



OPEN Automated segmentation of the fibula from CT imaging using two-stepped deep learning in 3D U-Net architectures

Jônatas de Souza Nascimento^{2,3}, Tobias Pankert^{1,3}, Florian Peters¹, Frank Hölzle¹, Ali Modabber^{1,3}, Mathias Wien² & Stefan Raith^{1,3}✉

This study proposes a fully automatic segmentation of the fibula bone from CT images for application in pre-operative planning of reconstructive surgery. The objective is to make use of new developments in the image segmentation field to optimize and reduce the costs of patient-specific surgery planning. Two different approaches are proposed to perform the fibula bone segmentation, both based on a two-step segmentation method using a 3D-UNet architecture. To account for the symmetry of the left and right fibula bones, input images of the right fibula are mirrored to the left side. The accuracy of the trained models is measured using common evaluation metrics, together with specific metrics focused on facial reconstructive surgery. Both of the described approaches achieve high-accuracy results. For the best-trained model, an average Dice score of 0.95 and Average Surface Distances below 0.31 mm is measured on the test set in the region of interest for the surgery. Both approaches are robust segmentation techniques and permit data pre-processing for further application in the context of preoperative surgical planning of procedures for facial reconstruction with bony transplants.

In recent years, Convolutional Neural Networks (CNNs) have been used for different tasks in the medical field, including detecting tumors¹, COVID-19 diagnosis², and improving patient-specific individualization of treatments³. These exemplary applications show the enormous potential of using Deep Learning methods to improve different tasks in the medical field. There are, however, several applications in this domain for which the possibilities of using Artificial Intelligence-based techniques have not been explored yet. The planning of facial reconstructive surgery is an example where essential tasks could be potentially improved with such novel methods⁷. Until now, few works on that topic have been published in literature, with only one publication known so far that deals with the segmentation of the iliac crest (i.e. a part of the pelvic bone) with the aim of use in reconstructive facial surgery⁴. There has been a remarkable increase in the use of computer tools to assist surgeons in the surgery planning procedure, which is leading to essential transformations in the field^{5,6}. The use of Deep Learning techniques, commonly used in medical imaging, has remained scarcely studied with respect to an application in a surgical context. This can be seen in the facial reconstructive surgery planning context. Even though the use of 3D models of the donor bones during the planning procedures is shown to yield positive outcomes⁷, the usual process to acquire those 3D models is still highly manual, and time-consuming⁸, and could potentially be automated with Neural Network-based techniques. Even though fibular segmentation may appear less challenging compared to more complex anatomical structures, such as the mandible, an accurate representation of the fibular bone from CT scans is of paramount importance for a precise patient individual planning of reconstructive surgery. Although manual segmentation is feasible, it remains time-consuming and specifically subject to inter-operator variability, and thus bottleneck in an otherwise largely automated process routine of digital surgical planning. Thus, an automatic and hence repeatable and fast segmentation method, is highly valuable for clinical application.

While various fully automatic CNN-based approaches for the segmentation of different bones are described in literature^{9–11}, segmentation of the fibula bone has not been investigated, despite its major significance as a donor bone for grafting procedures^{12,13}.

¹Department of Oral and Maxillofacial Surgery, RWTH Aachen University Hospital, Aachen, Germany. ²Institute for Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany. ³Inzipio GmbH, Aachen, Germany. ✉email: sraith@ukaachen.de

In this study, we propose a fully automatic CNN-based approach to perform the segmentation of the fibula bone from CT imaging using a two-step, 3D-Unet approach. This two-step based approach using 3D U-Nets was chosen given its showed robustness and applicability in a variety of usage scenarios in medical image processing. Its application proved to be well-suited for the segmentation of other anatomical structures, including mandible⁹, the carpal bone¹⁴, and the pelvic bone⁴.

Thus, the contribution of automated segmentation of the fibular bone is a closing link to the process chain of a fully automated pipeline of surgical planning comprising mandibular segmentation⁹, graft design^{15–18} and finally the design of surgical guides¹⁹ to enable a high-quality digital workflow in mandibular reconstruction using fibula-free flaps.

Virtual planning in reconstructive surgery

Preoperative planning is an essential part of complex surgical processes. Recently, there has been increasing research attention on this step^{6,20}, recognizing that well-planned procedures positively impact the surgery as a whole⁷. This focus has guided the growth in the search for patient-specific planning solutions in the field of facial reconstructive surgery in recent years. Consequently, the use of computer-assisted tools in surgical preparation has gained special attention, particularly in mandibular reconstruction surgeries²¹. With the assistance of 3D modeling of the bones, a patient-specific cutting guide can be developed beforehand, along with osteosynthesis plates²². These advancements optimize the surgery in ways that were not possible employing just a freehand surgical approach.

Various researchers compared the mandibular reconstruction surgery when made with and without the assistance of virtual planning guides, where 3D models of the relevant bones are made based on CT images. Results showed significant improvements in operative time and the overall surgery together with better comfort to the surgeon⁷, better reliability²³, and increased symmetry compared to freehand procedures²⁴. Other studies also made systematic reviews of the impact of computer-assisted surgery in maxillary reconstruction based on available literature^{6,20,21}. While there are relevant aspects of criticism on the heterogeneity of the evaluation metrics used to measure the success of the procedures⁶, they also highlight the advantages of computed assisted surgeries on the field^{20,21}.

Medical image segmentation

It is noted that, even though the use of virtual surgery planning has been showing great advantages in mandibular reconstruction surgery in general, the key limitation to this approach is the preparation of the 3D models to be used in the procedure. The first step of this process is to acquire three-dimensional image data of the anatomical regions of interest, typically with CT for bony structures, and then perform segmentation to annotate the bone to be modeled before the surgery. For instance, from the CT scan of the bottom region of the body, extract only the fibula that will be used as a donor bone for the surgery. The simplest and yet most commonly used segmentation method in the surgery planning context is threshold segmentation⁸. It consists of a selection of a threshold of intensity and asserting the value one for every pixel in the image above the intensity threshold, or zero otherwise²⁵. The threshold can be selected either empirically or using a quantitative technique. Van Eijnatten et al.⁸ made an overview of segmentation techniques for additive manufacturing of anatomical models. Their analysis shows that most segmentation applications used either a direct global threshold or some variation of this approach. However, even though this method was proven to be capable of achieving favorable results with decent geometric accuracy, the method includes an extensive manual step. This is because the cited method has a series of inconsistencies, such as not taking CT artifacts and noise into account or intensity variations between different CT scanners¹⁰. So far, the only way around such inconsistencies when creating 3D models suitable for surgery planning procedures or similar tasks includes manual post-processing, which is usually expensive, time-consuming, and requires a skilled and trained expert. Even in the best cases, due to subjectivity or personal variance, the final results of threshold-based segmentation may vary significantly.

To overcome these problems of purely threshold-based segmentations, atlas-based approaches showed remarkable performance, e.g. for femur segmentation²⁶. Deep learning techniques are more recently widely studied in the context of medical image segmentation. Different studies proposed CNN-based techniques to segment bones from CT^{27,28} or MRI images^{29,30}. Yosinski et al.³¹ showed that initializing weights of deep neural networks based on an already trained model improves the generalization performance of the model, even for non-related tasks. This feature is called Transfer Learning and was already used in the medical field, e.g., to train networks to identify COVID infection based on chest X-rays³², to classify brain tumor MRI images³³, and also to classify endoscopic colonoscopy images³⁴. Zhang et al. could show that even segmentation of different facial bones and corresponding landmark detection based on a heatmap approach is feasible³⁵.

Focusing on facial reconstructive surgery planning, Pankert et al.⁹ developed a U-Net-based two-step segmentation approach to segment the mandibular bone. They conducted an extensive discussion on the advantages of using a two-step approach instead of only one U-Net and reached Dice scores of over 0.94 and Average Surface Distances lower than 0.36 mm. The present work intends to investigate the possibility of adapting this two-step segmentation method to the donor bones of facial reconstructive surgery. This way contributes to the current main issue of the planning procedure, which is the high costs and time consumption of the segmentation step.

Materials and methods

Dataset overview

The RWTH Aachen University Hospital provided CT scans of 89 patients for this study along with corresponding manual segmentations of the fibula bone that were used in real surgery planning scenarios. Institutional approval

(EK 260/20) of the local ethics committee of RWTH Aachen University Hospital was obtained. All methods were carried out in accordance with relevant guidelines and regulations.

