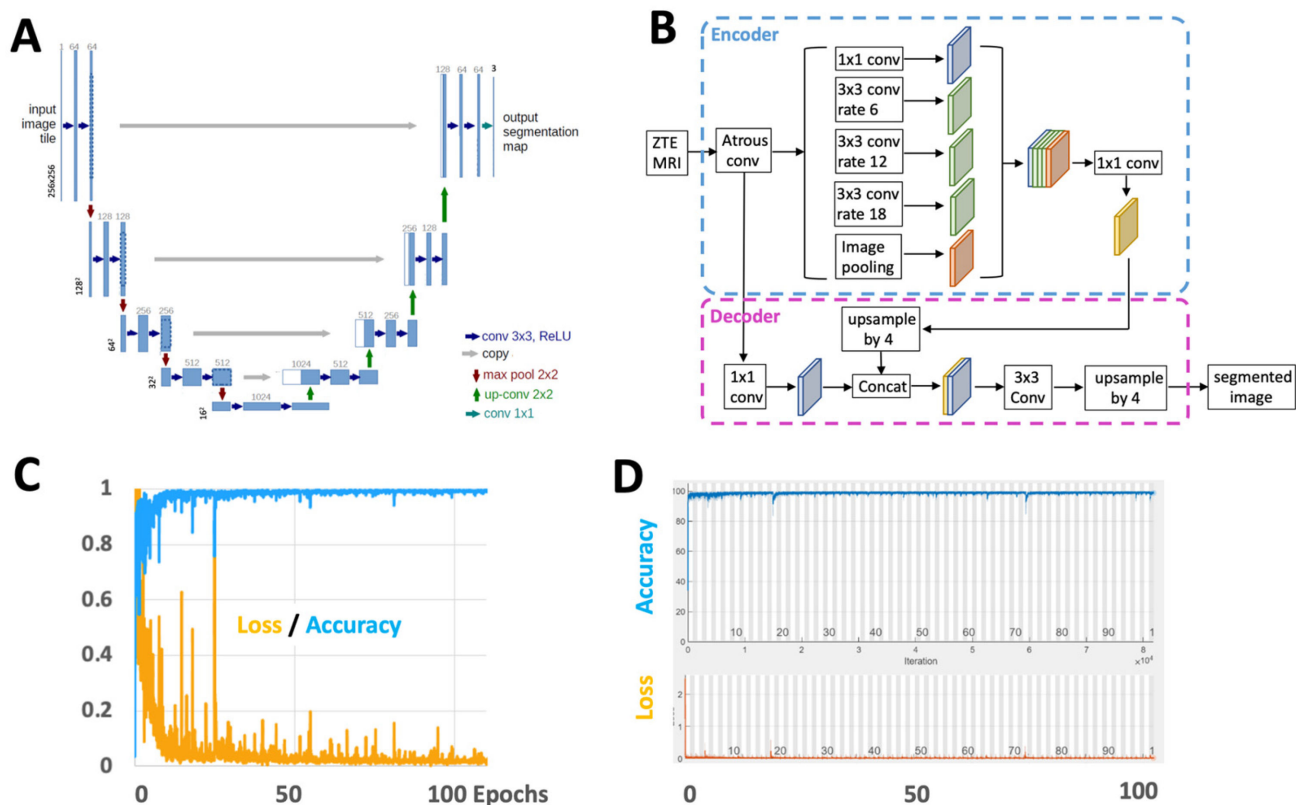


Figure 4. Architectures of (A) U-Net and (B) DeepLab used in this study. Adapted from [22] and [37], respectively. Training results showing accuracy and loss values for (C) U-Net and (D) DeepLab.

In both models, the final output consists of the pixel classification layer with 3 classes, each for the background (air), the humeral head/humerus, and the remaining other tissues. The models were trained to 120 epochs (~2 days) using the default setting (Adam optimizer, L2 regularization, population batch normalization, shuffling image every epoch, mini batch size of 8) using cross entropy as the loss function, on a Windows 10 PC with i7-10700K CPU, 32GB RAM, RTX3090 GPU with 24 GB VRAM. Training accuracy quickly converged to >99% (Figure 4C,D) and remained high. The weight that provided the lowest loss during training was kept and used for the remainder of the study. A MATLAB code is provided below in Section 2.4.1 to clarify the model building and training processes. (We also trained on augmented images using rotation, translation, and rescaling, but unfortunately this did not improve the test results. Additionally, the current approach did not use cross-validation, which is a moderate limitation.)

