Single Bone Modeler: deep learning bone segmentation for cone-beam CT

Eleonora Tiribilli
Dept. of Information Engineering
University of Florence
Florence, Italy
eleonora.tiribilli@unifi.it

Ernesto Iadanza
Chair, IFMBE Council of Societies
Dept. of Medical Biotechnologies
University of Siena
Siena, Italy
ernesto.iadanza@unisi.it

Leonardo Bocchi
Dept. of Information Engineering
University of Florence
Florence, Italy
leonardo.bocchi@unifi.it

Abstract—The accurate segmentation and modeling of bones play a crucial role in diagnosis and surgical planning in orthopedics. Traditional methods face challenges in capturing the fine details and complex structures present in cone-beam computed tomography (CBCT) scans. This paper introduces a novel deep learning-based workflow to precisely segment bone in CBCT scans of complex areas such as extremities, the Single Bone Modeler (SBM). It involves three main steps: bone segmentation, separation and 3D modeling. To achieve highly accurate bone segmentation, a dedicated U-Net architecture is developed and compared to a SegNet. Furthermore, we compare two different training strategies *axial* training and *multi-planar training*, when dealing with CBCT data. The separation of bones is performed through a watershed algorithm, and the structure of interest is subsequently modeled in 3D. The efficacy of proposed deep learning approaches is assessed, and outcomes are compared to benchmark techniques using two metrics: Jaccard Index (JI) and Dice Coefficient (DC). Results demonstrate the superior performance in bone segmentation of the proposed U-Net trained with *Multi-Planar* training, achieving a JI of 0.941 ± 0.031 and a DC of 0.970 ± 0.015. The entire workflow is further evaluated for its capacity to isolate specific bone, showcasing significant improvement over benchmark methods. In conclusion, the SBM enhances the precision of bone segmentation in high-resolution CBCT scans. The results suggest the potential for reliable and efficient extremity bone segmentation, with implications for improved applications in orthopedics.

Index Terms—Deep learning, segmentation, cone-beam CT, surgical planning, U-Net, multi-planar training, extremity segmentation, single bone modeler

I. INTRODUCTION

In contemporary medical practices, cone-beam computed tomography (CBCT) has become integral to oral and maxillofacial surgery and is showing promising expansion into orthopedics. CBCT offers advantages such as high-resolution imaging of hard tissues, cost-effectiveness, lower radiation exposure and a more compact size. These factors make CBCT effective in emergency departments and surgical rooms, especially for diagnosis and surgical planning [1], [2]. Accurate automatic segmentation of extremities, including hand, feet, ankle and foot, can be valuable in handling anatomical intricacies in diagnosis, pre-planning and post-planning phases. Extremities, unlike long bones, exhibit low contrast, weak bone boundaries, varying densities of cancellous tissue and narrow inter-bone

spacing, making bone segmentation a challenging task. Moreover, extremities like the foot and ankle, hand and wrist consist of numerous small, asymmetrical-shaped structures and densities. Image processing solutions must be designed to facilitate and expedite this work, minimizing manual interaction [3]. This paper proposes a deep learning-based approach to address these challenges and improve accurate bone segmentation in extremities.

Previous works by Klein et al. [4], Krawczyk et al. [5], and Nougachi et al. [6] demonstrated the feasibility of using deep learning techniques for the segmentation of bones in conventional CT scans. These studies highlighted the efficacy of neural networks in automatically delineating bone structures, aiding in various medical applications. However, when transitioning to CBCT pixel values may fluctuate because of artifacts such as shading, cupping and beam-hardening. In CBCT, grayscale values are directly associated with X-ray attenuation and lack the standardization provided by Hounsfield Unit (HU) calibration in conventional CT scans. This difference introduces nuances and variations in image intensity characteristics, leading to challenges in the segmentation process.

To the best of our knowledge, no study has been conducted that applies deep learning methodologies to extremity segmentation in CBCT. This paper introduces Single Bone Modeler (SBM), a novel workflow to separate and model individual bones in complex anatomical structures such as extremities. This work focuses primarily on the development of an accurate segmentation module, employing a deep learning approach.