These segmentations are referred to as ground truth in the following. The age of the patients ranged from 11 to 80 years with a mean of 55.9 and a median of 59 years. 37 patients were male and 52 female. The dataset provided was not fully homogeneous and partially contained inconsistencies, so the data was carefully investigated for those flaws, and sets were excluded from the subsequent study in case they did not meet the inclusion criteria, that was defined as a complete manual segmentation with correct spatial alignment to the corresponding CT imaging data. Additionally, not all patients had both fibula bones segmented, but only one side. In these cases, the available side was chosen for subsequent work. The included datasets were subsequently used in the convolutional neural networks training that is detailed in further sections.

After the exclusion of the inconsistent cases, there were 13 data sets with only the right fibula, 26 data sets with only the left fibula, and 36 data sets with both fibulae segmented, summing up to 111 fibulae from 75 subjects. From these, 10 data sets with both fibulae segmented were held back as an independent test set, to evaluate the proposed segmentation pipelines. This dataset was deemed to be sufficiently large to produce satisfying results, as other studies with two-staged approaches showed to be applicable with smaller datasets, e.g. on carpal bones¹⁴ and in a preliminary study on training set sizes on the mandibular bone⁹. However, considering the wide range of ages and the inclusion of both sexes, the relatively small dataset size is a limitation of the study.

Segmentation approach

In this work, we present two different methods to automatically segment the fibula bone from the CT data. The development of these two different approaches was done because of the dataset specificity that the bones of interest appear twice in the field of view. Both of these propositions are modifications of methods previously studied in literature. Wang et al., for instance, proposed a two-step segmentation framework to segment organs at risk of the head³⁶. The method uses a two-step segmentation process that starts with a localization step that is applied to the whole field of view of the imaging data. The purpose of this initial step is to locate the region of interest around the anatomical structures of relevance in order to crop the field of view around this detected region and to allow for a subsequent detailed segmentation. Subsequently, a second step, that may be called a refinement step, is applied that can focus on the structures of choice and allows to segment these particular structures in detail with high precision by optimally using the available model capability. This method proved to be beneficial in a variety of different use cases, such as the segmentation of mandibular bones⁹, the pelvic bone⁴ or the wrist bone¹⁴, as shown in previous research. In this application, two U-nets are independently trained for each step of the task. For both steps, the same input and output resolution was used.

Therefore, the present study proposes to address the fibula bone segmentation on the available dataset. The baseline method is similar to the approaches described previously: Two U-Nets are trained to perform two different tasks. The purpose of the first U-Net is to segment the bone of interest from the full CT image, while the second is trained to segment the bone only from the bounding box where the bone is present.

Bilateral segmentation

The first proposed method consists of a two-step segmentation, where the first step is responsible for identifying both of the fibulae at the same time. Thus, the first step is trained on the whole CT image, down-sampled to the resolution $144 \times 144 \times 288$ voxels, and the output is the left and the right fibulae, which will be used for the second step of segmentation. To identify the left and the right fibulae, the output of the U-Net is converted into 3D surface objects using the marching cubes algorithm³⁷. We assume that the two largest output meshes are the fibulae bones, while the rest are discarded. Based on their orientation in space, we label the fibulae meshes as left and right.

Similar to³⁶ and⁹, the neural networks for the second step are trained only on the region of interest of the CT, for this application resampled to $80 \times 80 \times 960$, not on the whole data. However, since there are two fibulae, one on the right side and one on the left, two different U-Nets would need to be trained for this step. To circumvent this issue, we propose to mirror all right fibulae along the mid-plane of the field of view to the left side to train the second-step segmentation model on all fibulae simultaneously.

Hence, the entire pipeline consists of the following: Initially, the first step of the segmentation identifies both fibulae based on the full CT. Then the output results are post-processed and transformed into 3D meshes. Then the original CT is cropped to the bounding box of both right and left outputs, and the right fibula and cropped CT are mirrored. Lastly, the second-step model segments each of the fibulae from the cropped CT, then once more the output is post-processed and the right fibula output is re-mirrored. A diagram of this approach is shown in Fig. 1.

Unilateral segmentation

Even though the method presented above is solid, and produces remarkable results, as will be shown in the Results chapter, a relevant issue with this implementation is that for the first step model to be trained, ground truth data for both left and right fibula need to be available, otherwise, it can not be used for training. This is specifically problematic for the available dataset since there are only 36 patients with both fibulae segmented in the whole set, due to the nature of the data coming from real clinical cases for surgical planning. As the second step is trained for each fibula individually, all the data can be used in the training procedure, and no samples have to be discarded. For the first step of the bilateral approach, data from 39 patients were excluded, which is 52% of the complete set.

Therefore, we propose an additional approach for the first step, where potentially all the data can be used for training the U-Net models. The proposed method consists of using all the fibulae individually for training, similar to what is already done in the second step. Instead of segmenting both fibulae from the complete CT scan,

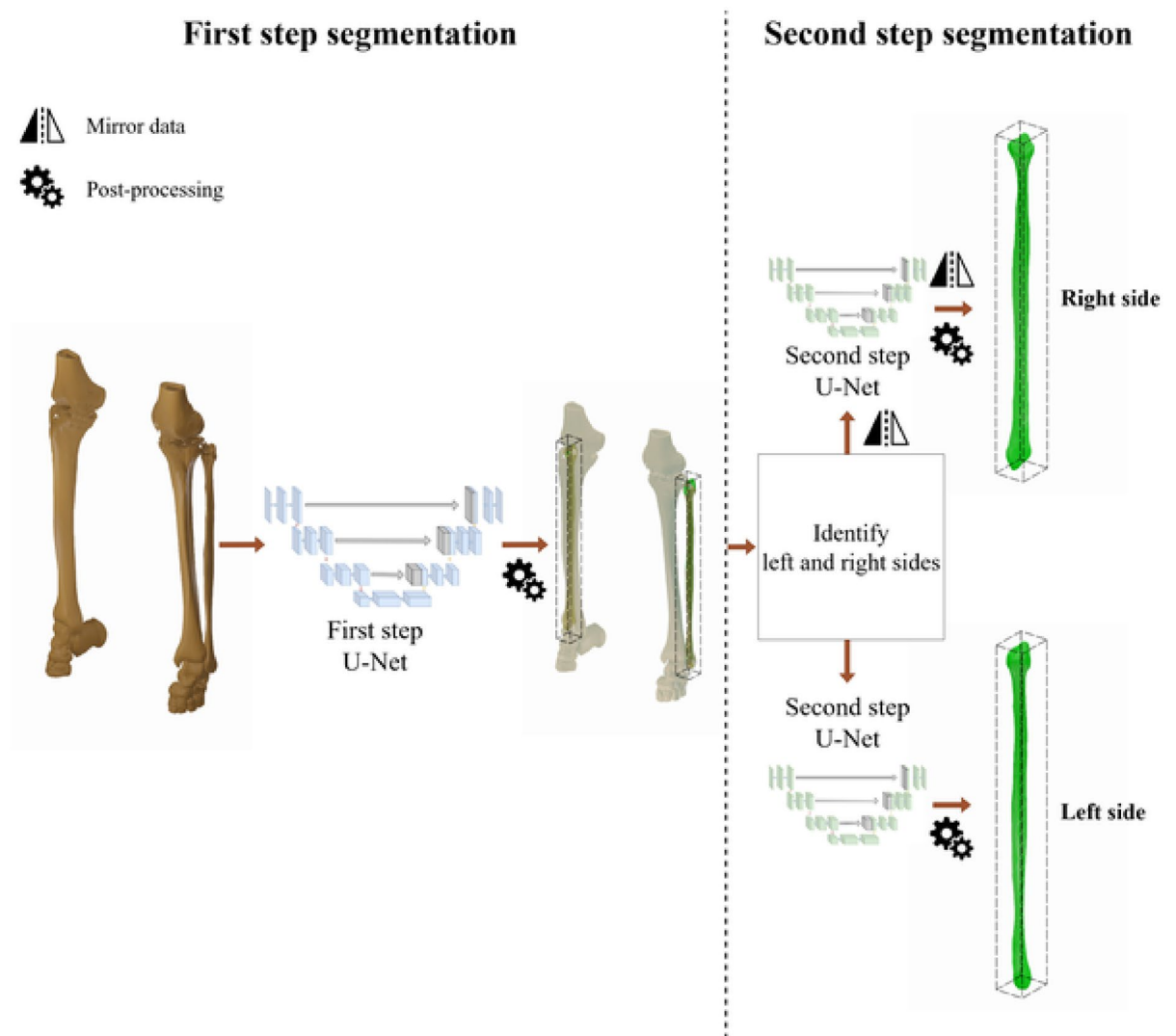


Fig. 1. Bilateral segmentation approach overview: The first step takes the original CT data (shown in the diagram as threshold segmentations for visual clarity) and segments both fibula bones from the whole CT, then the second step segments each fibula bone from the CT cropped to the region of interest.

