Benjamin Hohlmann*†, Peter Brößner†, Kristian Welle, Klaus Radermacher

# Segmentation of the Scaphoid Bone in Ultrasound Images

## A comparison of two machine learning architectures for in-vivo segmentation

**Abstract:** For the percutaneous fixation of scaphoid fractures, navigated approaches have been proposed to facilitate screw placement. Based on ultrasound imaging, navigation can be carried out in a cost-effective and fast manner, furthermore avoiding harmful radiation. For this purpose, a fast and efficient architecture for the automated segmentation of scaphoid bone in ultrasound volume images is needed.

Methods: For 2D segmentation of the scaphoid, two architectures are taken into account: 2D nnUNet and Deeplabv3+. These architectures are trained and evaluated on a newly created dataset consisting of 67 annotated in-vivo ultrasound volume images (4576 slice images).

Results: In terms of Dice coefficient, the 2D nnUNet achieves 0.67 compared to 0.57 for the Deeplabv3+. In terms of distance metrics, the 2D nnUNet shows an average symmetric surface distance error of 0.66mm, while the Deeplabv3+ achieves 0.55mm.

Conclusion: Fast and accurate segmentation of the scaphoid in ultrasound volumes is feasible. Both architectures show competitive results.

**Keywords:** Ultrasound imaging, machine learning, segmentation, scaphoid fixation

https://doi.org/10.1515/cdbme-2021-1017

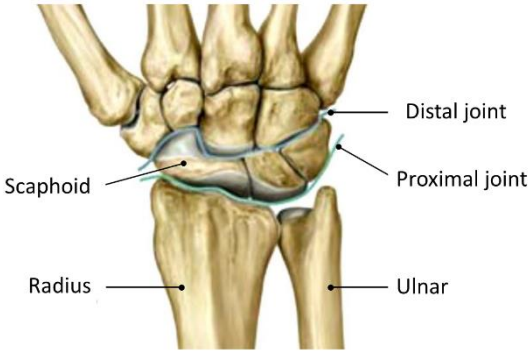


**Figure 1:** Anatomy of the human wrist. Adapted from [1].

## 1 Introduction

The scaphoid is the largest carpal bone in the human wrist, see Figure 1. It is prone to fractures, for example after a fall on the extended wrist [2]. Diagnosis involves different imaging techniques like ultrasound and computed tomography. Depending on the location and type of fracture, treatment of such an injury may be nonsurgical using a cast. In certain cases, an operative treatment offers an alternative that comes with faster recovery time. Here, an osteosynthesis screw is placed in the bone fragments, pulling them together and facilitating healing. Placement of this screw may be performed in a minimally invasive fashion, sparing the surrounding soft tissue. In this case, imaging is required and conventionally, fluoroscopy is used. However, this imaging technique exposes both, the patient and the surgeon, to ionizing radiation [3]. Further, it depicts a 2D projection of the 3D geometry, rendering positioning along the projection direction difficult [4].

Ultrasound offers an alternative: It is a cost-effective and widely available imaging technique with real-time capability. It furthermore allows for 3D imaging. However, its signal to noise ratio is very low and in contrast to fluoroscopy, soft tissue interfaces resemble bone surfaces. Furthermore, interpretation of volumetric images is a difficult task for

†**Both authors contributed equally**
***Corresponding author: Benjamin Hohlmann:** RWTH Aachen, Pauwelsstraße 20, Aachen, Germany, e-mail: hohlmann@hia.rwth-aachen.de
**Peter Brößner, Klaus Radermacher:** RWTH Aachen, Aachen, Germany
**Kristian Welle:** Orthopaedics and Traumatology, University Clinic Bonn, Germany

humans. In a previous work [5], our group presented an automated two-stage approach based on a concept of Beek et al. [6]: Pre-operatively, a 3D model of the scaphoid is acquired using computed tomography and the screw position is planned. Intra-operatively, volumetric ultrasound images are recorded and the scaphoid bone is segmented. Then the intra-operative model is registered to the preoperative one, allowing a transfer of the planned screw position to the current surgical setup. In a first study we proofed the feasibility of the concept in-vitro.