II. MATERIALS AND METHODS

The Single Bone Modeler (SBM) comprises three primary steps, as illustrated in fig. 1. Initially, bones undergo segmentation and are separated from the surrounding soft tissues. Subsequently, each bone is isolated through the application of the watershed algorithm. Finally, the isolated bone is modeled in three dimensions.

A. Deep learning-based bone segmentation

The core part of the Single Bone Modeler lies in accurate bone segmentation, which utilizes deep learning techniques.
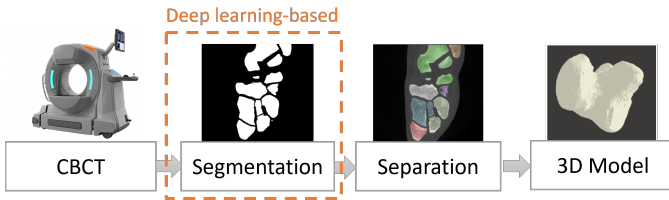
Fig. 1. The Single Bone Modeler workflow comprises three main steps: bone segmentation, bone separation, and modeling.

An in-house annotated dataset of CBCT scans was created for training and evaluating the proposed deep-learning model. Extremity anatomical preparations were scanned using the commercial machine See Factor CT3 (Imaginalis, Florence, Italy). Scan reconstruction was performed using the Feldkamp, Davis, and Kress (FDK) algorithm [7]. The volumetric data has an isotropic resolution of 0.2 mm and dimensions of $512 \times 512 \times 512$ pixels. A total of 20 CBCT scans were included in the study. Ground truth labels were generated by masking the scans using the software 3D Slicer. Each scan was acquired with varying acquisition parameters (kV and mA). For each volume, a different threshold was applied to isolate cancellous tissues. Further, segmentation refinement was performed manually using the 3D Brush tool. This tedious process resulted in a precise binary mask where hard tissues were designated as the foreground.

Bone segmentation was treated as a binary segmentation task. We developed and trained two different segmentation architectures: U-Net and SegNet. The U-Net [8] has a contracting path followed by an expansive path. The downstream part of the network is responsible for capturing high-level features and reducing spatial dimensions. It consists of five levels, each containing three convolutional blocks. The upstream part of the network is responsible for upsampling and combining high-level features from the downsampled levels. It starts from the last downsampled level and iteratively applies transposed convolution followed by concatenation with the corresponding skip connections from the downsampled path. The activation function is a sigmoid, suitable for binary segmentation tasks.

The second network is a SegNet [9]. It has an encoder-decoder architecture characterized by a symmetrical contracting and expanding structure. In contrast to U-Net, which employs skip connections, SegNet achieves upsampling by utilizing pooling indices. These indices store the locations of max pooling during the downsampling phase. The encoder part is composed by three sets of convolutional layers. Max pooling is applied after each set of convolutional layers, and the pooling indices are saved. In the decoder part max-unpooling with the saved pooling indices is applied to upsample the feature maps.. The final output is a sigmoid activation function, suitable for binary segmentation tasks. Both the U-Net and SegNet architectures were implemented in Python using the TensorFlow and Keras libraries.

The training was done on a workstation equipped with GeForce RTX 2070 SUPER (NVIDIA, Santa Clara, California).

In the domain of bone segmentation, there is an imbalance between the amounts of foreground and background. To address this imbalance, we employ a combined approach utilizing cross-entropy and dice loss as proposed by Klein et al. [4]. The Adam optimizer [10], with an initial learning rate of 10^{-3} was used. The networks undergo training for 250 epochs with a batch size of 16, incorporating learning rate decay and early stopping methods as callbacks in the network. To reduce overfitting, we implemented data augmentation, including a $\pm 15°$ rotation, a random horizontal flip and adjustments to the brightness range.