the proposition is to split the CT in half along the mid-plane of the field of view, with each half containing only one fibula. This does not imply an exact alignment with the anatomical mid-sagittal plane, but in our examples, as all patients were lying centric enough on the bench with sufficient distance between their legs, thus this approach showed to be valid on the given data. Subsequently, the first step segmentation model is trained to segment the fibula based only on this half section of the image. To avoid training two different neural networks for the same task, the same mirroring technique used in the second step is also applied to this modified first step. Hence, the right side of the CT, which was cropped, is also mirrored along with the ground truth in this step and only one neural network is trained in this step to identify left fibulae from the left part of CTs. A different resolution is also necessary for this first step, since the region that will be input for the U-Net is no longer based on the full CT scan. To maintain the relative proportion, the proportion of voxels from the y-axis to the x-axis is doubled, and the proportion from the y-axis to the z-axis is kept the same. The chosen resolution therefore is $96 \times 192 \times 320$.

In conclusion, the method presented here involves first splitting the CT data in half on the lateral axis, then using the same first-step model to segment both the left and the mirrored right side of the fibulae. Finally, to apply the second step model to segment both fibulae from the CT images cropped only to the bounding box of the fibulae found in the first step. Figure 2 presents an overview of this approach.

Dataset distribution

After preprocessing of the raw imaging data and a technical check for data integrity, all available data was used for the training of the respective networks and divided into a training and validation set. For the training of the bilateral first step of this model, 22 could be made available, split with a 9 to 1 ratio into training set ($n=20$) and a validation set ($n=2$). The first step of the unilateral segmentation could use a database of 87 fibulae, where 80

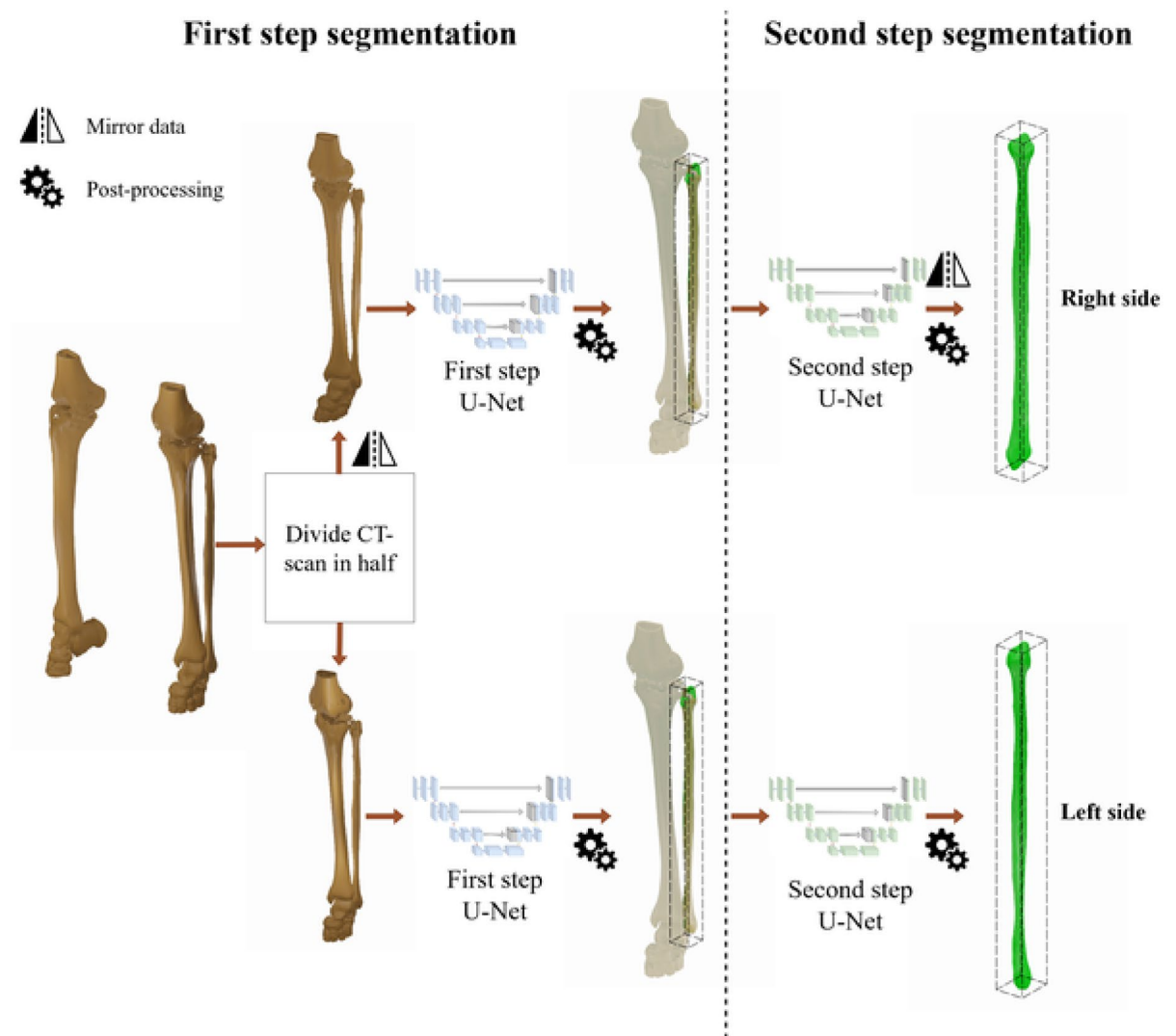


Fig. 2. Unilateral segmentation approach overview: The CT dataset is divided into left and right sections (shown in the diagram as threshold segmentations for visual clarity), and then the first step segments each fibula bone from its respective side, and finally the second step segments each fibula bone from the CT cropped to the region of interest.

could be used in training and 7 for validation. The second stage could rely on dataset of 85 fibulae, split into a dataset training (n=75) and validation (n=7).

Transfer learning

As transfer learning has shown to be beneficial in related research^{31–34}, both approaches proposed in this study use it. In this application, first, the second-step U-Net is trained from scratch, as this is expected to be a computationally easier task. Subsequently, transfer learning to the first step is used, as this showed superior results in preliminary work on other segmentation tasks. This allows for expanding the more detailed training results of the second step to the whole field of view of the CT, hence this approach is referred to as Expansion Transfer Learning (ETL)¹⁴.

U-Net architecture

The U-Net architecture used for the experiments is a variation of the three-dimensional U-Net described in³⁸. It has ten convolutional layers in the encoding path, with four max pooling layers in between. Following a common pattern seen in the literature^{9,39}, it doubles the number of feature maps for every pooling step. The network uses $3 \times 3 \times 3$ receptive fields and $2 \times 2 \times 2$ max pooling layers. Furthermore, the decoding path consists of nine convolutional layers and five upsampling layers, which simply repeat the elements of the input tensor by an upsampling factor in every dimension. The upsampling factor used in the described architecture was $2 \times 2 \times 2$, which means the output of every upsampling layer is double its input for every dimension. An overview of the 3D U-net described above can be seen in Table 1. All networks described in this paper were

trained using an Adam optimizer⁴⁰ and a Dice loss. We did not implement data augmentation techniques in our presented methodology. While preliminary experiments were conducted with various augmentation approaches including spatial deformation, scaling, intensity variations, and noise addition, we did not observe significant improvements in model performance that would have justified their inclusion in our final implementation.

No individual image normalization was used to standardize intensity values. However, the standard range of Hounsfield units (from -1024 to +3071 HU) was clipped and mapped to floating point numbers in the range from 0.0 to 1.0 for further processing in the neural network.

Post-processing

Both proposed segmentation methods include a post-processing step of the model’s predictions. This stage occurs after all the neural networks outputs and transform the binary image output to the 3D model mesh, that is subsequently used in the surgical procedure. The first step of post-processing is a binary erosion of the data⁴¹ and is followed by the application of the Gaussian filter⁴² to add smoothness. Finally, the binary data is transformed to a mesh using the marching cubes algorithm on the predicted data³⁷.