2.4.1. MATLAB Code

```
% data directories
imageDir = "directory for pre-processed MRI images"
labelDir = "directory for segmented MRI images"

% create image and label datastores
imds=imageDatastore (imageDir)
pxds=imageDatastore(labelDir)

% combined training datastore
dsTrain=combine (imds, pxds)

% create deep network
imageSize=[256 256]
encoderDepth=5;
classNames=["background" , "humerus" , "remaining"];
numClasses=3; % one for each of the segmentation

% create U-Net
network = unetLayers (imageSize, numClasses, 'EncoderDepth', encoderDepth)

% alternatively, create Deeplab (but not both in the same program)
network = deeplabv3plusLayers(imageSize,numClasses, 'resnet18');

% training options
train_options=trainingOptions ('adam' ...
    'LearnRateDropFactor',0.1, ...
    'LearnRateDropPeriod', 3, ...
    'Shuffle', 'every-epoch', ...
    'MaxEpochs', 120, ...
    'MiniBatchSize', 8); ...

% start training
Segmentation_Network = trainNetwork ( dsTrain, network, train_options )

%%% inference after training %%%
segmented_image = semanticseg( "input MRI image", Segmentation_Network )
```

2.5. Inference Accuracy

After training the U-net model, DL segmentation was performed on the validation images (each dataset taking ~2 min) and compared against the manual annotation. The following similarity metrics were determined. The Dice score [38,39] provides a measure of image (i.e., segmentation mask) overlap, defined as Equation (1):

$$\text{Dice} = (2 \text{ TP}) / (2 \text{ TP} + \text{FP} + \text{FN}) \quad (1)$$

where TP is the number of true positive voxels (i.e., a value of 1 in both DL and manual segmentations), FP is the number of false positive voxels (a value of 1 in DL, a value of 0 in manual segmentation), and FN is the number of false negative voxels (a value of 0 in DL, a value of 1 in manual segmentation).

Sensitivity [40] was determined as Equation (2):

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

Specificity [40] was determined as Equation (3):

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (3)$$

where TN is the number of true negative voxels (a value of 0 in both DL and manual segmentations).

2.6. Clinically Applicable DL Segmentation Workflow

In addition to the development of the DL segmentation model, additional processing steps were developed to create a workflow that can be used on an MRI scanner, without the need for off-line processing. As shown in the flowchart in Figure 5A, a MATLAB runtime (R2021b for Linux) program was compiled as binary and installed on an MRI console to: (1) read inverted ZTE DICOM images; (2) perform DL segmentation (using the same routine as the off-line version) to create masks of the humeral head/humerus and the remaining tissues; and (3) multiply with input ZTE images to create a new series of images showing the separated structures viewable on the MRI console or any PACS (Figure 5B). This enabled a push-button DL segmentation of ZTE shoulder images directly on the MRI scanner for an immediate evaluation following acquisition. As a quality control measure, we compared the DICOM images from the MRI console to the those from off-line programs and found no difference in the images.



Figure 5. (A) Flow chart of ZTE DL processing, which reads raw digital imaging and communications in medicine (DICOM) images, and performs DL segmentation to create masks for the humerus and the remaining tissues. The masks are then multiplied with the raw image to create segmented DICOM images that are saved as a new series in the exam. (B) Segmented DICOM images viewed in a PACS viewer, showing the original image on the left, the segmented humeral bone in the middle, and segmented remaining tissues on the right.

2.7. Statistics

For the validation data, Dice scores, sensitivity, and specificity for segmentation of the humerus and the remaining tissues were tabulated, and the mean and standard deviation of the values were determined and plotted as box plots. To determine the effect of different models in the segmentation performance, we compared the mean values between U-Net and DeepLab models using a *t*-test with a significance level set at 0.05, using Systat (v10, Grafiti LLC, Palo Alto, CA, USA).

3. Results

3.1. Training

The training results were good, converging to >99% accuracy (Figure 4C,D). The model weight that achieved the highest accuracy was kept and used for the remainder of the study.

3.2. Testing

Figure 6A–F show segmented ZTE test images. Ground truth segmentations (Figure 6A,D) are shown next to DL segmentation performed by U-Net (Figure 6B,E) and DeepLab (Figure 6C,F). While both models performed well by including the important areas (i.e., bone) in the humeral head segmentation, U-net slightly over-estimated segmentation for the humeral head while DeepLab slightly under-estimated. The segmented ZTE images could be used for 3D morphologic evaluation of the humeral head (Figure 6H) and the glenoid (Figure 6I).