The dataset was organized by volumes to ensure that each slice from a specific volume was either in the training set or the test set, preventing data leakage. This contributes to preserving the data independence across the training and test sets, which is necessary for a trustworthy model evaluation and accurate generalization of the model's performance on unknown data. In particular, fourteen volumes, comprising seven feet and seven hands, were used for training the models. Four additional volumes, two feet and two hands, were dedicated to testing purposes and excluded from the training phase. Furthermore, two volumes (one foot and one hand) were set aside as a validation set. This validation set is distinct from both the training and test sets and was used during model training to prevent over-fitting. Each volume has been normalized relative to the maximum gray level of the volume itself.

Two different training strategies were assessed and compared during the training process. The first strategy is the conventional 2D training approach widely employed for training medical image volumes, referred to as *axial* training. We use axial slices of 512×512 pixels as inputs. To harness the isotropic voxel nature of cone-beam CT, we experimented with an alternative method that we called *Multi-Planar training* (MPT). This approach involves using images obtained from all three views, axial sagittal and frontal, thereby providing the opportunity to triple the number of slices while maintaining the same number of CBCT scans used in the training phase.

To quantitatively assess the performance of the first stage of our SBM and facilitate comparison, we compared the predictions of deep learning networks to the ground truth by calculating the Dice coefficient (DC) and the Jaccard index (JI). Moreover, to assess the quality of bone segmentation and the performance of deep learning in this binary segmentation task, we compared the results obtained with the developed networks with a thresholding method and with a graph-cut algorithm proposed by Boykov et al. [11] and implemented as described by Tiribilli et al. [12].

B. Bone separation and modeling

The segmented bones were separated using a watershed algorithm, that isolated individual bones within the anatomical region under study. The algorithm executed a distance transform over the binary mask, then a threshold was applied to find markers of each object. These markers served as seeds for the watershed algorithm, which filled each bone with a

TABLE I
COMPARISONS OF PERFORMANCE METRICS

	Thresholding	Graph-cut	U-Net AT [1]	U-Net MPT [2]	SegNet AT[1]	SegNet MPT[2]
Jaccard Index	0.663 ± 0.089	0.794 ± 0.048	0.917 ± 0.058	**0.941 ± 0.031**	0.903 ± 0.120	0.932 ± 0.018
Dice Coefficient	0.792 ± 0.110	0.885 ± 0.120	0.957 ± 0.015	**0.970 ± 0.015**	0.948 ± 0.018	0.964 ± 0.003

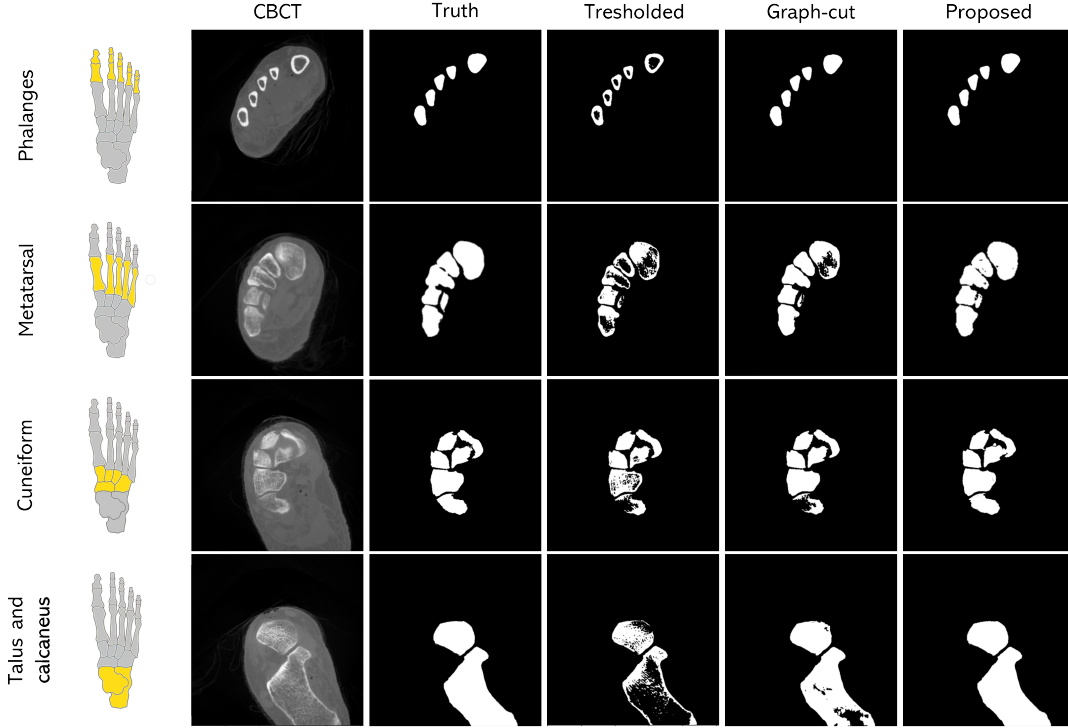[1] Axial training   [2] Multi Planar training