Evaluation metrics

To evaluate the trained models in this study, we use three metrics commonly used in medical image analysis in similar applications. The first metric is the Dice Score (DS)⁴³, an overlapping metric that compares the result segmentation and the ground truth. For a ground truth segmentation represented by the set of voxels *A* and a predicted segmentation represented by the set of voxels *B*, the Dice Score can be formalized according to Equation 1.

$$\mathcal{D}\mathcal{S} = 2 \cdot \frac{A \cap B}{A + B} \tag{1}$$

Additionally, we use two distance metrics, the Average Surface Distance (ASD)⁴⁴, and the 95% Hausdorff Distance (95% HD)⁴⁵ which measure the average, and the 95% highest distance between the ground truth and the model prediction respectively.

Restricted evaluation of transplantable bone

During the facial reconstructive surgery with the use of the fibula-free flap, 6 cm or 7 cm from the top and bottom of the fibula are preserved to avoid knee or ankle instability^{46–48}. Therefore, only the central part of the fibula is effectively used in the procedure and needs to be as well segmented as possible for the present application. Consequently, the proposition here is to evaluate with the same metrics as defined beforehand, removing the top and bottom 7 cm of the fibula bone. This removal can be seen in Fig. 3. This removal is done on both ground truth segmentation and the output of the proposed framework. Then the cropped meshes are again compared using Dice Score, Average Surface Distance, and 95% Hausdorff Distance. In this study, these metrics are additionally labeled as Region of Interest (ROI) metrics to differentiate from the standard implementations on the whole fibular bone.

Results

The results section is split into a qualitative part and a quantitative part. Both refer to the final prediction, i.e. the output of the respective second U-Nets.

Qualitative results

Bilateral segmentation approach

	Type	Output feature maps
Encoding path	Input layer	1
	Conv	16
	Conv + Pool + Conv	32
	Conv + Pool + Conv	64
	Conv + Pool + Conv	128
	Conv + Pool + Conv	256
	Conv	512
Decoding path	Conv	512
	Conv + Upsampling + Conv	256
	Conv + Upsampling + Conv	128
	Conv + Upsampling + Conv	64
	Conv + Upsampling + Conv	32
	Conv	1
	Output Layer	1

Table 1. 3D U-Net architecture.



Fig. 3. Fibula bone cropping example: the image shows the top and bottom 7cm being removed from the evaluation procedure.

Figure 4 shows predictions from the trained models, each aside from its respective ground truth. The two neural networks combined succeed in identifying not just the fibula bone region and a fibula-like object, but the fibulae are visually close to the corresponding ground truth segmentation.

Visually, all the bottom edges and central regions look well segmented, whereas the top parts are not perfect for all cases. The first and second fibulae, especially, have flaws in the segmentation of its top regions, where the ground truth has it well segmented.

Unilateral segmentation approach

The predictions obtained using the unilateral segmentation approach can be seen in Fig. 5. These results were generated for the same fibulae as the ones in Fig. 4, therefore a direct comparison is possible.

Visually, the predictions are closely aligned with the ground truths, especially on the central region, which is the one of biggest importance for facial reconstructive surgery, and also for the bottom region. The first and second fibulae show minor flaws in the top region segmentation, even though the segmentation looks relatively better than the ones achieved using the bilateral segmentation approach. Figure 6 displays a closer view of one example of prediction and highlights the poor results on the top region and the low deviation results on the central and bottom regions.

Quantitative results

Table 2 displays the quantitative evaluation of the trained models on the test set.

The unilateral approach performs overall better in all metrics. When evaluating the metrics on the whole fibula, the achieved Dice Scores are 0.85 and 0.87 for the bilateral segmentation and the unilateral method respectively, while the Average Surface Distance and 95% Hausdorff Distance show an even bigger difference between the two approaches. It is also important to note the high standard deviation of the bilateral segmentation evaluation.

However, for the Region of Interest (ROI) metrics, the difference between the two approaches is less significant. Both methods achieve high DS values of 0.94 (bilateral approach) and 0.95 (unilateral approach), the ASD and 95% HD metrics are at 0.36 mm to 0.31 mm and 0.89 mm to 0.79 mm, respectively (Table 2).

The bilateral approach suffers from a robustness problem visible in the high standard deviations. For one of the test cases, the first step of the bilateral approach failed to predict one of the fibulae as a continuous bone. Due to filtering in the post-processing, the inferior part of the fibula was discarded, resulting in a too small field of view for the second segmentation step (Fig. 7).

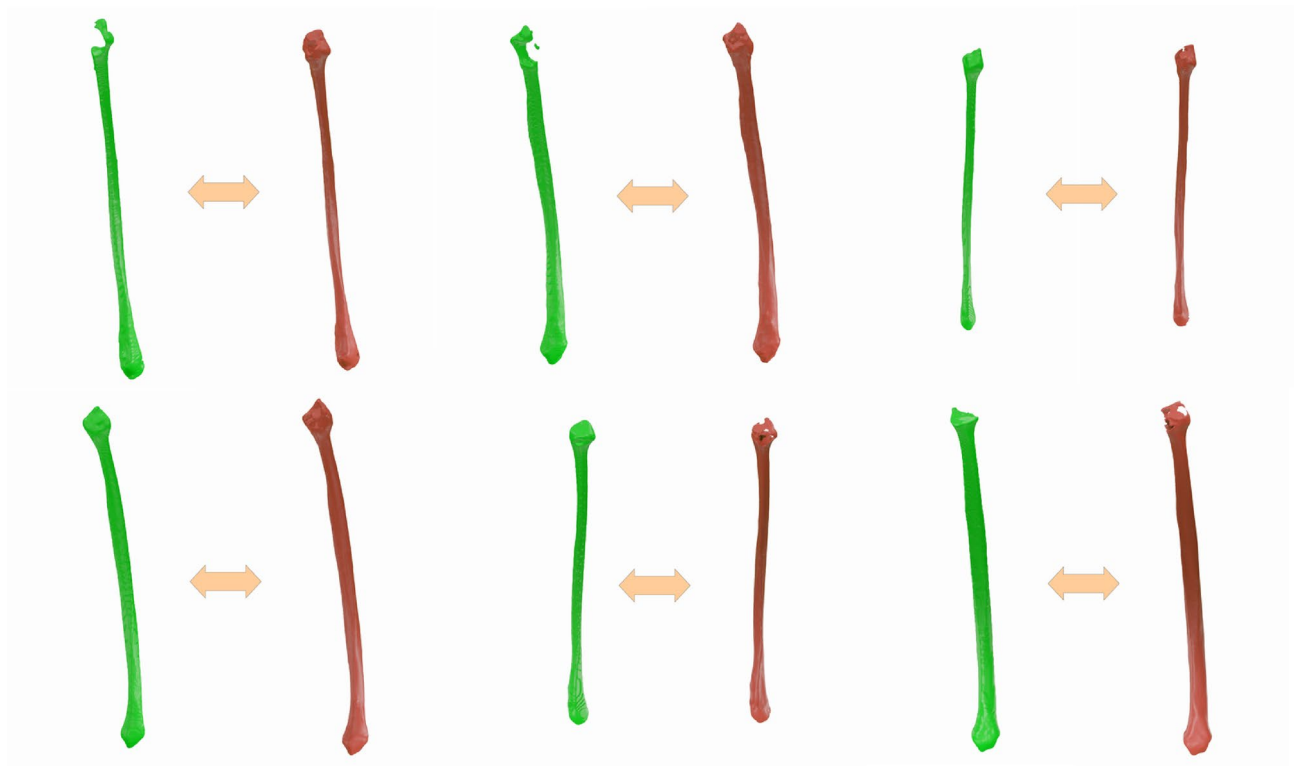


Fig. 4. Bilateral Segmentation approach: Predictions (green) and ground truths (red) examples.