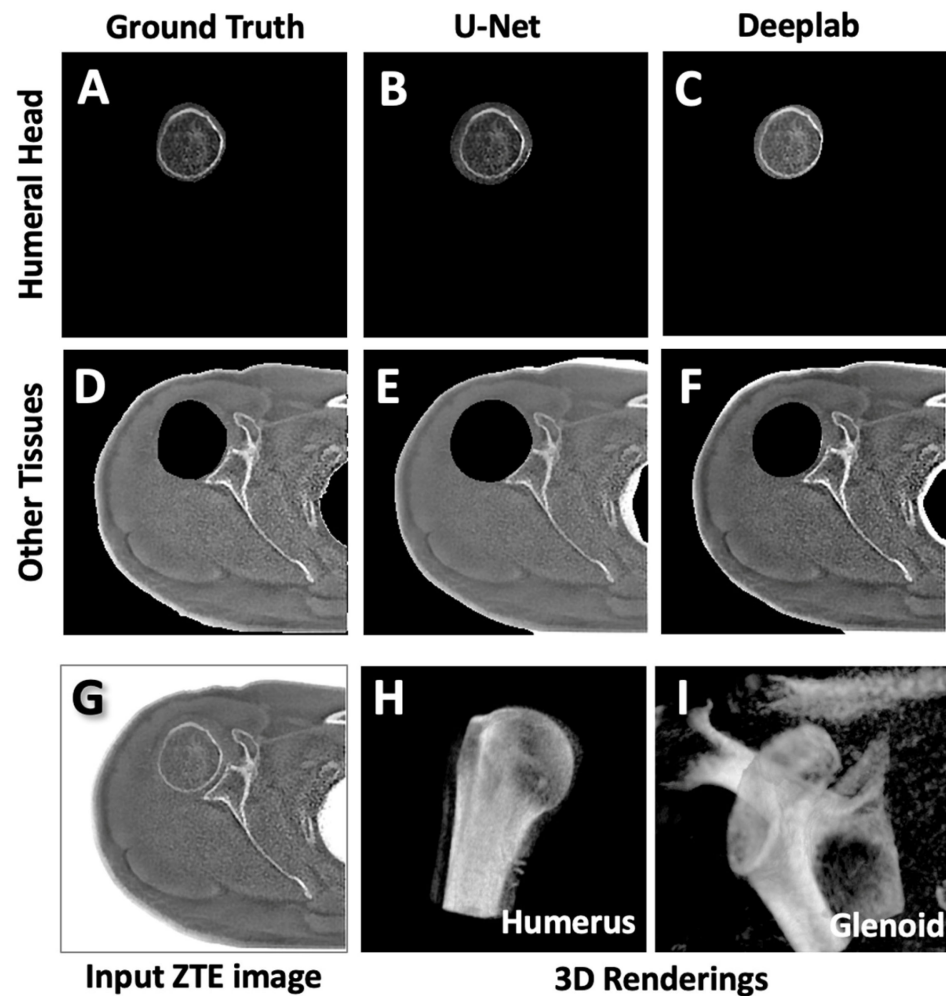


Figure 6. Segmentation results on test images. (A,D) Ground truth or manually segmented images of the humeral bone and the remaining other tissues shown for comparison. Output segmented images of (B,C) the humeral head and (E,F) the remaining tissues after DL segmentation performed by (B,E) U-Net and (C,F) DeepLab. Qualitatively, U-Net slightly over-estimated the area for the humeral head while DeepLab slightly under-estimated. (G) Input ZTE MRI image is shown. (H,I) Segmented ZTE images (from the ground truth; (A) and (D)) were used to create separate 3D renderings of the (H) humerus and (I) glenoid/scapular bone.

Figure 7 shows the testing results quantified using the Dice score, the sensitivity, and the specificity. Figure 7A is the boxplot for the humerus/humeral head and Figure 7B is the boxplot for the remaining other tissues; the blue color represents U-Net and the red color represents DeepLab.

Table 1 presents the means and standard deviations, and the p -values suggesting the difference in values between the models. U-Net had consistently high Dice scores throughout, averaging 0.88 for the humerus/humeral head and 0.94 for the remaining tissues. The sensitivity and specificity values were also very high, ranging from 0.88 to 0.99. In comparison, DeepLab had a significantly lower Dice score (0.81 vs. 0.88, $p = 0.027$) and sensitivity (0.71 vs. 0.91, $p < 0.001$), but greater specificity (0.999 vs. 0.995, $p < 0.001$), for the humerus/humeral head segmentation. Other metrics were not statistically different. This, combined with visual comparison, suggested that U-Net was generally more sensitive and slightly over-included areas of humeral head, while DeepLab was more conservative in this regard.