Fig. 2. Segmentation of bone over an anatomical preparation of a human foot. Comparison between thresholding, graph-cut and the proposed method that achieves best results in terms of JI and DC, U-Net trained with MPT.

different label. Subsequently, through a graphical interface, the user selected a bone of interest and generated a 3D mesh using the marching cubes algorithm [13].

To assess the ability to isolate a single bone, six bones of interest from the hand and foot were chosen and separated using our method. We evaluated the efficacy of the novel SBM by utilizing a threshold, a graph-cut, and the proposed deep learning technique as the segmentation module. We compared the results in terms of Dice coefficient.

III. RESULTS

Quantitative and qualitative assessments are conducted on the dedicated test dataset, comparing the two networks' performances against ground truth annotations using JI and DC.

U-Net, SegNet, conventional thresholding and graph-cut are compared in order to verify the efficacy of the deep learning algorithms in bone segmentation. Metrics are calculated for each of the four volumes under test and then mean and standard deviation are computed and reported in table I. Results indicate that *Multi-Planar Training* (MPT) performs best in both architectures. The U-Net trained with MPT achieved the
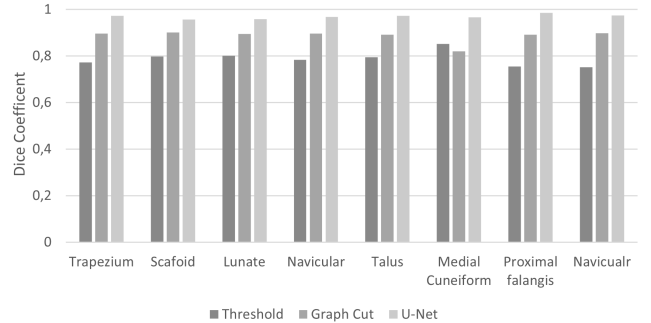

Fig. 3. Performances of the presented workflow over six bony structures. Comparison of segmentation performed with thresholding, grap-cut and the proposed U-Net.

best results, with a JI of 0.941 ± 0.031 and a DC of 0.970 ± 0.015.

Qualitative results are shown in fig 2. The proposed network, as well as the thresholding and graph-cut approaches, work well for cortical bone. The primary challenges emerge when attempting to segment spongy bone and bone marrow.

Due to their lower density, the gray level values of these tissues closely resemble those of soft tissue. As predicted, a thresholding-based approach struggles to segment these tissues effectively. Moreover, this approach is not suitable due to the pixel intensity instability of CBCT. The mentioned problems are partly solved by the graph-cut approach, as it does not rely only on gray-level values. As expected, the best performances are achieved with deep learning approaches. Also qualitative results confirm the superior performance of the proposed U-Net trained with MPT in handling complex anatomical structures, emphasizing its practical advantages in extremity segmentation.