The contribution of this paper is the adaption of the segmentation to in-vivo ultrasound images, while maintaining its speed and automation.

# 2 Related Work

Beek et al. [6] presented the first work on the segmentation of the scaphoid in ultrasound images. While their pipeline proofed to be accurate with a surface distance error (SDE) of 0.5mm, the process incorporated manual placement of landmarks. Starting from the landmarks, they interpolate splines that are adapted to the highest intensity pixels. Given the high variability of intensity in ultrasound images, Anas et al. [4] improved on this by incorporating phase symmetry, as well as bone shadow as additional features. Still, the process required manual interaction resulting in processing times of several minutes per volume image for both methods.

While there are no other publications on the segmentation of the scaphoid, a high number of closely related methods have been published. Noble et al. [7] provide a comprehensive overview on ultrasound segmentation methods, not limited to orthopaedics. Hacihaliloglu [8] and Pandey [9] more specifically address bone segmentation and volumetric ultrasound in their survey publications. In contrast to our work, only few of the methods presented base on machine learning. In a previous publication [10], our group compared a number of state-of-the-art semantic segmentation architectures on segmentation of the femur, namely High-Resolution Net [11], Pyramid Scene Parsing Net [12], Deeplabv3+ [13] and UNet [14]. The SDE ranged from 0.56mm to 0.88mm, with Deeplabv3+ achieving the lowest error. As data set size, ultrasound machine, annotation type and other related aspects are very similar to the problem at hand, its findings most likely apply, too. Accordingly, we trained Deeplabv3+ in an in-vitro study [5] on segmentation of the scaphoid. However, this publication does not include an isolated evaluation of the segmentation error.

A critical problem for identification of well-suited methods on bone segmentation is the lack of comparability: There is no common dataset that serves as a benchmark.

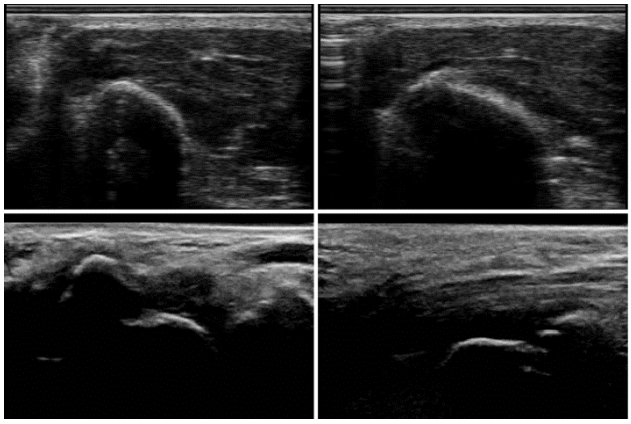

**Figure 2:** Example images of the SonixTOUCH (top row) and Affinity 50 systems (bottom row). Notice the image on the bottom right, which does not show the scaphoid, but the lunatum.

Accordingly, the reported accuracies vary strongly. On top of this, the actual performance of machine learning algorithms depends strongly on other aspects like hyperparameter tuning. Isensee et al. [15] addressed these problems by proposing a data set specific architecture, termed nnUNet. It builds on the well-known UNet but adapts the network and training parameters to the dataset at hand. Its self-adapting nature predestines it as an out-of-the-box benchmark tool.

# 3 Methods

## 3.1 Architectures for Segmentation

For the use of automated segmentation in real-time applications, fast inference times and computational efficiency are key factors. Therefore, we decided to focus on architectures for 2D segmentation instead of volumetric segmentation. For the task at hand, we selected and trained two different architectures: First, we employed the Deeplabv3+ architecture [13] an encoder-decoder structure with spatial pyramid pooling. As a backbone, we used the Mobilenetv2 [16], which is optimized for low capacity computation. This combination of performance and efficiency showed promising results in our previous work. In total, the network has 2.141.762 trainable parameters. As a reference, we utilized the 2D nnUNet [15], which is part of a framework for the automated design of medical segmentation models based on the UNet architecture [14]. This architecture includes 29.966.112 trainable parameters.

