Towards a learning-based CT segmentation of acetabular fractures

Andy Zhang, Mehdi Boudissa, Maxime Nemo, Jérôme Tonetti, Matthieu Chabanas

Towards a learning-based CT segmentation of acetabular fractures

Andy Zhang[a], Mehdi Boudissa[a,b], Maxime Nemo[a], Jérôme Tonetti[a,b], and Matthieu Chabanas[a]

[a]Université Grenoble Alpes, CNRS, Grenoble INP, TIMC; F-38000 Grenoble, France
[b]Grenoble University Hospital, Department of Orthopaedic and Traumatology Surgery; F-38000 Grenoble, France

ABSTRACT

Fractures of the *acetabulum*, the cavity of the hip that hosts the femoral head, are complex to understand, plan, and surgically reduce. Segmenting bone fragments in CT scans is fundamental for assisting surgeons in their therapeutically process, and can benefit from recent learning-based advances. In this paper, we extended a learning-based network for the semantic segmentation of 6 pelvic bones: left and right hip, left and right femur, sacrum, and lumbar spine. This semantic segmentation is then process by a surgeon to separate fracture fragments, similarly to an existing baseline process. Results on 6 fracture cases show a qualitative improvement of the final fragment segmentation quality. Mostly, the segmentation time is statistically significantly reduced from 94 min to 18 min, in mean, which is a promising step towards using such learning-based method in preoperative clinical routine.

Keywords: Semantic segmentation, Neural networks, Fracture segmentation, Acetabulum, Orthopaedic surgery.

1. INTRODUCTION

Medical image segmentation is an essential process in many image-guided procedures. This is especially the case for acetabular surgery, where the segmentation of bone fragments is a core step to assist in analysing the fracture, planning the surgery, and for guiding procedures.[1–5] Numerous methods, usually semi-automatic, have been proposed to segment fractured pelvis with standard segmentation method like region growing, wavelets, or active shape modeling.[6–8] With the advent of Deep Learning as a state-of-the-art segmentation approach, several authors have addressed bone segmentation in non-fracture contexts.[9] Still, only a few works are dedicated to segmenting the pelvic area[10] or localizing pelvic fracture lines.[11,12]

In previous studies[3,5] we proposed a semi-automatic segmentation method. Bones were first segmented using region growing, followed by significant manual refinements. After this *semantic segmentation*, the different bone fragments were manually separated which corresponds to an *instance segmentation*. The objective of this study was to replace the semantic segmentation phase by a fully-automated learning-based segmentation, then to evaluate its impact on the instance segmentation phase and in terms of total segmentation time.

We extended the cascade 3D U-Net framework proposed by Liu et al. (2021)[10] for the automatic semantic segmentation of 6 bone structures of the pelvis: left & right hip, left & right femur, sacrum, and lumbar spine. The fracture fragments were thus segmented manually as in our baseline procedure. Results on 6 fracture cases show a clear qualitative improvement of the final fragment segmentation quality. Mostly, the segmentation time is statistically significantly reduced from 94 min to 18 min, in mean, which is a promising step towards using such learning-based method in preoperative clinical routine.

Contact: mboudissa@chu-grenoble.fr, matthieu.chabanas@univ-grenoble-alpes.fr

2. MATERIAL AND METHODS

2.1 Datasets

The initial neural network from Liu et al. (2021)[10] was re-trained/tested with 110 CT scans and labels from their open source CLINIC and CLINIC-metal datasets available at https://github.com/ICT-MIRACLE-lab/CTPelvic1K. Most of the cases from these datasets have a pelvic ring fracture, and about 10% of them have an acetabular fracture. The CLINIC-metal set also contains post-operative CT with metal artifacts generated by the fixation screws and plates.

In addition to these publicly available datasets, we included 6 pre-operative CT scans from the Grenoble University Hospital, France. All cases have a fracture of the *acetabulum*, sometimes combined with a pelvic ring fracture. Patients signed an informed consent prior to surgery, with the approval of the local ethics committee (RCB 2022-A01472-41).

2.2 Baseline segmentation method

In our previous studies,[5,13] we proposed a semi-automatic procedure based on the open-source application ITK-SNAP (www.itksnap.org).[14] This procedure consists in successive stages:

1. the segmented volume is first cropped around the region of interest: the fracture fragments and the bones connected to these fragments through ligaments. Non-relevant structures are ignored.