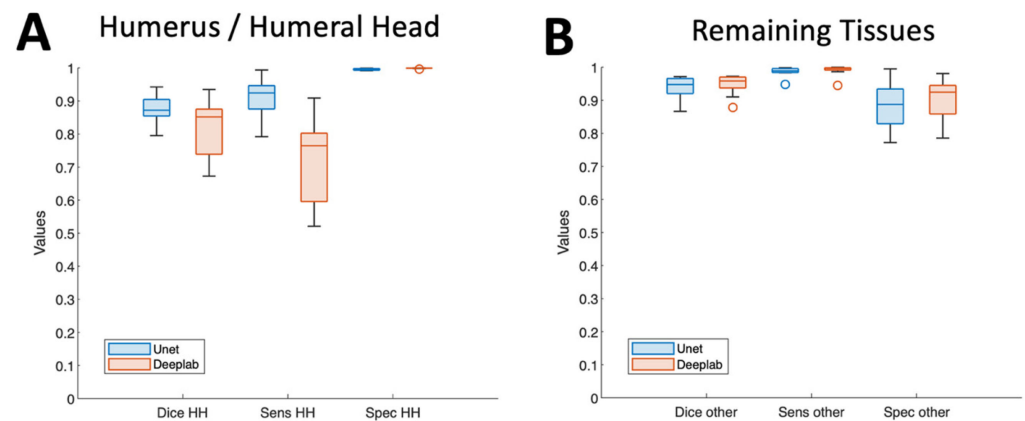


Figure 7. DL model performances compared. Boxplots of inference accuracy (Dice score, sensitivity, specificity) quantified on the humeral bone (A) and the remaining tissue (B), determined using U-Net (blue) and DeepLab (red) models. Marked differences in the accuracy metrics for the humeral bone were noted.

Table 1. Mean and standard deviation of the Dice scores, sensitivity, and specificity values. *p*-values from *t*-tests indicate statistical difference between the mean values obtained using U-Net vs. DeepLab.

		Dice HH	Dice Other	Sens HH	Sens Other	Spec HH	Spec Other
Deep U-Net Lab	Mean	0.876	0.940	0.910	0.987	0.995	0.883
	SD	0.043	0.031	0.061	0.013	0.003	0.070
	N	13	13	13	13	13	13
	Mean	0.811	0.949	0.715	0.992	0.999	0.903
	SD	0.088	0.030	0.132	0.015	0.001	0.055
	N	13	13	13	13	13	13
<i>p</i> -value		0.027	0.467	<0.001	0.426	<0.001	0.424

3.3. Comparison vs. CT

In the opportunistic data where both ZTE MRI (Figure 8A) and CT (Figure 8B) data were available from the same subject, the Dice score of the humerus segmentation using DL was 97%. We created a fused 3D rendering (Figure 8C), which shows an excellent overlap (white) between ZTE (purple) and CT (green), and is likely to yield similar values when measured for length, etc. However, this needs to be validated in additional samples.

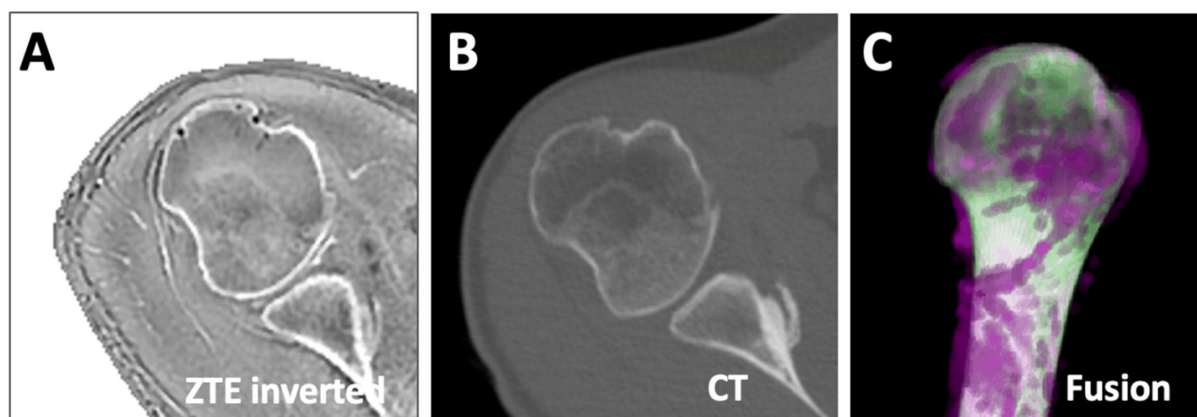


Figure 8. Comparison of MRI vs. CT segmentation. ZTE MRI (A) and CT (B) data of the same subject were registered and segmented (using U-Net for MRI, manually for CT). The segmented images were fused (C), showing the overlapping regions as white, and the non-overlapping regions in magenta for MRI and green for CT.