The capacity of the suggested approach to isolate a single bone in the proposed SBM workflow was then assessed. Fig. 3 shows the performance comparison in terms of DC employing a threshold, graph-cut, and U-Net trained with MPT as the segmentation module.

In the segmentation of all assessed single bones, we find that the deep learning strategy greatly outperforms both the graph-cut and thresholding methods in terms of individual bone isolation. The best results are achieved in the proximal phalanges.

IV. CONCLUSIONS

In conclusion, our study introduces a novel workflow, the Single Bone Modeler which leverages deep learning techniques for precise extremity bone segmentation in high-resolution CBCT. We compared a U-Net and a SegNet that we developed and trained from scratch using an in-house dataset. We used the Adam optimizer, learning rate decay and data augmentation to prevent overfitting. Furthermore, we analyzed the effects of two different training methods, *axial* and the *MPT*, on the segmentation results. We evaluated the performance of deep learning approaches using two metrics DC and JI and compared the results with a simple thresholding and graph-cut. U-Net trained with MPT achieves the best results in the segmentation module. Moreover, we evaluated the entire SBM workflow by applying it to six different anatomical parts. The results of our study demonstrate that the deep learning approach, employing the proposed U-Net trained with MPT significantly enhances the isolation of individual bones in extremity bone segmentation when compared to traditional methods such as graph-cut and thresholding.

The enhanced precision achieved with deep learning models signifies the potential for more reliable and efficient extremity bone segmentation in diagnosis, pre-planning and post-planning phases. Moreover, our results support the use of CBCT technology in orthopedics, which leads to lower radiation exposure, cost-effectiveness and its use in emergency rooms due to its small size.

It is feasible to apply SBM to additional anatomical regions such as vertebrae, shoulders, knees, etc. because the network is trained for binary segmentation rather than to offer a label for each unique bone. We plan to extend our project to identify and label bones developing an instance segmentation network. However, it is important to note that for each application on a new target, a new labeled dataset and new network training are required. Given the accurate segmentation results achieved with this method, another potential future application consists of the segmentation and modeling of complex fractures.

REFERENCES

[1] T. Jacques, V. Morel, J. Dartus, S. Badr, X. Demondion, and A. Cotten, "Impact of introducing extremity cone-beam CT in an emergency radiology department: A population-based study," *Revue de Chirurgie Orthopedique et Traumatologique*, vol. 107, pp. 210–217, 2021.

[2] R. Grassi, E. Guerra, and D. Berritto, "Bone fractures difficult to recognize in emergency: May be cone beam computed tomography (CBCT) the solution?" *Radiologia Medica*, vol. 128, no. 1, pp. 1–5, 2023.

[3] E. Tiribilli, L. Manetti, L. Bocchi, and E. Iadanza, "Bones segmentation techniques in computed tomography a survey," in *IFMBE Proceedings*, vol. in press, 2022.

[4] A. Klein, J. Warszawski, J. Hillengaß, and K. H. Maier-Hein, "Automatic bone segmentation in whole-body CT images," *IJCARS*, vol. 14, no. 1, pp. 21–29, 2019.

[5] Z. Krawczyk and J. Starzynski, "Segmentation of bone structures with the use of deep learning techniques," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 69, 2021.

[6] S. Noguchi, M. Nishio, M. Yakami, K. Nakagomi, and K. Togashi, "Bone segmentation on whole-body ct using convolutional neural network with novel data augmentation techniques," *Computers in Biology and Medicine*, vol. 121, 2020.

[7] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America*, vol. 1, no. 6, pp. 612–619, 1984.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, vol. 9351, pp. 234–241, 2015.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on PAMI*, 2017.

[10] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 2015.

[11] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in n-d images," vol. 1, 2001, pp. 105–112.

[12] E. Tiribilli, L. Manetti, L. Bocchi, and E. Iadanza, "Extremity bones segmentation in cone beam computed tomography, a novel approach," in *IFMBE Proceedings*, vol. 93, 2024, pp. 278–284.

[13] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.