2. A *semantic segmentation* is then performed to identify bone tissue. After a thresholding initialization, seeds are manually set to initiate a region growing algorithm.[14] This phase has proven to be quite difficult as the method tends to stop on every discontinuity, not only the fracture lines but the discontinuities artifically created by the thresholding operation in areas where the bone is very thin. Thus, a significant number of initial seeds is required. Many leeks also occur along cancellous bone, where the signal is never clearly defined.

3. Manual refinement is always necessary to clean the semantic segmentation, expand under-segmented areas, and trim over-segmented ones.

4. Finally, the different fragments are manually separated. This stage is analogous to an *instance segmentation*.

As reported over 10 patients,[5] the total segmentation time was 82 ± 18 minutes [range 60–120]. For the 6 cases selected for this study, the mean segmentation time was 94 minutes.

2.3 Learning-based semantic segmentation

For the semantic segmentation of the pelvic bones, we adapted the open-source deep learning network proposed by Liu et al. (2021).[10] This network is composed a two 3D nnU-Net[15] working in cascade on low resolution and then full resolution CT volumes. The 4-class output is then post-processed based on a Signed Distance Function (SDF) filtering[16] to avoid the removal of isolated fractured bone fragments from the segmentation.

The initial multi-class network labels 4 classes referred as the left hip, right hip, sacrum, and lumbar spine. A fundamental structure to plan surgery is the femoral head, which is tighly connected to the *acetabulum* via the acetabular capsule. In the case of a fracture, this ligament capsule can be partially torn off and the femoral head can be in contact with surrouding bone fragments or even be dislocated from the joint. We therefore added two classes to the network's output and annotation set: left femur and right femur.

The network modification was straightforward as we simply increased the number of classes from 4 to 6 in the decoder of the low resolution 3D nnU-Net and in both the encoder and decoder of the full resolution U-Net. However, due to the change of dimension of the layers, we could not simply refine or transfer the network weights from the original paper and instead had to fully retrain the network.

Similarly to Liu et al. (2021),[10] an iterative procedure was followed to create the femoral annotations. Both femurs were first manually segmented on 20 cases using ITK-SNAP (www.itksnap.org).[14] The network was then trained with these cases, and used to infer femur annotations in 10 new cases. After manual refinement, the

10 corrected annotations were added to the training set. Note that only the femur annotations were conserved from the last inference, and merged with the original 4-class open source annotations. This process was repeated to obtain a 6-class annotation on 70 cases, when basically no manual refinement were needed for the femurs. These 70 cases came from the publicly available CLINIC and CLINIC-metal datasets. The 40 remaining cases from these datasets were left for testing.

2.4 Experiments and evaluation protocol

We studied how the learning-based semantic segmentation could benefit the overall segmentation process. The semantic network replaced the first steps of our baseline method (section 2.2) and we simply performed the fragments identification and separation starting from the network output inference.

After training the 6-class segmentation network on public images, we used the network to automatically infer the classes on 6 cases from the Grenoble University Hospital. Then, an orthopaedic surgeon manually segmented the instances from the network output. The 6 cases had already been segmented for a previous study, by the same surgeon, so we could compare the different results. Three models were compared:

baseline manual instance segmentation after a semi-automatic, region-growing segmentation (section 2.2).

1-class semantic manual instance segmentation from the automatic, 1-class semantic segmentation. The 6 output classes of the network were re-labelled with a single "bone" label, to mimic the output of the baseline region-growing method.

6-class semantic manual instance segmentation from the automatic, 6-class semantic segmentation (left and right hip, left and right femur, sacrum, lumbar spine).

Two metrics were considered. We first evaluated the quality of the different models. We could not use a quantitative metric such as the DICE coefficient since the baseline models were not complete (only the regions of interest were segmented with refinement, others regions were roughly curated or totally discarded). We thus only verified that the number of fracture fragments was the same, and then we qualitatively evaluated the quality of the segmented instances. The second metric was the total curation time. The semantic segmentation was automatically performed on a DELL Ampere server, then when added the time needed for the manual instance segmentation.

3. RESULTS AND DISCUSSION

3.1 Semantic segmentation

The modified network was first evaluated. The model was retrained on 70 cases of the 110 cases segmented with 6 classes, and tested on the remaining 40 cases. The Dice coefficients for the initial classes (left and right Hip, sacrum, and lumbar spine) were all between .98 and .99 in mean, which is no different than the results reported in.[10] This was unexpected as only 70 volumes were used to retrain the model instead several hundreds. Similar Dice coefficients were obtained for the left and right Femur, which validated the extension of the model from 4 to 6 classes.

While these results are already excellent, we plan to continue the iterative process to segment femurs on all public images to fully benefit from the size of the datasets and increase the robustness of the semantic network.