While both models exhibit an encoder-decoder structure with skip-connections, the Deeplabv3+ incorporates recent advances into its architecture: Via spatial pyramid pooling, it

incorporates multi-scale context. The pyramid features can be computed efficiently using atrous convolutions, also known as dilated convolution. By disentangling convolution along the channel and spatial dimension into a depthwise seperable convolution, the number of trainable parameters and with it the model complexity can be reduced significantly.

**Table 1:** Overview of data set splits.

| Data set | Probands | Machine | # Images |
|---|---|---|---|
| **Training** | A B | SonixTOUCH | 1620 |
| | D E F | Affinity | 1566 |
| **Validation** | C (left) | SonixTOUCH | 405 |
| | G | Affinity | 290 |
| **Test** | C (right) | SonixTOUCH | 405 |
| | H | Affinity | 290 |

## 3.2 Datasets

For the training of scaphoid segmentation in US images, we created an in-vivo dataset. This dataset consists of two subsets in order to represent different US imaging techniques. This allows for an analysis of the robustness against different technical setups for US image acquisition.

The first subset was created on a *SonixTOUCH* (Ultrasonix, Richmond, BC) device, equipped with a motorized 3D probe. We captured images of three male subjects, with five different probe poses for left and right wrists respectively. This lead to a total of 30 volume images or 2430 slice images. These were split according to subjects: two subjects for training (1620 images), one subject for validation and testing (405 images both).

For the second subset, we acquired images with an *Affinity 50* (Philips, Amsterdam, Netherlands) device equipped with a phased array 3D probe. We captured images of five different subjects, with five to ten different probe poses per subject. This resulted in a total of 37 volume images or 2146 slice images. Again, these were split corresponding to subjects: three subjects for training (1566 images), one subject for validation (290 images), and one subject for testing (290 images). These images cover a bigger volume and also include neighbouring bones. See Figure 2 for an example.

The combined dataset of 67 volume images thus consisted of 3186 images slices for training, 695 images slices for validation and 695 images slice for testing. These images were annotated manually, supported by an expert annotator. See Table 1 for additional details on proband distribution.

Besides, we used an additional in-vitro dataset as described in our previous publication [5]. This dataset consists of 2376 automatically annotated phantom images, split in 1782 images for training and 594 images for validation.

## 3.3 Training

We trained the Deeplabv3+ starting from weights pretrained on the in-vitro dataset. The models were trained for 300 epochs on a combination of dice loss and cross entropy loss, using Adam for optimization with a learning rate of 0.0001. Final models were obtained by early stopping after 189 epochs, based on best validation results regarding foreground dice. For nnUNet, training of the model was executed as specified by the framework, using stochastic gradient descent optimizer and a combination of dice loss and cross entropy loss with an initial learning rate of 0.01. The final model were obtained after training for three days, corresponding to about 900 epochs. Pretraining on the in-vitro dataset showed no benefit.

## 3.4 Evaluation

All segmentation models were evaluated using the metric implementation of the nnUNet. This includes voxel-wise metrics like precision, recall and dice as well as distance-based metrics including the average symmetric surface distance (ASSD) and the 95-percentile of the average symmetric Hausdorff distance (HD). Both distance metrics have been computed in 3D per volume image, in order to better reflect effects on an intended use case of a volumetric pipeline.