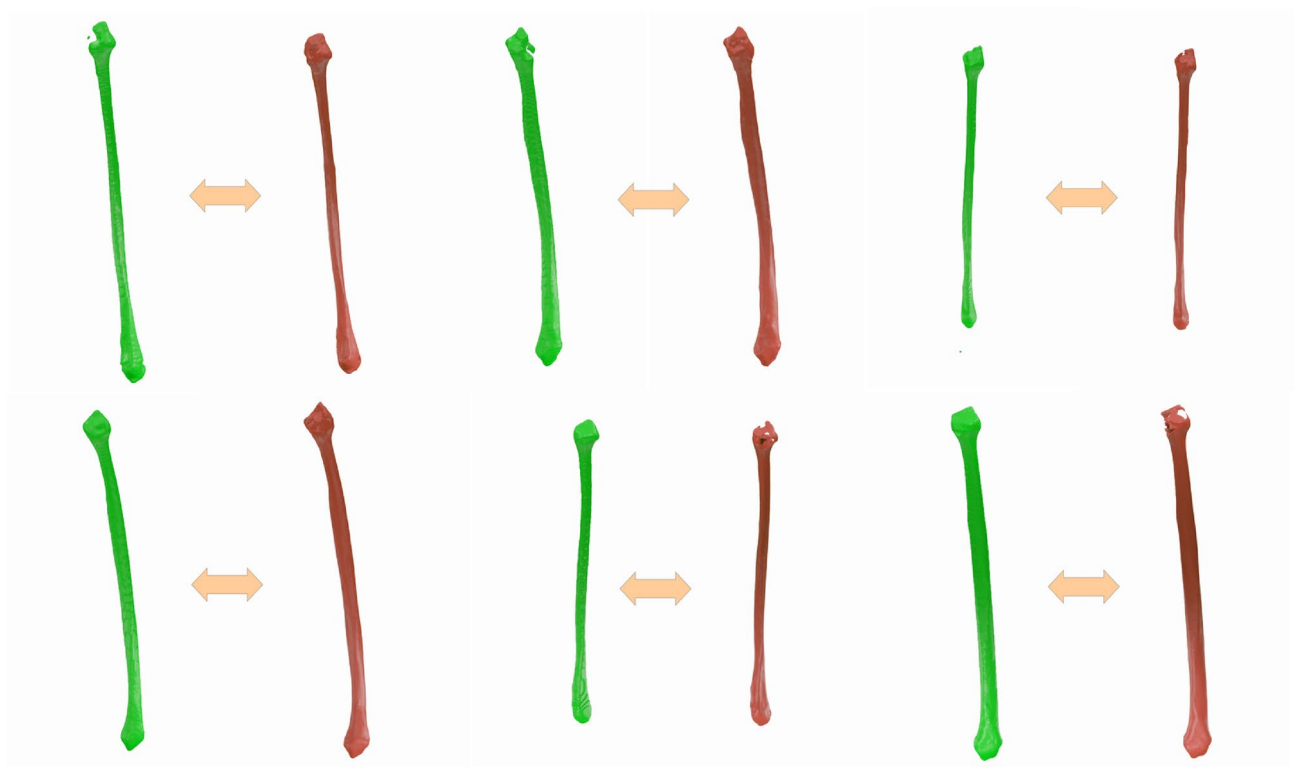


Fig. 5. Unilateral segmentation approach: predictions (green) and ground truths (red) examples.

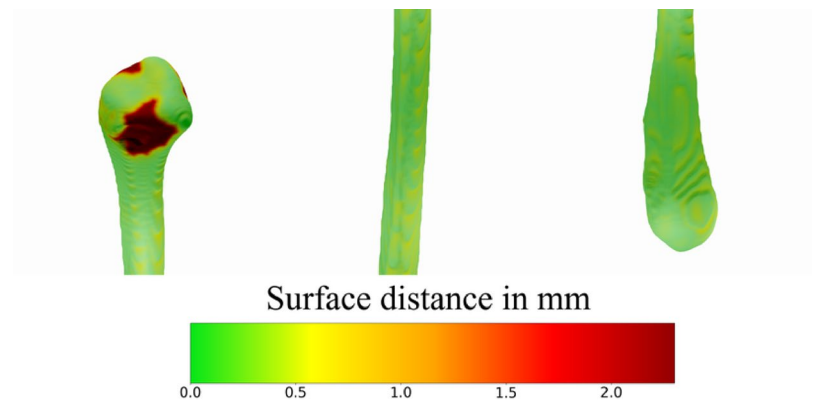


Fig. 6. Detail view of a prediction in relation to ground truth data in color-coded 3D surface comparison. (left: proximal region of the fibula (head), middle: shaft of the fibula, right: lateral region of the fibula (ankle). The image shows an example result of the bilateral approach.

Approach	DS	ASD	95% HD	ROI DS	ROI ASD	ROI 95% HD
Bilateral segmentation	0.85 ± 0.09	0.91 ± 1.55	5.45 ± 14.81	0.94 ± 0.02	0.36 ± 0.12	0.89 ± 0.59
Unilateral segmentation	0.87 ± 0.05	0.58 ± 0.25	2.15 ± 0.84	0.95 ± 0.02	0.31 ± 0.13	0.79 ± 0.63

Table 2. Evaluation of all models on the whole test set based on Dice Score (DS), Average Surface Distance (ASD) 95% Hausdorff Distance (HD) both on the whole fibular bone, and restricted to the Region of Interest (ROI).

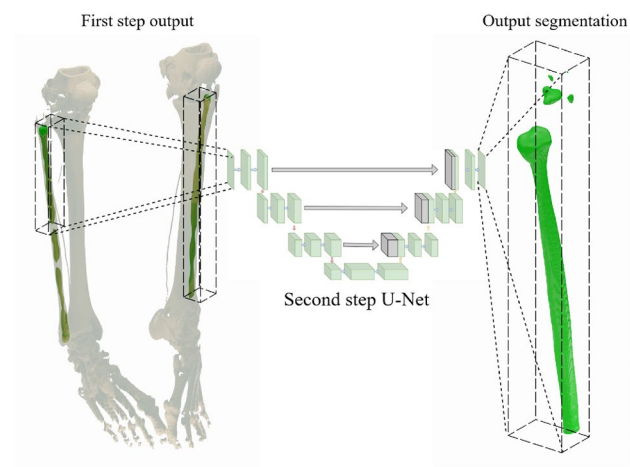


Fig. 7. The low accuracy of the first-step segmentation in the bilateral approach leading to an incorrect bounding box for the second segmentation step. The outputs are displayed after conversion to mesh representations and before filtering out unconnected parts. Note: this is an exceptionally bad case and not the standard of our study.

Figure 8 shows the statistical distribution of the ASD evaluation on the test set for both the bi- and unilateral approaches. It is important to note that evaluating the same prediction only in the region of interest provides better results.

The unilateral approach solves the problem of high differences in the outliers. This variant achieves an average ASD and 95% HD value of 0.58 mm and 2.15 mm, respectively. Both standard deviations are drastically lower as well, which are 0.25 mm and 0.84 mm for the whole bone. Running the segmentations, including preprocessing of the DICOM datasets and all post-processing steps took on average 50 s for the bilateral and 58 s for the unilateral approach per dataset on a PC with an *Nvidia GeForce RTX 2080 Ti* and an *Intel Core i9-7900X*.

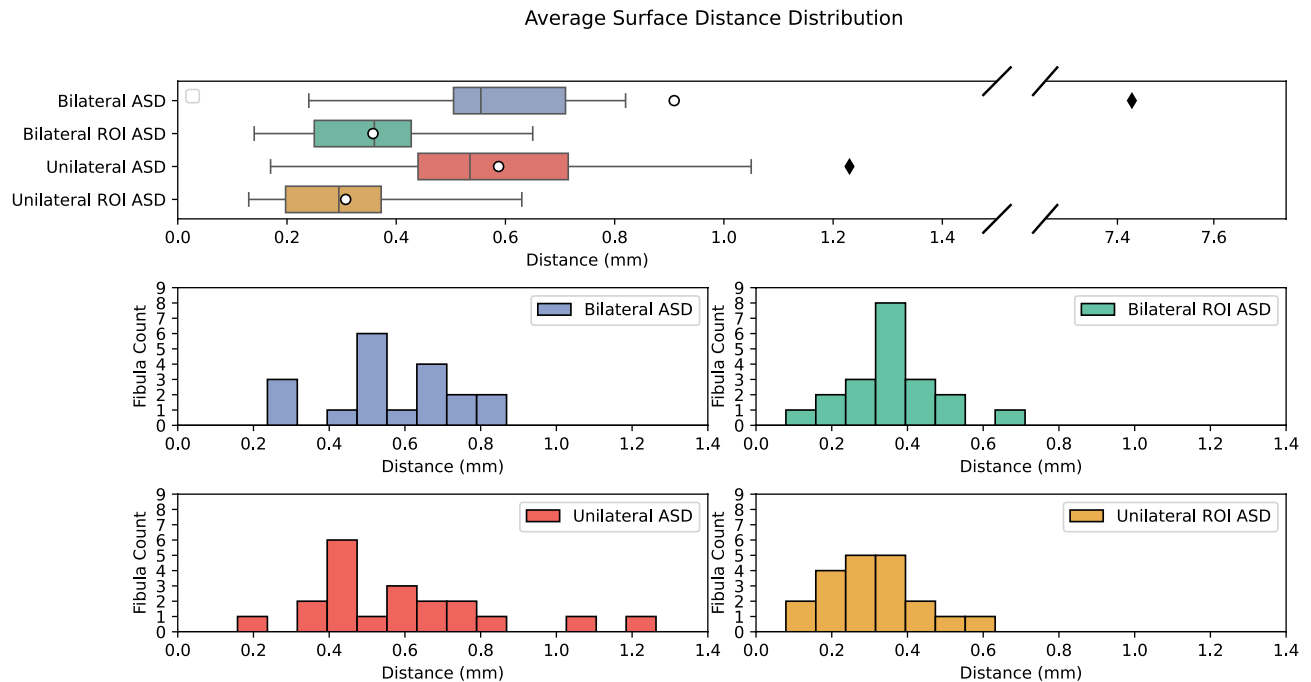


Fig. 8. Distribution of the average surface distances (ASD) on the full bone or within the region of interest (ROI) for bi- and unilateral segmentations. For the histogram of the bilateral segmentations (in red), the outlier at 7.43 mm has been omitted.