3.2 Qualitative comparison of the segmentations

Unfortunately, the baseline and network-based results could not be quantitatively compared with metrics such as the Dice coefficient or Haussdorf distance. The reason is that baseline segmentations are limited to an area around the fracture, while the whole pelvis is systematically segmented with the network. In addition, areas far the fracture or irrelevant for the simulation planning were not refined during the baseline process. For example, the femur was never separated from the *acetabulum* on the side opposite the fracture.

The different results were thus qualitatively compared by the surgeon, yielding the following comments:

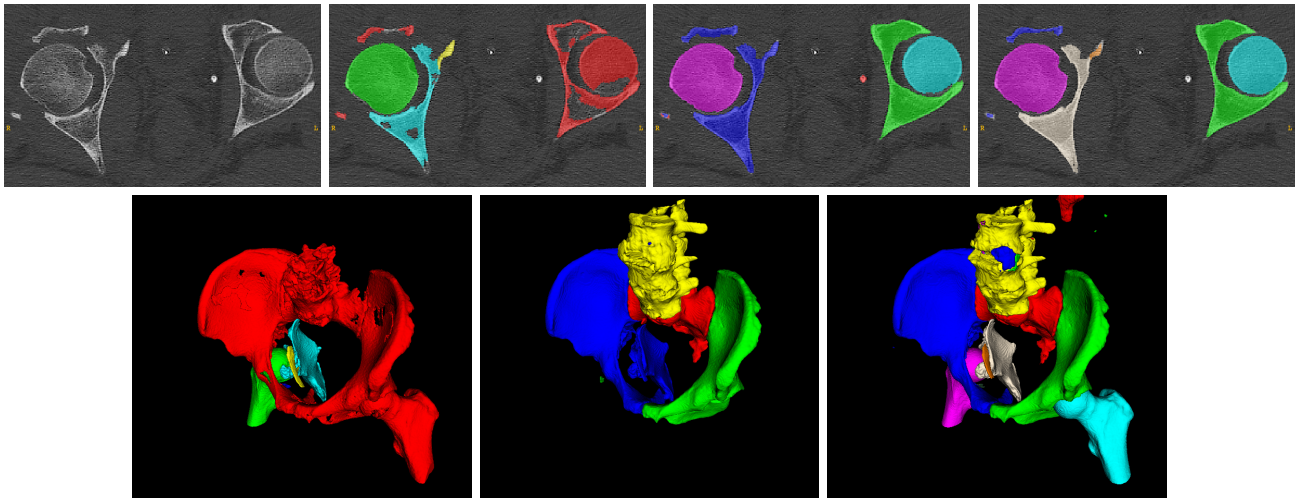

Figure 1. Segmentation of case 2. Top row, from left to right: CT images, baseline segmentation, automatic 6-class segmentation, and final fracture segmentation after fragment separation. Bottom row: 3D view of the baseline, 4-class segmentation from Liu et al. (2021), and final segmentation.

- The segmentation is always clearly improved for all cases.
- First, classes already labelled do not have to be manually separated which is a direct benefit. This is especially relevant for the femurs, the sacrum, and the pubic symphysis. Note that the femur was correctly separated even in cases with a severe dislocation (see Figure 2), even if such configuration was never present in the training dataset.
- The overall 3D aspect is more pleasing since the pelvis is totaly segmented, and "holes" in areas where the cortical bone is thin (e.g. in the middle of the iliaque) are much smaller if not absent (see Figure 1).
- The semantic segmentation was not perfect, though, and several small fragments or extremities were either over- (cases 1 & 3) or under-segmented (cases 4 & 5). These defects still had to be manually refined similarly to the baseline procedure.
- Finally, unseparated fragments with a common label after the 6 class-segmentation (e.g. left hip) were easier to separate than from the baseline region growing output.

3.3 Computation times

Table 1 reports the computation times for all segmentation models, on 6 fracture cases from our center. The time analysis yields:

- The automatic 6-class semantic network took 3 to 6 minutes to perform a case, with a strong correlation with the number of slices in the CT scan.
- Separating the fragment instances from the semantic 1-class inferences took 24.5 minutes in mean. Even after adding the network computation time, this is a statistically significant reduction compared to the baseline time.
- this reduction is even more significant when fragments are segmented from the 6-class inferences, with a segmentation time reduced to 13 minutes in mean from more than 90 minutes.