4. Discussion

We have successfully implemented a DL model to segment the humerus and other tissues from ZTE MRI shoulder images. Using a small amount of training data, a reasonable level of performance was achieved both visually and quantitatively with moderately high Dice scores, sensitivity, and specificity using both U-Net and DeepLab models. Comparison of segmented ZTE vs. CT showed an excellent overlap, suggesting that ZTE MRI could become a useful modality for imaging bony tissues in the body. The DL model has also been implemented on an MRI scanner to perform a push-button automated segmentation in 2 to 3 min. This work could be useful clinically when it is desired to evaluate 3D bone morphology at the glenohumeral joint for bone defects [41] or dysplasia [42], or for pre-surgical measurement [3–6], by eliminating the need for manually segmenting the humeral bone from the glenoid.

While there have been a few studies on DL segmentation of shoulder MRI, this is the first study utilizing ZTE images, and using DeepLab architecture. Compared to other DL segmentation models trained on conventional MRI, the performance of our model may appear underwhelming at first. Rodrigues et al. [27] used a 2D U-Net model to achieve a sensitivity and specificity of 95% and 99%, respectively, for the humerus, and 86% and 92%, respectively, for the glenoid. Mu et al. [29] and Wang et al. [28] both achieved an average sensitivity of ~95% for segmenting the two bones. One study utilizing a conventional thresholding method combined with a manual selection [43] found that the MRI measure of the glenoid area correlated strongly (an intraclass correlation coefficient of 0.94) with that from CT. In contrast, our models achieved a somewhat lower average sensitivity of 71% (DeepLab) and 92% (U-Net), but a high specificity of >99% (both models) for the humerus segmentation. This may be related to our approach of using a loosely oversized region (Figure 6A) for annotating the humerus bone, unlike in past studies that needed to use a precisely defined region for each bone [27]. While our approach may have resulted in a lowered sensitivity, minor deviations in the segmentation mask do not appear to cause problems for the simple purpose of isolating the humeral head and glenoid for visualization. For additional quantifications such as volume measurement, conventional image processing techniques (e.g., thresholding [18,44–46]) may be applied to further segment the bone from surrounding soft tissues.

Comparison between U-Net and DeepLab suggested that while both yielded visually satisfactory results, there was a tendency of U-Net to slightly over-estimate the area for the humerus segmentation, while DeepLab tended to slightly under-estimate. A consequence of this can be seen in Figure 6, where U-Net segmentation included a sliver of glenoid bone (Figure 6B), while DeepLab missed a sliver of humeral bone (Figure 6C). This was also apparent in the significantly higher sensitivity of humeral segmentation for U-Net, and a significantly higher specificity of humeral segmentation for DeepLab. Although the exact reasons for such differences are unclear, DeepLab's architecture that uses atrous (i.e., dilated) convolutions tends to broaden the receptive field, which may be advantageous for detection and inclusion, but may make the segmentation less precise.

This early study has several limitations. First, we used only axial images for the training and testing. Given the 2D nature of the models, shoulder images in other planes (coronal, sagittal) will not provide the expected results in the present study. However, ZTE images are often acquired as an isotropic 3D stack [16], so axial reformatting will not degrade the image quality. The variations in scan parameters of the training images, while lowering segmentation accuracy, would have been beneficial in terms of the generalizability of the model. Segmentation was performed by a single observer, which is less desirable than using an average of multiple observers. However, this may not be critical as we utilized loosely fitting masks unlike most other studies. U-Net and DeepLab models were introduced some time ago and there are now newer models such as the Segment Anything Model. However, the application of the existing models to the shoulder segmentation and the comparison of the results still provide useful insight, despite our study not specifically using transfer learning [30,47] from existing weights and models. Additionally, U-Net

and DeepLab models are readily available on the MATLAB platform for easier adaptation, unlike the newest models that may require expertise in a computer science discipline. The number of training datasets (7935 images from 31 subjects), while sufficient to yield a functional model, was too small to capture large variations found in normal shoulder anatomy and did not include any shoulders with known abnormalities in bone morphology [48–51]. Additional and varied datasets in the future will likely improve segmentation performance. ZTE MRI, while providing superior bone contrast compared to conventional MR sequences, still falls short compared to CT scanning for bone evaluation. Tissues other than bone can appear iso-intense with bone (Figure 8A), and this may adversely affect visualization (Figure 8C, purple signal) using our loosely encompassing segmentation masks.

5. Conclusions

In conclusion, we developed and deployed a fully automated methodology based on two popular deep learning models to segment the humerus and other bones on novel ZTE MR images of the human shoulder. Although this is an early attempt with limitations, with additional training data and model refinement, this study has the potential to improve clinical practice, by improving clinical workflow such as evaluation of 3D bone morphology or pre-surgical measurement of the glenohumeral joint by providing rapid and automated segmentation of the humeral bone.

