

# Ultrasound-based Navigation of Scaphoid Fracture Surgery

Peter Broessner<sup>1,2</sup>, Benjamin Hohlmann<sup>1,2</sup>, Klaus Radermacher<sup>2</sup>

<sup>1</sup>Both authors contributed equally

<sup>2</sup>Chair of Medical Engineering, RWTH Aachen

`hohlmann@hia.rwth-aachen.de`

**Abstract.** For minimally-invasive surgery of the scaphoid, navigation based on ultrasound images instead of fluoroscopy could reduce costs as well as prevent exposure to ionizing radiation. We present a machine learning based two-stage approach that tackles the tasks of image segmentation and point cloud registration individually. For this, Deeplabv3+ as well as the PRNet architecture were trained on two newly generated datasets. An evaluation on in-vitro data results in an average surface distance error of 1.1 mm and a mean rotational deviation of 6.2° with a processing time of 9 seconds. We conclude that near real-time navigation is feasible.

## 1 Introduction

Of all carpal bones, the scaphoid is the most frequently fractured one, accounting for about 60% of all fractures [1]. For diagnosis of fractures a comprehensive exam including bi-planar radiography as well as computed tomography (CT) and possibly magnetic resonance imaging (MRI) and ultrasound (US) is standard. Given this rich image based pre-operative information, the decision upon conservative treatment using a cast, or operative treatment is based on the stability of the fracture. Fractures of the proximal third as well as displaced fractures indicate an operative treatment. Stable or non-displaced cases may also be treated operatively to fasten the recovery [1].

Surgery can be performed in an open as well as minimally-invasive fashion. While strongly dislocated cases require an open surgery, minimally-invasive surgery (MIS) is recommended whenever possible due to minimized operative trauma, preservation of carpal ligaments and faster recovery. During surgery, the bone fragments are united using an osteosynthesis screw. The exact placement of this screw is crucial for surgical success. In MIS, placement and validation, which takes place under continuous fluoroscopy, is a challenging task due to the limited spatial perception of the three-dimensional position in the two-dimensional projected radiographs. Furthermore, the patient and surgeon are exposed to ionizing radiation. Therefore, this work investigates ultrasound as a cheap and readily available alternative to fluoroscopy. Yet, ultrasound is limited in terms of signal-to-noise ratio as well as occlusion.

Intra-operative registration of ultrasound images to pre-operatively acquired models is a common concept in navigated surgery. For surgery of the scaphoid, which poses a hard problem due to the small size of the bone, several authors proposed concepts and validated them in in-vitro, ex-vivo as well as in-vivo studies. The earliest procedure, proposed by Beek et al., involves a semi-automatic heuristic, requiring the user to set seed points. Subsequently, the pre-operative plan is manually aligned to the intra-operative ultrasound image and the position is refined using the iterative closest point algorithm (ICP). While the method proved viable regarding realization of the surgical plan, it requires manual interaction with reported times of 5-10 minutes [2]. Following a breakthrough of ultrasound segmentation techniques, Anas et al. improved the procedure by incorporating phase symmetry pre-processing as well as statistical shape and pose models into the segmentation process [3]. They enhance the symmetric high intensity interfaces, like the bone surface, in the ultrasound image by computing the phase symmetry. To distinguish bone from soft-tissue interfaces, the bone's shadow is incorporated as an additional feature. After that, a statistical shape and pose model of all carpal bones is manually aligned to the ultrasound image. The alignment is optimized in an Expectation Maximization framework using Gaussian Mixture Models. This algorithm reduces the manual interaction while at the same time improving the registration accuracy. They evaluated their technique in in-vitro [3], ex-vivo [4] and in-vivo [3] studies and achieved a processing time of about 90 seconds.

In recent years, the computer vision community achieved great advances in automatic semantic segmentation of the bone surface in in-vivo ultrasound images. Pandey et al. reviewed 56 articles on this specific task [5]. Most of the publications included fully automatic methods with a clear tendency to machine learning based approaches in the recent past.

Given the success of Convolutional Neural Networks (CNN) on images the concept was transferred to point sets. Wang et al. used a graph based approach for convolution-like computations [6]. Their Dynamic Graph CNN (DGCNN) is the backbone used in the Partial Registration Network (PRNet), a machine learning based architecture for partial point set registration [7].

In this work, we present the first fully automatic as well as near-real time capable algorithm for ultrasound based navigation of scaphoid fracture surgery. We further proof its feasibility in an in-vitro study. We propose a two-stage architecture, tackling the problems of segmentation and registration individually. As machine learning based segmentation is a well studied problem, we focus on evaluating the registration.