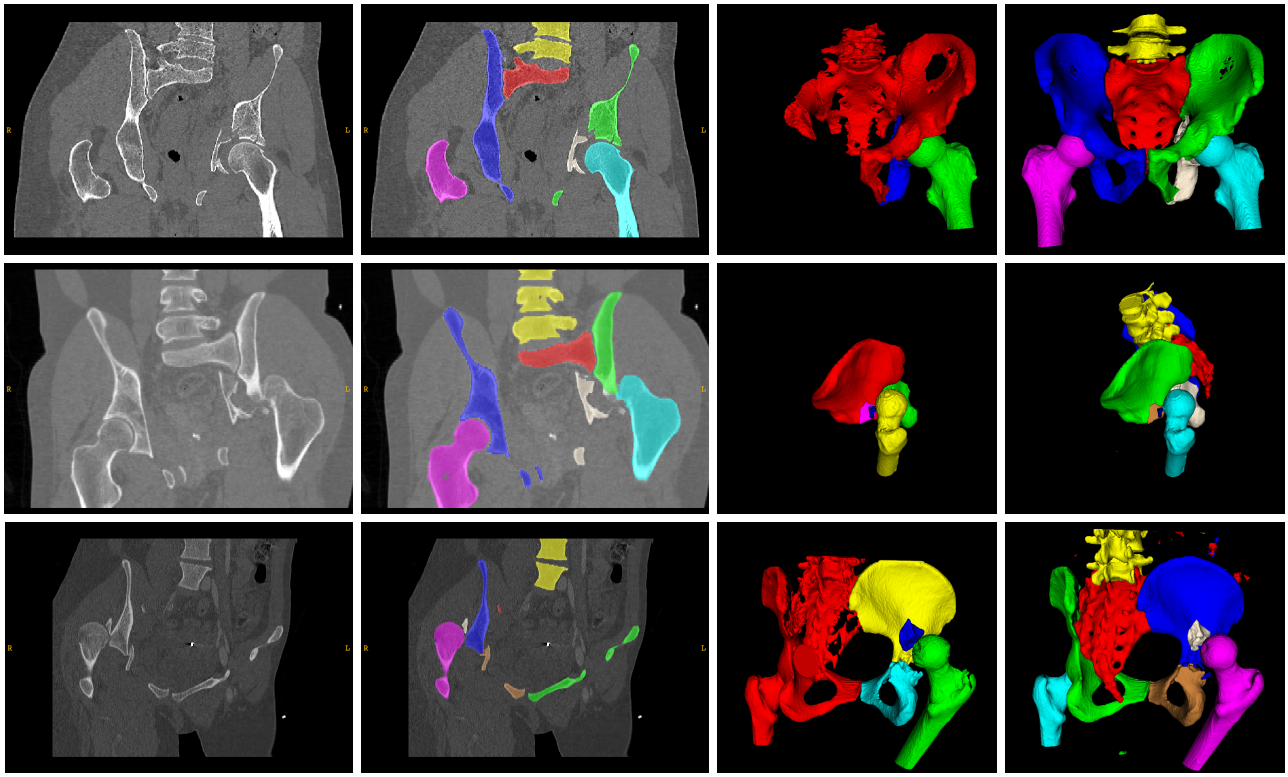

Figure 2. Segmentation of cases 1, 6, 3. From left to right: CT images with final segmentation, 3D view of the baseline and final segmentation.

Table 1. Computation times for the different segmentation methods: baseline, instance segmentation from the automatic 1-class semantic segmentation, instance segmentation from the automatic 6-class semantic segmentation.

Case	# slices	# labels	baseline (min)	semantic network (min)	from 1-class semantic (min)	from 6-class semantic (min)
1	458	7	90	6.12	30	18
2	489	9	70	5.36	22	15
3	236	9	120	3.35	20	8
4	259	8	100	5.17	25	12
5	430	7	90	5.43	35	17
6	137	9	-	3.19	15	8
mean computation time (min)			94	4.77	24.5	13

4. CONCLUSION

Automatizing the semantic segmentation of pelvic bones with a learning-based network has proven to be a significant step for the segmentation of bone fractures. First, the quality of the fracture segmentation is consistently improved. However, the main benefit is the significant reduction of the total segmentation time, bringing the method towards a use in routine clinical planning.

Future works include completing the femur segmentation of the public datasets, and potentially improving the SDF post-processing[10] to better segment small comminuted fragments and thin extremities. In a longer term, we plan to investigate the automatization of fragments separation using graph cut methods or panoptic segmentation networks such as Mask R-CNN.[17]

ACKNOWLEDGMENTS

This work has been supported by the French National Research Agency through the frameworks "Computer Assisted Medical Interventions" (ANR-11-LABX-0004 CAMI Labex) and by the "Multidisciplinary Institute in Artificial Intelligence" (ANR-19-P3IA-0003 MIAI@Grenoble Alpes).