Author Contributions: Funding Acquisition: W.C.B., A.B. and C.B.C.; project Supervision: W.C.B. and M.F.; conceptualization: W.C.B. and C.B.C.; data collection: W.C.B., M.C., D.P., E.Y.C., S.S., A.B. and M.F.; methodology: W.C.B., M.C. and K.L.; software: W.C.B., M.C. and K.L.; validation: W.C.B., M.C., K.L. and D.P.; administrative: W.C.B., S.S. and M.F.; manuscript writing: W.C.B.; reviewing and editing: All authors. All authors have read and agreed to the published version of the manuscript.

Funding: This article was made possible in part by research grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (R01 AR066622 and P30 AR073761), and from General Electric Healthcare in support of Dr. Bae.

Institutional Review Board Statement: The parts of this study involving MR imaging of human subjects were approved by the institutional review board. The remaining images were obtained after de-identification.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to privacy issues.

Conflicts of Interest: This work was funded by a research grant from General Electric Healthcare. Drs. Carl, Brau, and Fung are employees of General Electric Healthcare.

References

1. Bakshi, N.K.; Cibulas, G.A.; Sekiya, J.K.; Bedi, A. A Clinical Comparison of Linear- and Surface Area-Based Methods of Measuring Glenoid Bone Loss. *Am. J. Sports Med.* **2018**, *46*, 2472–2477. [\[CrossRef\]](#)
2. Mahylyis, J.M.; Entezari, V.; Jun, B.J.; Iannotti, J.P.; Ricchetti, E.T. Imaging of the B2 Glenoid: An Assessment of Glenoid Wear. *J. Shoulder Elb. Arthroplast.* **2019**, *3*, 2471549219861811. [\[CrossRef\]](#)
3. Bizzotto, N.; Tami, I.; Santucci, A.; Adani, R.; Poggi, P.; Romani, D.; Carpeggiani, G.; Ferraro, F.; Festa, S.; Magnan, B. 3D Printed replica of articular fractures for surgical planning and patient consent: A two years multi-centric experience. *3d Print. Med.* **2015**, *2*, 2. [\[CrossRef\]](#)
4. Jehan, S.; Akhi Baig, N.M.; Tavitian, J. Treatment of Shoulder Dislocation with Greater Tuberosity and Glenoid Fractures. *J. Coll. Physicians Surg. Pak.* **2016**, *26*, 997–999.
5. Goetti, P.; Becce, F.; Terrier, A.; Farron, A. Three-dimensional surgical planning, patient-specific instrumentation and intraoperative navigation in shoulder arthroplasty. *Rev. Medicale Suisse* **2019**, *15*, 2299–2302. [\[CrossRef\]](#)
6. Moroder, P.; Resch, H.; Schnaitmann, S.; Hoffelner, T.; Tauber, M. The importance of CT for the pre-operative surgical planning in recurrent anterior shoulder instability. *Arch. Orthop. Trauma. Surg.* **2013**, *133*, 219–226. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Macias-Hernandez, S.I.; Morones-Alba, J.D.; Miranda-Duarte, A.; Coronado-Zarco, R.; Soria-Bastida, M.L.A.; Nava-Bringas, T.; Cruz-Medina, E.; Olascoaga-Gomez, A.; Tallabs-Almazan, L.V.; Palencia, C. Glenohumeral osteoarthritis: Overview, therapy, and rehabilitation. *Disabil. Rehabil.* **2017**, *39*, 1674–1682. [\[CrossRef\]](#)