Discussion

Even though the fibular bone is an important anatomy for diverse applications in the medical field, it is still not covered by research works on automatic segmentation of this bone, especially using AI techniques.

Though targeting a long bone of the lower extremity, the present work is particularly relevant for the field of maxillofacial reconstruction, where the fibula is commonly used as a donor site for autologous bone transplantation, and fast and reliable segmentation can significantly improve the preoperative workflow. Hence, the experiments conducted in this work are of high value to this clinical domain, as the proven applicability of this method in the given context demonstrates the efficiency for segmenting the desired parts of the fibula from CT images.

Therefore, the experiments conducted in this work are of high value to the field, since their remarkable results attest to the efficiency of the proposed methods in segmenting the desired bone from CT images. These results are especially good not just for the use case but have the potential to be used in different applications, such as fracture detection^{49,50} or disease prediction⁵¹. Both methods proposed are modifications of the approach described in^{9,36}. Overall, the results showed a high quality of the segmentation for both proposed approaches. When evaluating the predictions on the whole bone, the average of the Dice Score metric found for the two cases is over 0.84. Overall, the bilateral approach showed decreased robustness, as visible in the high standard deviations, due to the presence of outliers. The unilateral segmentation approach, by contrast, solves this problem and shows a significantly improved robustness.

This result is a strong indication of how the unilateral approach deals better with difficult cases than the bilateral segmentation approach. Even in CT scans for which the bilateral approach has difficulties in performing the first step segmentation, this method provides a segmentation that is sufficient to correctly detect the region of interest for the second step input. That the proximal part of the fibula shows weaker segmentation accuracy can be treated as irrelevant for the intended purpose of autologous transplantation. However, for potential application of this segmentation to other medical indications, the head of the fibula could deserve a more careful investigation, either by improving the ground truth segmentations in this particular region or by specific post-processing with shape-based refinement.

Therefore, the visual analysis corroborates the statistics, which attest to an overall improvement of the metrics when using the unilateral approach when evaluating the whole bone. This is expected for two main reasons: The first one is directly related to the relative resolution of the down-sampling for the first step. Since the CT scan is divided in half for the unilateral approach, and the resolution used is kept proportionally, the relative down-sampling is lower for this case. Thus, for the bilateral method, more information is lost, which affects directly the training procedure. The second reason is related to the dataset availability. As described beforehand, not every patient in the data used for training has both fibulae segmented as a ground truth. This is a problem for the bilateral approach since it needs both bones to have ground truths to allow better training. Thus, the number of samples used for training in this method is 67.5% lower in comparison to the unilateral method. These factors

combined explain how the split approach can deal better with the outliers, and have overall better results, even though the second step network is the same.

All these results together show that the unilateral approach is overall a better method to perform the fibula bone segmentation task than the bilateral approach, especially because of its higher robustness, and its remarkably better results for the metrics that evaluate the whole bone. Nevertheless, both the quantitative and qualitative analyses suggest that the main difference overall between the two methods is outside the region of interest. When evaluating both approaches only on the region of interest, the difference between them is shown to be significantly lower. Even the worst results of both metrics show Average Surface Distances below 1 mm in the fibula region used for the surgery. The other metrics show similar results, since the average of the Dice Score and 95% Hausdorff distance metrics are respectively 0.94 and below 1 mm for both approaches, and even the worst outliers are within a good range considering the task. These errors are all within the order of magnitude of the resolution of the underlying CT scans.

Thus, in summary, both two-step approaches showed to be applicable and yield remarkable accuracy, leveraging the full potential of the available hardware by applying a coarse segmentation in a first stage, followed by a detailed segmentation around the detected area of interest. In this approach, the advantages of using one common model architecture and training dataset could be combined with the gain in relative resolution in the focused second step.

For the application of the presented segmentation method in the context of surgical planning, the required accuracy depends on many different factors. Generally, the more accurate the segmentation the better for the surgical outcome. Accuracy is limited by several factors, including imaging, segmentation and manufacturing of parts created based on this data. Van den Broeck et al. report a mean error of 0.48 mm between CT scans and optical scans of the tibial bone diaphysis⁵², showing that the imaging already contains some errors in comparison to the real clinical application. Wallner et al. report an inter-observer Dice score of 0.94 between the manual segmentation of the mandible bone from CT imaging from two different human annotators⁵³. Dice scores much higher than this might indicate an overfitting of the AI models to a specific human annotator instead of correctly generalizing to the segmentation problem. In a work by Kim et al., the manufacturing accuracy of surgical guides based on different manufacturing techniques was measured, showing mean errors ranging from 0.09 mm for SLA printing to 0.31 mm for Multijet printing (MJP)⁵⁴.

These results are of high importance for the present work, since they show that even though the bilateral segmentation approach has problems regarding robustness when considering only the region that is considered in the facial reconstructive surgery, the results are outstanding for all cases. Even though there are no studies that apply a similar approach to perform the segmentation of fibular bones, it is possible to compare the achieved results to other bone segmentation from volumetric images. For instance, when performing the vertebral cortical segmentation from chest CT scans using a 3D U-Net-based approach,⁵⁵ obtains Dice Scores of only 0.71, which is considerably lower compared to the results achieved in the present work.⁵⁶ proposed a cascaded approach to segment condyles from CT scans, also using 3D U-Nets. The results achieved in this study were of high quality, achieving over 0.93 Dice Score values and Hausdorff Distances lower than 2.45 mm. Focusing on the reconstructive facial surgery planning application, Pankert et al. developed a fully automatic framework to segment the mandibular bone and obtained averages of 0.94, 0.35 mm, and 0.97 mm for the Dice Score, Average Surface Distance, and 95% Hausdorff Distance respectively, and attested its significance and possible applicability in surgical planning scenarios⁹. All the discussed comparisons have to be done carefully since even though the metrics are the same, the segmented bones and the available datasets are significantly different. However, literature results attest to what is a good segmentation in the medical image field. In fact, for the region of interest, the results obtained for this work are either better or at least similar for both presented methods.

The complete process of segmentation (including pre- and post-processing) took less than one minute on average on a workstation PC with contemporary hardware. Even though the fibular bone might be an easier task for manual segmentation than other anatomical structures, the time for manual segmentation is expected to be much longer. Though there is no published data available on the average times for manual segmentation of the fibula from CT scans, and comparisons with the segmentation of other anatomical structures are not directly applicable. Based on internal consultation with personnel experienced in manual segmentation, an estimated duration of 5 to 10 minutes per fibula was reported. It is important to note that this estimate refers to trained professionals with substantial experience; for less experienced users or those performing the task only occasionally, a significant learning curve can result in considerably longer times. While this working time is lower than for more complex anatomical structures or regions with extensive imaging artifacts, it still represents a relevant burden in an already packed clinical workflow. Moreover, with an automated segmentation using a CNN, any inter-observer differences may be omitted.

Thus, the AI approach presented in this work, enables the integration into a comprehensive process chain of surgical planning, which involves more cumbersome tasks, and profits significantly from an objective and reproducible acquisition of bony geometry from CT data. The segmented virtual representation of fibular bones can be directly integrated into computer-assisted planning workflows, such as virtual preoperative planning of surgical reconstruction of the mandible or midface bones in continuity defects, as is required e.g. after tumor removal, bone necrosis, or excessive trauma, where an autologous transplant from the fibular bone in a so-called fibula free flap proved its clinical value^{6,7,20,21}. The geometric accuracy of the virtual surfaces of the digital bone models is essential for the planning of the transplant to match the defect dimensions and thus allow for optimal functional and aesthetic rehabilitation.