REFERENCES

[1] Cimerman, M. and Kristan, A., "Preoperative planning in pelvic and acetabular surgery: the value of advanced computerised planning modules," *Injury* **38**(4), 442–449 (2007).

[2] Fornaro, J., Keel, M., Harders, M., Marincek, B., Székely, G., and Frauenfelder, T., "An interactive surgical planning tool for acetabular fractures: initial results," *J Orthop Surg Res* **5**(1), 50–55 (2010).

[3] Boudissa, M., Orfeuvre, B., Chabanas, M., and Tonetti, J., "Does semi-automatic bone-fragment segmentation improve the reproducibility of the letournel acetabular fracture classification?," *Orthopaedics & Traumatology: Surgery & Research* **103**(5), 633–638 (2017).

[4] Boudissa, M., Courvoisier, A., Chabanas, M., and Tonetti, J., "Computer assisted surgery in preoperative planning of acetabular fracture surgery: state of the art," *Expert Review of Medical Devices* **15**(1), 81–89 (2018). PMID: 29206497.

[5] Boudissa, M., Noblet, B., Bahl, G., Oliveri, H., Herteleer, M., Tonetti, J., and Chabanas, M., "Planning acetabular fracture reduction using a patient-specific biomechanical model: a prospective and comparative clinical study," *Int J CARS* **16**, 1305–1317 (2021).

[6] Vasilache, S., Ward, K., Cockrell, C., Ha, J., and Najarian, K., "Unified wavelet and gaussian filtering for segmentation of CT images; application in segmentation of bone in pelvic CT images," *BMC Medical Informatics and Decision Making* **9**, S8 (Dec. 2009).

[7] Fornaro, J., Székely, G., and Harders, M., "Semi-automatic segmentation of fractured pelvic bones for surgical planning," in [*Biomedical Simulation*], Bello, F. and Cotin, S., eds., 82–89, Springer Berlin Heidelberg, Berlin, Heidelberg (2010).

[8] Wu, J., Davuluri, P., Ward, K. R., Cockrell, C., Hobson, R., and Najarian, K., "Fracture Detection in Traumatic Pelvic CT Images," *International Journal of Biomedical Imaging* **2012**, 1–10 (2012).

[9] Klein, A., Warszawski, J., Hillengaß, J., and Maier-Hein, K. H., "Automatic bone segmentation in whole-body CT images," *Int. Journal of Computer Assisted Radiology and Surgery* **14**, 21–29 (Jan. 2019).

[10] Liu, P., Han, H., Du, Y., Zhu, H., Li, Y., Gu, F., Xiao, H., Li, J., Zhao, C., Xiao, L., Wu, X., and Zhou, S. K., "Deep learning to segment pelvic bones: large-scale CT datasets and baseline models," *International Journal of Computer Assisted Radiology and Surgery* **16**, 749–756 (May 2021).

[11] Wang, Y., Lu, L., Cheng, C.-T., Jin, D., Harrison, A. P., Xiao, J., Liao, C.-H., and Miao, S., "Weakly supervised universal fracture detection in pelvic x-rays," in [*Lecture Notes in Computer Science*], 459–467, Springer International Publishing (2019).

[12] Ukai, K., Rahman, R., Yagi, N., Hayashi, K., Maruo, A., Muratsu, H., and Kobashi, S., "Detecting pelvic fracture on 3D-CT using deep convolutional neural networks with multi-orientated slab images," *Scientific Reports* **11**, 11716 (Dec. 2021).

[13] Boudissa, M., Oliveri, H., Chabanas, M., and Tonetti, J., "Computer-assisted surgery in acetabular fractures: Virtual reduction of acetabular fracture using the first patient-specific biomechanical model simulator," *Orthopaedics & Traumatology: Surgery & Research* **104**, 359–362 (May 2018).

[14] Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., and Gerig, G., "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage* **31**(3), 1116–1128 (2006).

[15] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H., "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods* **18**(2), 203–211 (2021).

[16] Perera, S., Barnes, N., He, X., Izadi, S., Kohli, P., and Glocker, B., "Motion segmentation of truncated signed distance function based volumetric surfaces," in [*2015 IEEE Winter Conference on Applications of Computer Vision*], 1046–1053 (2015).

[17] Abdulla, W., "Mask r-cnn for object detection and instance segmentation on keras and tensorflow." https://github.com/matterport/Mask_RCNN (2017).