## 2 Materials and Methods

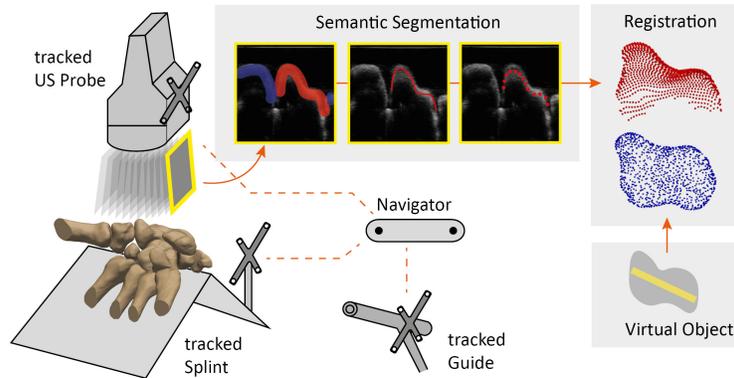
In order to allow a navigated fixation of scaphoid fractures, a preoperative virtual object, including the surgical plan, has to be intra-operatively registered to the therapeutical object. For this purpose, the aforementioned two-stage approach is proposed (see fig. 1): in a first step, a tracked 3D US probe is used for

the acquisition of slice images, which are subsequently segmented by a neural network. The pixels labeled as scaphoid surface are then skeletonized and uniformly sampled to obtain a surface point set (depicted in red). In a second step, the point set of the source model (depicted in blue) is registered to this sampled point set, again using a neural network.

## 2.1 Architectures

For the task of semantic segmentation, the DeepLabv3+ [8] architecture is selected. It is characterized by an encoder-decoder structure with atrous separable convolution for spatial pyramid pooling. In combination with a MobileNetv2 [9] backbone it offers a compromise between performance on the one hand and a reduced number of trainable parameters on the other hand, which is favorable given the rather small size of the training data set.

The subsequent task of registration is quite challenging: Points sampled from a partial surface have to be registered to points representing the complete surface, without real point correspondences and disturbed by errors of segmentation. The task is further complicated by the fact that the use of shared architectures for feature extraction requires the point sets to be of equal size, which in our case leads to different spatial resolutions. To meet these challenges, the PRNet architecture in combination with a DGCNN backbone seems most promising and is employed in the course of this work. The DGCNN utilizes a convolution-like learning of filters on dynamically updated k-nearest neighbors for the extraction of local and global point features; in PRNet, these feature vectors are co-contextualized by a Transformer, which in combination addresses the difficulty of partial-to-full registration. Furthermore, PRNet aims at establishing non-bijective correspondences with variable sharpness by using gumbel softmax, which addresses the lack of real correspondences and the difficulty of different spatial resolutions.



**Fig. 1.** Intra-operative procedure for the registration of therapeutical object and virtual object: the US image slices acquired by a 3D US probe are fed into a segmentation network. The resulting masks are thinned and sampled to a partial point set, which is registered to the point set obtained from CT.

## 2.2 Datasets

In order to train models for the tasks of semantic segmentation and registration, two datasets are created. The first dataset is created for semantic segmentation of carpal bones in US images. It is based on four printed carpal phantoms, two male and two female, and consists of automatically annotated US phantom images. For the automated annotation, tracked phantoms are placed in a water basin, where a tracked 3D US probe is then used for the acquisition of 22 volume images per phantom, with 81 image slices each. By transforming the respective carpal model to these US volume images of the carpal phantom, a surface annotation is generated. Since neighboring slice images are very similar, only every third slice is included, resulting in a total of 2376 annotated US images. These are split according to the underlying wrist phantom for the creation of similar composed and hence comparable datasets. With four wrist phantoms available, images are split in 1782 images (three phantoms) for training and 594 images (one phantom) for validation and testing.

The second dataset is created for the training of point-based scaphoid registration. It is based on 105 scaphoid models provided by Moore et al. [10], which were generated from CT images of both male and female patients. From these 105 models, a statistical shape model (SSM) is derived in order to obtain a greater variety of data. The resulting dataset consists of pairs of aligned point sets with equal sizes of 1024 points. For each of these pairs of point sets, the first set is derived from the SSM with variances in the range of  $\pm 2$  standard deviations (SD), while the second is generated by synthetic sampling of the first set, which imitates US imaging. The dataset contains about 74,000 pairs of data, divided into about 41,000 pairs for training and about 16,500 pairs for validation and testing respectively. Fig. 1 exemplarily shows a pair of point sets from the created dataset in the registration section.

## 2.3 Training

For training of the segmentation model, weights pretrained on the PASCAL VOC dataset are used as initialization. Using Adam for optimization and a set of hyperparameters derived from grid search, the final segmentation model is obtained by early stopping after 156 epochs based on results on the combined validation/test set.

For training of the registration model, ground truth (GT) has to be generated from the aligned pairs of point sets by applying a random transform to the sampled point set. This random transform consists of a rotation around each axis uniformly sampled from  $[0^\circ, 45^\circ]$ , and a translation uniformly sampled from  $[-25\%, 25\%]$  of object size. Again, Adam is employed for optimization, with hyperparameters determined by grid search; the final registration model is obtained by early stopping after 14 epochs based on results on the validation set.

**Table 1.** Rotational and translational registration errors on point sets derived from segmentation GT and segmentation results, with mean and SD respectively.