**Table 2:** Segmentation results: The 2D nnUNet outperforms the Deeplabv3+ architecture for voxel-wise metrics. For distance metrics, the Deeplabv3+ achieved lower errors.

| Metric | Deeplabv3+ | 2D nnUNet |
|---|---|---|
| **Precision** | 0.77 | 0.77 |
| **Recall** | 0.50 | 0.63 |
| **Dice** | 0.57 | 0.67 |
| **ASSD (mm)** | 0.50 | 0.66 |
| **Sym. HD 95% (mm)** | 2.39 | 3.76 |

# 4 Results

Table 2 shows test results for Deeplabv3+ and 2D nnUNet. The foreground dice coefficient appears rather low with 0.57 for the Deeplabv3+ and 0.67 for the 2D nnUNet. The distance-based metrics demonstrate a very good localization of the

prediction, with an average symmetric surface distance of 0.5mm in case of Deeplabv3+ and 0.66mm for the 2D nnUNet. Moreover, Figure 3 shows a qualitative comparison of predicted segmentations for both architectures. The Deeplabv3+ tends to miss individual slices of the scaphoid surface in both datasets, but more often in the *Affinity* data. For the nnUNet, a tendency to segmenting adjacent carpal bones as scaphoids could be observed on the *Ultrasonix* data set. For the *Affinity* data, the opposite was true: The prediction was missing slices. Segmentation of the scaphoid itself was highly precise. This holds true for images of both ultrasound machines.

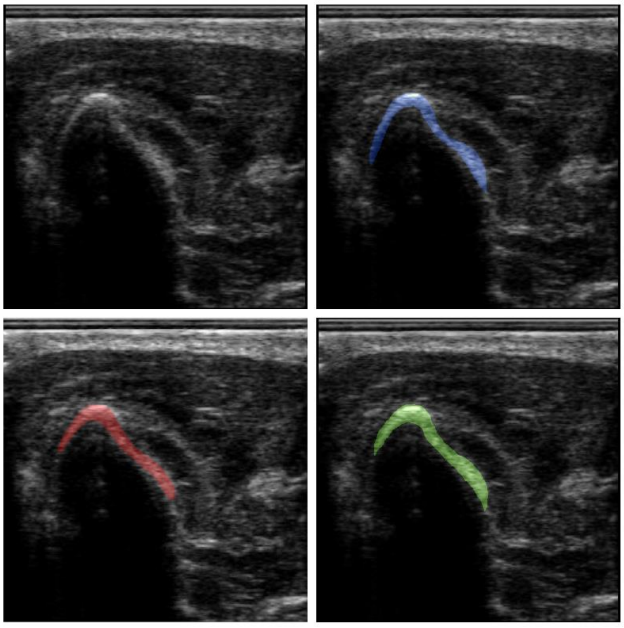

**Figure 3:** Ultrasound image depicting the scaphoid, recognizable due to its characteristic “bone shadow”. Image (upper left), GT annotation (blue, upper right), prediction of 2D nnUNet (red, lower left), prediction of Deeplabv3+ (green, lower right).

# 5 Discussion

The architectures demonstrated strong performance in segmentation of the scaphoid bone. Even more, most of the errors found originate either from isolated segmented components adjacent to the scaphoid bone or from missing the first and last slices of the scaphoid. This is due to the limitation of 2D segmentation. In some of these cases, the networks may have outperformed the human annotators: Given the small gaps between carpal bones, their delineation is challenging.

As the annotation resembles a thin line-like structure, metrics based on overlap appear to be rather low. The distance based metrics reveal promising results, which is confirmed by the qualitative analysis: Image slices depicting the scaphoid are segmented with an extremely high accuracy.

Regarding robustness to different ultrasound system, both architectures proofed to be reliable. In a direct comparison, the nnUNet dominates in terms of voxel-wise metrics, while the Deeplabv3+ shows better performance in terms of distance-based metrics. However, the nnUNet comes with an increase in inference time, which may not be feasible in the intended clinical use case.

# 6 Conclusion and Outlook

The segmentation results found are in line with previous work on segmentation of the scaphoid in ultrasound images by Beek et al. [6] and Anas et al [4]. In contrast to their approach, we presented a fully automatic and real-time capable framework. Regarding network architecture, the lightweight Deeplabv3+ achieved competitive results while offering faster inference.