This work focuses on the surgical planning of fibula bone transplants in reconstructive surgery. A critical factor for clinical success is the accurate identification and integration of supporting blood vessels, as visible in CT angiography. While the present study concentrates on bone geometry, a natural extension of the automated segmentation approach would thus be the integration of vascular structures. Enabling automatic detection of

relevant vessels would allow this information to directly inform clinical decision-making without requiring additional manual input. However, this task is considerably more challenging due to the high anatomical variability of vascular structures and would likely necessitate a different study design, potentially with access to larger and more diverse datasets.

It is important to note some limitations of the study. The used dataset was prepared to focus on a facial reconstructive surgery context, therefore flaws in unimportant regions for the surgery were ignored when modeling the segmentations. This directly affected the predictions achieved by the trained models. Furthermore, the low number of data available for training, especially on the bilateral segmentation approach, was probably one of the causes of the model's lack of robustness for this method. Thus, we expect the overall results to improve significantly when more data is available for training. More data could also be generated by applying data augmentation to the existing datasets. Even though within the limitation of the preliminary studies conducted prior to the present work did not yield a discernible improvement in performance, a further ablation on augmentations techniques and settings could achieve an improvement of the training performance. In future work, we intend to apply similar methods to segment other donor bones used in the facial reconstructive surgery procedure. Furthermore, comparisons of our CNN-based approach to Atlas-based methods²⁶ might yield relevant insights to the technical specifics of the different approaches in this particular field.

Conclusion

This study presented two methods to automatically segment fibula bones from full-resolution CT scans. Both propositions are based on two-step segmentations, where the second step predicts the output based on the region of interest found by the first step. The bilateral approach's first step does a coarse segmentation of both fibulae from a full-resolution CT at once, while the unilateral approach first splits the CT into two lateral halves to simplify the task by increasing the relative image resolution and the usable training data.

Both alternatives presented in this paper have the potential to be used in clinical surgical planning scenarios after integration into approved software as a medical device. The framework is capable of producing accurate segmentation of the fibula bone from the CT scans with a significant improvement in speed in comparison to established (semi-) manual segmentation techniques, and can be directly used as input data for the computer-assisted planning procedure.

Even eventual imprecision can be easily manually corrected before the procedure, which would still be a better and overall cheaper approach than the state-of-the-art threshold segmentation-based techniques.

Data availability

The raw medical imaging datasets analyzed during this study are not publicly available due to privacy concerns, as per the guidelines of the ethics committee, which permits access to the original data only for a specific group of named researchers. However, evaluation data and statistical analyses derived from it are available from the corresponding author upon reasonable request.

Code availability

The code for evaluation of the geometry data is available as part of the supplementary material of this publication. The base implementation of the utilized 3D U-Net is available under <https://github.com/smatzek/3DU-netCNN>. Further details of the implementation and methodological description, as well as specific code parts used in this study for the training and inference pipelines, are available from the corresponding author upon reasonable request.

Received: 24 July 2025; Accepted: 14 November 2025

Published online: 25 November 2025

References

1. Mahmud, M. I., Mamun, M. & Abdelgawad, A. A deep analysis of brain tumor detection from MR images using deep learning networks. *Algorithms* **16**, 176. <https://doi.org/10.3390/a16040176> (2023).
2. Poongodi, M. et al. Diagnosis and combating COVID-19 using wearable oura smart ring with deep learning methods. *Pers. Ubiquit. Comput.* **26**, 25–35. <https://doi.org/10.1007/s00779-021-01541-4> (2021).
3. Wen, Z. et al. Deep learning in digital pathology for personalized treatment plans of cancer patients. *Semin. Diagn. Pathol.* **40**, 109–119. <https://doi.org/10.1053/j.semdp.2023.02.003> (2023).
4. Raith, S. et al. Segmentation of the iliac crest from CT-data for virtual surgical planning of facial reconstruction surgery using deep learning. *Sci. Rep.* **15**, 1097 (2025).
5. van Baar, G. J. et al. Accuracy of computer-assisted surgery in maxillary reconstruction: A systematic review. *J. Clin. Med.* **10**, 1226 (2021).
6. Chan, T. J., Long, C., Wang, E. & Prisman, E. The state of virtual surgical planning in maxillary reconstruction: A systematic review. *Oral Oncol.* **133**, 106058 (2022).
7. Culié, D. et al. Virtual planning and guided surgery in fibular free-flap mandibular reconstruction: A 29-case series. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **133**, 175–178. <https://doi.org/10.1016/j.anorl.2016.01.009> (2016).
8. van Eijnatten, M. et al. Ct image segmentation methods for bone used in medical additive manufacturing. *Med. Eng. Phys.* **10**, 24–25. <https://doi.org/10.1016/j.medengphys.2017.10.008> (2017).
9. Pankert, T. et al. Mandible segmentation from CT data for virtual surgical planning using an augmented two-stepped convolutional neural network. *Int. J. Comput. Assist. Radiol. Surg.* <https://doi.org/10.1007/s11548-022-02830-w> (2023).
10. Minnema, J. et al. Ct image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput. Biol. Med.* **103**, 130–139. <https://doi.org/10.1016/j.compbiomed.2018.10.012> (2018).
11. Raudaschl, P. F. et al. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Med. Phys.* **44**, 2020–2036 (2017).
12. Hidalgo, D. A. & Pusic, A. L. Free-flap mandibular reconstruction: A 10-year follow-up study. *Plast. Reconstr. Surg.* **110**, 438–449. <https://doi.org/10.1097/00006534-200208000-00010> (2002).