	Point sets derived from GT		Point sets derived from Segmentation	
	MAE(R) / °	MAE(t) / mm	MAE(R) / °	MAE(t) / mm
Initial	23.17 ± 7.14	3.29 ± 1.09	22.02 ± 7.35	3.72 ± 1.34
ICP	24.68 ± 14.05	2.23 ± 1.42	22.77 ± 13.62	2.88 ± 1.82
PRNet	5.29 ± 3.79	0.92 ± 0.47	10.22 ± 7.37	1.73 ± 1.16
PRNet+ICP	1.42 ± 3.94	0.13 ± 0.25	6.20 ± 8.80	0.72 ± 1.50

## 2.4 Testing

Test results for registration are reported for two different test scenarios, which are based on the segmentation validation/test dataset: registration results on point sets derived from GT, in comparison to results of registration on point sets derived from segmentation results. For each of the two test scenarios, initial errors are compared to registration results of ICP and PRNet. Moreover, results of a combination of PRNet and ICP are included, with ICP starting from the estimated transformation of PRNet. Results of registration are measured by means of mean absolute error (MAE) between GT and predicted transformation, decomposed into a rotational error MAE(R) and a translational error MAE(t). Furthermore, the surface distance error (SDE) is computed as a point to surface distance. All experiments are repeated 10 times, results are reported as mean and SD.

## 3 Results

Test results for registration can be seen in Tab. 1, with results on point sets derived from GT in the left column, and results of registration on point sets derived from segmentation results in the right column. The SDE after registration of GT is  $0.49\text{ mm} \pm 0.02\text{ mm}$  and after registration of predicted segmentations is  $1.10\text{ mm} \pm 0.86\text{ mm}$ . Computation times for the whole process add up to  $9.09\text{ s} \pm 0.89\text{ s}$ , of which the major part is attributed by segmentation with  $7.70\text{ s} \pm 0.75\text{ s}$ , while only  $0.21\text{ s} \pm 0.03\text{ s}$  are needed for registration.

## 4 Discussion

The proposed two-stage approach removes the need for manual interaction while simultaneously reducing the processing time to 9 seconds. This is an at least ten-fold improvement over previous methods [2,3,4]. Our evaluation results in an axis deviation of  $6.2^\circ$  MAE and  $1.1\text{ mm}$  SDE, which is roughly equal to  $5^\circ$  absolute deviation and  $1\text{-}1.2\text{ mm}$  SDE reported by Anas et al [3]. In an ex-vivo evaluation of their method, Anas et al. successfully performed 10 out of 13 screw placements [4]. Thus, for clinical application, the overall error needs to be reduced further. Additionally, given the limitation to non-displaced fractures, only few patients

could receive an ultrasound-based treatment, yet. Extending the application to displaced fractures requires an evaluation of bone fragment registration as well as visibility. An obvious limitation of this work is the evaluation on in-vitro data, which is a comparably simple task. Finally, the absence of an independent test set limits the significance of the segmentation evaluation of this study.

As shown in Tab. 1, scaphoid registration poses a difficult task, as the gold standard algorithm ICP is not able to converge to the global minimum solution. Combining it with a machine learning based global prior registration however, our approach achieves significant improvements. The segmentation on the other hand is not yet sufficiently fast and precise, as can be concluded from the high errors when processing segmented point sets. Future work will therefore focus on improving the first stage: A preceding classifier may reduce the number of false positive segmentations. Lightweight architectures designed for real-time segmentation could speed up the computation. Additionally, the pipeline needs to be adapted to and evaluated on in-vivo data.

**Acknowledgement.** We would like to thank Moore et al. for providing carpal and scaphoid models.

## References

1. Mehling IM, Sauerbier M. Skaphoidfrakturen und Skaphoidpseudarthrosen. *Z Orthop Unfall*. 2013;151(6):639–660.
2. Beek M, Abolmaesumi P, Luenam S, et al. Validation of a new surgical procedure for percutaneous scaphoid fixation using intra-operative ultrasound. *Med Image Anal*. 2008;12(2):152–162.
3. Abu Anas EM, Seitel A, Rasoulian A, et al. Bone enhancement in ultrasound using local spectrum variations for guiding percutaneous scaphoid fracture fixation procedures. *Int J Comput Dent*. 2015;10(6):959–969.
4. Anas EMA, Seitel A, Rasoulian A, et al. Registration of a statistical model to intraoperative ultrasound for scaphoid screw fixation. *Int J Comput Assist Radiol Surg*. 2016;11(6):957–965.
5. Pandey PU, Quader N, Guy P, et al. Ultrasound bone segmentation: A scoping review of techniques and validation practices. *Ultrasound Med Biol*. 2020;46(4).
6. Wang Y, Sun Y, Liu Z, et al. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans Graph*. 2019;38(5):1–12.
7. Wang Y, Solomon JM. PRNet: Self-Supervised Learning for Partial-to-Partial Registration. In: Wallach H, editor. *Adv Neural Inf Process Syst* 32. Curran Associates, Inc.; 2019. p. 8814–8826.
8. Chen LC, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*; 2018. p. 801–818.
9. Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*; 2018. p. 4510–4520.
10. Moore DC, Crisco JJ, Trafton TG, et al. A digital database of wrist bone anatomy and carpal kinematics. *J Biomech*. 2007;40(11):2537–2542.