Future work will focus on incorporating 3D information while maintaining a fast implementation. A rather simple approach could be a connected component analysis. However, precise delineation of the individual carpal bones may proof intractable. Alternatively, a combination of 2D CNNs with 3D point-based architectures may be able to process comprehensive spatial information with fast inference times.

As a next step, the presented work will be integrated into the full framework for intra-operative segmentation and registration of the scaphoid bone, enabling an ultrasound-based navigated surgery. Given this framework, an evaluation of the osteosyntheses screw placement as well as processing times will be performed.

Finally, as the network is trained on healthy bones only, its application is limited to certain, non-displaced fractures. The performance on displaced fractures with a visible gap in between the fragments is subject to future work.

**Author Statement**

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration.

## References


[1] Schünke M, 2014. Prometheus Lernatlas - Allgemeine Anatomie und Bewegungssystem, 4th ed., Thieme, Stuttgart.

[2] Langer MF, Oeckenpöhler S, and Breiter S, et al., 2016,“Anatomie und Biomechanik des Kahnbeins,” Orthopäde, 45(11), pp. 926–937.

[3] Singer G, 2005,“Radiation exposure to the hands from mini C-arm fluoroscopy,” The Journal of Hand Surgery, 30(4), pp. 795–797.

[4] Anas EMA, Seitel A, and Rasoulian A, et al., 2016,“Registration of a statistical model to intraoperative ultrasound for scaphoid screw fixation,” Int J CARS, 11(6), pp. 957–965.

[5] Broessner P, Hohlmann B, and Radermacher K, 2021,“Ultrasound-based Navigation of Scaphoid Fracture Surgery,” Bildverarbeitung für die Medizin 2021, 1st ed., Springer Fachmedien Wiesbaden; Imprint: Springer Vieweg, Wiesbaden, pp. 28–33.

[6] Beek M, Abolmaesumi P, and Luenam S, et al., 2008,“Validation of a new surgical procedure for percutaneous scaphoid fixation using intra-operative ultrasound,” Medical Image Analysis, 12(2), pp. 152–162.

[7] Noble JA, and Boukerroui D, 2006,“Ultrasound image segmentation: a survey,” IEEE transactions on medical imaging, 25(8), pp. 987–1010.

[8] Hacihaliloglu I, 2018,“3D Ultrasound for Orthopedic Interventions,” Advances in experimental medicine and biology, 1093.

[9] Pandey PU, Quader N, and Guy P, et al., 2020,“Ultrasound Bone Segmentation: A Scoping Review of Techniques and Validation Practices,” Ultrasound in Medicine & Biology, 46(4), pp. 921–935.

[10] Hohlmann B, Glanz J, and Radermacher K, 2020,“Segmentation of the distal femur in ultrasound images,” Current Directions in Biomedical Engineering, 6(1).

[11] Jingdong Wang, Ke Sun, and Tianheng Cheng, et al., 2019,“Deep High-Resolution Representation Learning for Visual Recognition,” TPAMI.

[12] Zhao H, Shi J, and Qi X, et al., 2017,“Pyramid Scene Parsing Network,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[13] Chen L-C, Zhu Y, and Papandreou G, et al., 2018, Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, Computer Vision - ECCV 2018, Ferrari V., Hebert M., Sminchisescu C., and Weiss Y., eds., Springer International Publishing, Cham, pp. 833–851.

[14] Ronneberger O, Fischer P, and Brox T, 2015, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015: proceedings, Navab N., Hornegger J., Wells W. M., and Frangi A. F., eds., Springer, Cham, pp. 234–241.

[15] Isensee F, Jaeger PF, and Kohl SAA, et al., 2021,“nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” Nat Methods, 18(2), pp. 203–211.

[16] Sandler M, Howard A, and Zhu M, et al., 2018,“MobileNetV2: Inverted Residuals and Linear Bottlenecks,” CVPR 2018, IEEE Computer Society, Los Alamitos, California.