13. Hölzle, F., Franz, E. P., von Diepenbroick, V. H. & Wolff, K.-D. Evaluation der unterschenkelarterien vor mikrochirurgischem fibulatransfer. *Mund Kiefer Gesichtschir.* **7**, 246–253. <https://doi.org/10.1007/s10006-003-0486-8> (2003).
14. Raith, S. et al. Multi-label segmentation of carpal bones in mri using expansion transfer learning. *Phys. Med. Biol.* (2025).
15. Raith, S. et al. Introduction of an algorithm for planning of autologous fibular transfer in mandibular reconstruction based on individual bone curvatures. *Int. J. Med. Robot. Comput. Assist. Surg.* **14**, e1894. <https://doi.org/10.1002/racs.1894> (2018).
16. Modabber, A. et al. Evaluation of a novel algorithm for automating virtual surgical planning in mandibular reconstruction using fibula flaps. *J. Craniomaxillofac. Surg.* **47**, 1378–1386. <https://doi.org/10.1016/j.jcms.2019.06.013> (2019).
17. Hagen, N. et al. A user-friendly software for automated knowledge-based virtual surgical planning in mandibular reconstruction. *Journal of Clinical Medicine* **14**, <https://doi.org/10.3390/jcm14134508> (2025).
18. Guo, Y. et al. Automated planning of mandible reconstruction with fibula free flap based on shape completion and morphometric descriptors. *Med. Image Anal.* **102**, 103544. <https://doi.org/10.1016/j.media.2024.103544> (2025).
19. Šimić, L., Kopačin, V., Mumlek, I., Butković, J. & Zubčić, V. Improved technique of personalised surgical guides generation for mandibular free flap reconstruction using an open-source tool. *Eur. Radiol. Exp.* **5**, 30. <https://doi.org/10.1186/s41747-021-00229-x> (2021).
20. van Baar, G. J. C. et al. Accuracy of computer-assisted surgery in maxillary reconstruction: A systematic review. *J. Clin. Med.* **10**, <https://doi.org/10.3390/jcm10061226> (2021).
21. Powcharoen, W., Yang, W.-F., Li, K. Y., Zhu, W. & Su, Y.-X. Computer-assisted versus conventional freehand mandibular reconstruction with fibula free flap: a systematic review and meta-analysis. *Plast. Reconstr. Surg.* **144**, 1417–1428 (2019).
22. Wilde, F. et al. Multicenter study on the use of patient-specific cad/cam reconstruction plates for mandibular reconstruction. *Int. J. Comput. Assist. Radiol. Surg.* **10**, 2035–2051. <https://doi.org/10.1007/s11548-015-1235-2> (2015).
23. Al-Sabahi, M. E. et al. Aesthetic reconstruction of onco-surgical mandibular defects using free fibular flap with and without CAD/CAM customized osteotomy guide: A randomized controlled clinical trial. *BMC Cancer* **22**, <https://doi.org/10.1186/s12885-022-10322-y> (2022).
24. Bartier, S., Mazzaschi, O., Benichou, L. & Sauvaget, E. Computer-assisted versus traditional technique in fibular free-flap mandibular reconstruction: A ct symmetry study. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **138**, 23–27. <https://doi.org/10.1016/j.anorl.2020.06.011> (2021).
25. Al-Amri, S. S., Kalyankar, N. V. et al. Image segmentation by using threshold techniques. *arXiv preprint arXiv:1005.4020* (2010).
26. Besler, B. A., Michalski, A. S., Forkert, N. D. & Boyd, S. K. Automatic full femur segmentation from computed tomography datasets using an atlas-based approach. In *Computational Methods and Clinical Applications in Musculoskeletal Imaging*, vol. 10734 of *Lecture Notes in Computer Science*, 120–132, https://doi.org/10.1007/978-3-319-74113-0_11 (Springer, 2018).
27. Liu, P. et al. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 749–756 (2021).
28. Hemke, R., Buckless, C. G., Tsao, A., Wang, B. & Torriani, M. Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. *Skeletal Radiol.* **49**, 387–395 (2019).
29. Liu, X. et al. Fully automated pelvic bone segmentation in multiparametric mri using a 3d convolutional neural network. *Insights into Imaging* **12** (2021).
30. Milletari, F., Navab, N. & Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571 (IEEE, 2016).
31. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inform. Process. Syst.* **27** (2014).
32. Das, N. N., Kumar, N., Kaur, M., Kumar, V. & Singh, D. Automated deep transfer learning-based approach for detection of COVID-19 infection in chest x-rays. *IRBM* **43**, 114–119. <https://doi.org/10.1016/j.irbm.2020.07.001> (2022).
33. Deepak, S. & Ameer, P. Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* **111**, 103345. <https://doi.org/10.1016/j.combiomed.2019.103345> (2019).
34. Wang, Y., Feng, Z., Song, L., Liu, X. & Liu, S. Multiclassification of endoscopic colonoscopy images based on deep transfer learning. *Comput. Math. Methods Med.* **1–12**, 2021. <https://doi.org/10.1155/2021/2485934> (2021).
35. Zhang, R. et al. Craniomaxillofacial bone segmentation and landmark detection using semantic segmentation networks and an unbiased heatmap. *IEEE J. Biomed. Health Inform.* **28**, 427–437. <https://doi.org/10.1109/JBHI.2023.3337546> (2024).
36. Wang, Y., Zhao, L., Wang, M. & Song, Z. Organ at risk segmentation in head and neck ct images using a two-stage segmentation framework based on 3d u-net. *IEEE Access* **7**, 144591–144602. <https://doi.org/10.1109/ACCESS.2019.2944958> (2019).
37. Lewiner, T., Lopes, H., Vieira, A. W. & Tavares, G. Efficient implementation of marching cubes' cases with topological guarantees. *J. Graph. Tools* **8**, 1–15. <https://doi.org/10.1080/10867651.2003.10487582> (2003).
38. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II* 19, 424–432 (Springer, 2016).
39. Wang, C., MacGillivray, T., Macnaught, G., Yang, G. & Newby, D. A two-stage 3d unet framework for multi-class segmentation on full resolution image. *arXiv preprint arXiv:1804.04341* (2018).
40. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. [arxiv:https://arxiv.org/abs/1412.6980](https://arxiv.org/abs/1412.6980) (2014).
41. Ji, L., Piper, J. & Tang, J.-Y. Erosion and dilation of binary images by arbitrary structuring elements using interval coding. *Pattern Recognition Letters* **9**, 201–209. [https://doi.org/10.1016/0167-8655\(89\)90055-x](https://doi.org/10.1016/0167-8655(89)90055-x) (1989).
42. Deng, G. & Cahill, L. An adaptive gaussian filter for noise reduction and edge detection. In *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, vol. 3, 1615–1619, <https://doi.org/10.1109/NSSMIC.1993.373563> (1993).
43. Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302. <https://doi.org/10.2307/1932409> (1945).
44. Heimann, T. et al. Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging* **28**, 1251–1265. <https://doi.org/10.1109/tmi.2009.2013851> (2009).
45. Huttenlocher, D., Klanderman, G. & Rucklidge, W. Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 850–863. <https://doi.org/10.1109/34.232073> (1993).
46. Wolff, K.-D., Ervens, J., Herzog, K. & Hoffmeister, B. Experience with the osteocutaneous fibula flap: an analysis of 24 consecutive reconstructions of composite mandibular defects. *J. Cranio-Maxillofac. Surg.* **24**, 330–338 (1996).
47. Shpitzer, T. et al. Leg morbidity and function following fibular free flap harvest. *Ann. Plast. Surg.* **38**, 460–464 (1997).
48. Chhabra, A. et al. Treatment of chronic nonunions of the humerus with free vascularized fibula transfer: A report of thirteen cases. *J. Reconstr. Microsurg.* **25**, 117–124. <https://doi.org/10.1055/s-0028-1090624> (2008).
49. Ruikar, D. D., Santosh, K. C. & Hegadi, R. S. Automated fractured bone segmentation and labeling from CT images. *Journal of Medical Systems* **43**, <https://doi.org/10.1007/s10916-019-1176-x> (2019).
50. Ma, Y. & Luo, Y. Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network. *Inform. Med. Unlocked* **22**, 100452. <https://doi.org/10.1016/j.imu.2020.100452> (2021).
51. Kim, H., Lee, K., Lee, D.-C. & Baek, N. 3d reconstruction of leg bones from x-ray images using cnn-based feature analysis. *2019 International Conference on Information and Communication Technology Convergence (ICTC)* 669–672 (2019).
52. den Broeck, J. V., Vereecke, E., Wirix-Speetjens, R. & Sloten, J. V. Segmentation accuracy of long bones. *Med. Eng. Phys.* **36**, 949–953. <https://doi.org/10.1016/j.medengphys.2014.03.016> (2014).

53. Wallner, J., Mischak, I. & Egger, J. Computed tomography data collection of the complete human mandible and valid clinical ground truth models. *Scientific Data* 2019 6:1 6, 1–14. <https://doi.org/10.1038/sdata.2019.3> (2019).
54. Kim, T. et al. Accuracy of a simplified 3D-printed implant surgical guide. *J. Prosthet. Dent.* **124**, 195–201.e2. <https://doi.org/10.1016/j.prosdent.2019.06.006> (2020).
55. Li, Y. et al. Automated segmentation of vertebral cortex with 3d u-net-based deep convolutional neural network. *Front. Bioeng. Biotechnol.* **10**, 996723 (2022).
56. Jha, N. et al. Fully automated condyle segmentation using 3D convolutional neural networks. *Sci. Rep.* **12**, 20590 (2022).

Author contributions

S.R. and T.P. designed the study. S.R., T.P., and M.W. elaborated the methodology. J.S.N. performed technical implementation of the two-stage 3D U-Net architecture. F.P., F.H., and A.M. supervised the clinical aspects of the study, ensuring the relevance to surgical applications. M.W. provided guidance on data analysis. S.R. and J.S.N. wrote the main manuscript text, while S.R., M.W. and S.J. provided critical revisions. J.S.N and S.R. prepared the figures. All authors reviewed and approved the final manuscript. S.R., T.P., and A.M. are co-founders of Inzipio, a spin-off from the department of Oral and Maxillofacial Surgery at RWTH Aachen University Hospital.

Funding

Open Access funding enabled and organized by Projekt DEAL. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest

S.R., T.P., and A.M. are co-founders of Inzipio GmbH, a spin-off from the department of Oral and Maxillofacial Surgery at RWTH Aachen University Hospital. J.S.N., S.J., M.W., F.P., and F.H. have no financial or proprietary interests in any material discussed in this article.

Ethical approval

Institutional approval (EK 260/20) of the local ethics committee of RWTH Aachen University Hospital was obtained. All methods were carried out in accordance with relevant guidelines and regulations.

Informed consent

Due to the retrospective nature of the study, the Independent Ethics Committee of the Faculty of Medicine of RWTH Aachen University Hospital waived the need of obtaining informed consent.

Competing interests

S.R., T.P., and A.M. are co-founders of Inzipio, a spin-off from the department of Oral and Maxillofacial Surgery at RWTH Aachen University Hospital.

Additional information

Correspondence and requests for materials should be addressed to S.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025