8. Jun, B.J.; Ricchetti, E.T.; Haladik, J.; Bey, M.J.; Patterson, T.E.; Subhas, N.; Li, Z.M.; Iannotti, J.P. Validation of a 3D CT imaging method for quantifying implant migration following anatomic total shoulder arthroplasty. *J. Orthop. Res.* **2022**, *40*, 1270–1280. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Imhoff, A.B.; Hodler, J. Correlation of MR imaging, CT arthrography, and arthroscopy of the shoulder. *Bull. Hosp. Jt. Dis.* **1996**, *54*, 146–152.
10. Kuhlman, J.E.; Fishman, E.K.; Ney, D.R.; Magid, D. Complex shoulder trauma: Three-dimensional CT imaging. *Orthopedics* **1988**, *11*, 1561–1563. [\[CrossRef\]](#)
11. Bae, W.C.; Biswas, R.; Chen, K.; Chang, E.Y.; Chung, C.B. UTE MRI of the Osteochondral Junction. *Curr. Radiol. Rep.* **2014**, *2*, 35. [\[CrossRef\]](#)
12. Geiger, D.; Bae, W.C.; Statum, S.; Du, J.; Chung, C.B. Quantitative 3D ultrashort time-to-echo (UTE) MRI and micro-CT (muCT) evaluation of the temporomandibular joint (TMJ) condylar morphology. *Skelet. Radiol.* **2014**, *43*, 19–25. [\[CrossRef\]](#)
13. Cheng, K.Y.; Moazamian, D.; Ma, Y.; Jang, H.; Jerban, S.; Du, J.; Chung, C.B. Clinical application of ultrashort echo time (UTE) and zero echo time (ZTE) magnetic resonance (MR) imaging in the evaluation of osteoarthritis. *Skelet. Radiol.* **2023**, *52*, 2149–2157. [\[CrossRef\]](#)
14. Bharadwaj, U.U.; Coy, A.; Motamedi, D.; Sun, D.; Joseph, G.B.; Krug, R.; Link, T.M. CT-like MRI: A qualitative assessment of ZTE sequences for knee osseous abnormalities. *Skelet. Radiol.* **2022**, *51*, 1585–1594. [\[CrossRef\]](#)
15. Eley, K.A.; Delso, G. Automated 3D MRI rendering of the craniofacial skeleton: Using ZTE to drive the segmentation of black bone and FIESTA-C images. *Neuroradiology* **2021**, *63*, 91–98. [\[CrossRef\]](#)
16. Jang, H.; Carl, M.; Ma, Y.; Searleman, A.C.; Jerban, S.; Chang, E.Y.; Corey-Bloom, J.; Du, J. Inversion recovery zero echo time (IR-ZTE) imaging for direct myelin detection in human brain: A feasibility study. *Quant. Imaging Med. Surg.* **2020**, *10*, 895–906. [\[CrossRef\]](#)
17. Breighner, R.E.; Endo, Y.; Konin, G.P.; Gulotta, L.V.; Koff, M.F.; Potter, H.G. Technical Developments: Zero Echo Time Imaging of the Shoulder: Enhanced Osseous Detail by Using MR Imaging. *Radiology* **2018**, *286*, 960–966. [\[CrossRef\]](#)
18. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Sys. Man. Cyber.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
19. Biratu, E.S.; Schwenker, F.; Debelee, T.G.; Kebede, S.R.; Negera, W.G.; Molla, H.T. Enhanced Region Growing for Brain Tumor MR Image Segmentation. *J. Imaging* **2021**, *7*, 22. [\[CrossRef\]](#)
20. Spinczyk, D.; Krason, A. Automatic liver segmentation in computed tomography using general-purpose shape modeling methods. *Biomed. Eng. Online* **2018**, *17*, 65. [\[CrossRef\]](#)
21. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
22. 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, Munich, Germany, 5–9 October 2015*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
23. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv* **2018**, *1802*, 02611.
24. Liu, J.; Sun, X.; Li, R.; Peng, Y. Recognition of Cervical Precancerous Lesions Based on Probability Distribution Feature Guidance. *Curr. Med. Imaging* **2022**, *18*, 1204–1213. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Shia, W.C.; Hsu, F.R.; Dai, S.T.; Guo, S.L.; Chen, D.R. Semantic Segmentation of the Malignant Breast Imaging Reporting and Data System Lexicon on Breast Ultrasound Images by Using DeepLab v3. *Sensors* **2022**, *22*, 5352. [\[CrossRef\]](#)
26. 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. [\[CrossRef\]](#)
27. Cantarelli Rodrigues, T.; Deniz, C.M.; Alaia, E.F.; Gorelik, N.; Babb, J.S.; Dublin, J.; Gyftopoulos, S. Three-dimensional MRI Bone Models of the Glenohumeral Joint Using Deep Learning: Evaluation of Normal Anatomy and Glenoid Bone Loss. *Radiol. Artif. Intell.* **2020**, *2*, e190116. [\[CrossRef\]](#)
28. Wang, G.; Han, Y. Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint. *Comput. Methods Programs Biomed.* **2021**, *200*, 105862. [\[CrossRef\]](#)
29. Mu, X.; Cui, Y.; Bian, R.; Long, L.; Zhang, D.; Wang, H.; Shen, Y.; Wu, J.; Zou, G. In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks. *Comput. Methods Programs Biomed.* **2021**, *211*, 106325. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Conze, P.H.; Brochard, S.; Burdun, V.; Sheehan, F.T.; Pons, C. Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders. *Comput. Med. Imaging Graph.* **2020**, *83*, 101733. [\[CrossRef\]](#)
31. Khan, S.H.; Khan, A.; Lee, Y.S.; Hassan, M.; Jeong, W.K. Segmentation of shoulder muscle MRI using a new region and edge based deep auto-encoder. *Multimed. Tools Appl.* **2022**, *82*, 14963–14984. [\[CrossRef\]](#)
32. Wakamatsu, Y.; Namiya, N.; Zhou, X.F.; Kato, H.; Hara, T.; Fujita, H. Automatic Segmentation of Supraspinatus Muscle via Bone-Based Localization in Torso Computed Tomography Images Using U-Net. *IEEE Access* **2021**, *9*, 155555–155563. [\[CrossRef\]](#)
33. Ortiz-Posadas, M.R.; Benitez-Graniel, M.A.; Pimentel-Aguilar, A.B. PACS: Reengineering workflow in the Imaging Department of a National Health Institute in Mexico. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2007**, *2007*, 3585–3588. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Hagland, M. Reshaping radiology: Change management and workflow optimization give PACS new punch. *Healthc. Inform.* **2004**, *21*, 24–26, 28. [\[PubMed\]](#)

35. Fillicelli, T. Future of PACS: Advanced integration with RIS and workflow management. *Radiol. Manag.* **2001**, *23*, 12–13.
36. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]
37. Zheng, S.-X.; Zhang, J.; Che, X.; Li, Y. Road information detection method based on deep learning. *J. Phys. Conf. Ser.* **2021**, *1827*, 012181. [[CrossRef](#)]
38. Dice, L. Measure of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
39. Eelbode, T.; Bertels, J.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. *IEEE Trans. Med. Imaging* **2020**, *39*, 3679–3690. [[CrossRef](#)] [[PubMed](#)]
40. Shreffler, J.; Huecker, M.R. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2023.
41. Park, I.; Oh, M.J.; Shin, S.J. Effects of Glenoid and Humeral Bone Defects on Recurrent Anterior Instability of the Shoulder. *Clin. Orthop. Surg.* **2020**, *12*, 145–150. [[CrossRef](#)] [[PubMed](#)]
42. Abboud, J.A.; Bateman, D.K.; Barlow, J. Glenoid Dysplasia. *J. Am. Acad. Orthop. Surg.* **2016**, *24*, 327–336. [[CrossRef](#)]
43. Lansdown, D.A.; Cvetanovich, G.L.; Verma, N.N.; Cole, B.J.; Bach, B.R.; Nicholson, G.; Romeo, A.; Dawe, R.; Yanke, A.B. Automated 3-Dimensional Magnetic Resonance Imaging Allows for Accurate Evaluation of Glenoid Bone Loss Compared With 3-Dimensional Computed Tomography. *Arthroscopy* **2019**, *35*, 734–740. [[CrossRef](#)]
44. Zhang, Y.; He, Z.; Fan, S.; He, K.; Li, C. Automatic Thresholding of Micro-CT Trabecular Bone Images. In Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics, Sanya, China, 27–30 May 2008; pp. 23–27.
45. Burghardt, A.J.; Kazakia, G.J.; Majumdar, S. A local adaptive threshold strategy for high resolution peripheral quantitative computed tomography of trabecular bone. *Ann. Biomed. Eng.* **2007**, *35*, 1678–1686. [[CrossRef](#)] [[PubMed](#)]
46. Buie, H.R.; Campbell, G.M.; Klinck, R.J.; MacNeil, J.A.; Boyd, S.K. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. *Bone* **2007**, *41*, 505–515. [[CrossRef](#)] [[PubMed](#)]
47. Marin, R.; Chang, V. Impact of transfer learning for human sperm segmentation using deep learning. *Comput. Biol. Med.* **2021**, *136*, 104687. [[CrossRef](#)] [[PubMed](#)]
48. Alentorn-Geli, E.; Wanderman, N.R.; Assenmacher, A.T.; Cofield, R.H.; Sanchez-Sotelo, J.; Sperling, J.W. Reverse shoulder arthroplasty for patients with glenohumeral osteoarthritis secondary to glenoid dysplasia. *Acta Orthop. Belg.* **2019**, *85*, 274–282.
49. Inui, H.; Nobuhara, K. Glenoid osteotomy for atraumatic posteroinferior shoulder instability associated with glenoid dysplasia. *Bone Jt. J.* **2018**, *100-B*, 331–337. [[CrossRef](#)] [[PubMed](#)]
50. Sewell, M.D.; Al-Hadithy, N.; Higgs, D.; Bayley, I.; Falworth, M.; Lambert, S. Complex shoulder arthroplasty in patients with skeletal dysplasia can decrease pain and improve function. *J. Shoulder Elb. Surg.* **2014**, *23*, 1499–1507. [[CrossRef](#)] [[PubMed](#)]
51. Seaggar, R.M.; Loveridge, J.; Crowther, M.A. Beware of glenoid dysplasia mimicking bone trauma in the injured shoulder. *Int. J. Shoulder Surg.* **2009**, *3*, 37–40. [[CrossRef](#)]

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.