

Article

Deep Learning-Based Workflow for Bone Segmentation and 3D Modeling in Cone-Beam CT Orthopedic Imaging

Eleonora Tiribilli ^{1,2,*} and Leonardo Bocchi ^{1,2}¹ Information Engineering Department, University of Florence, 50139 Firenze, Italy; leonardo.bocchi@unifi.it² Eido Lab, University of Florence, 50139 Firenze, Italy

* Correspondence: eleonora.tiribilli@unifi.it

Abstract: In this study, a deep learning-based workflow designed for the segmentation and 3D modeling of bones in cone beam computed tomography (CBCT) orthopedic imaging is presented. This workflow uses a convolutional neural network (CNN), specifically a U-Net architecture, to perform precise bone segmentation even in challenging anatomical regions such as limbs, joints, and extremities, where bone boundaries are less distinct and densities are highly variable. The effectiveness of the proposed workflow was evaluated by comparing the generated 3D models against those obtained through other segmentation methods, including SegNet, binary thresholding, and graph cut algorithms. The accuracy of these models was quantitatively assessed using the Jaccard index, the Dice coefficient, and the Hausdorff distance metrics. The results indicate that the U-Net-based segmentation consistently outperforms other techniques, producing more accurate and reliable 3D bone models. The user interface developed for this workflow facilitates intuitive visualization and manipulation of the 3D models, enhancing the usability and effectiveness of the segmentation process in both clinical and research settings. The findings suggest that the proposed deep learning-based workflow holds significant potential for improving the accuracy of bone segmentation and the quality of 3D models derived from CBCT scans, contributing to better diagnostic and pre-surgical planning outcomes in orthopedic practice.

Keywords: cone beam CT; segmentation; 3D models; bones; orthopedic; neural networks; U-Net



Citation: Tiribilli, E.; Bocchi, L. Deep Learning-Based Workflow for Bone Segmentation and 3D Modeling in Cone-Beam CT Orthopedic Imaging. *Appl. Sci.* **2024**, *14*, 7557. <https://doi.org/10.3390/app14177557>

Academic Editor: Claudio Belvedere

Received: 28 June 2024

Revised: 8 August 2024

Accepted: 14 August 2024

Published: 27 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cone beam computed tomography (CBCT) is a state-of-the-art 3D medical imaging technology that employs a cone-shaped X-ray beam to produce high-resolution images of anatomical structures. One of the standout attributes of CBCT is its high spatial resolution, enabling the visualization of fine details, particularly in hard tissues. This feature makes CBCT emerge in orthopedics, where it supports implant planning, joint assessment, and the evaluation of traumatic injuries, ultimately enhancing patient outcomes and delivering personalized care [1]. Additionally, compared to traditional CT imaging, CBCT can deliver detailed images at a significantly lower radiation dose, making it a valuable tool in emergency departments and surgical rooms, particularly for diagnosis and pre-surgical planning [2,3].

Currently, CBCT is especially beneficial in maxillofacial and oral surgery. For instance, Bhat et al. [4] proposed a workflow for dental pre-surgical planning using immersive virtual reality and CBCT data. Lalonde et al. [5] utilized 3D-printed models derived from CBCT data to treat a rare type of dens invaginatus in a mandibular incisor, highlighting the utility of 3D models in guiding proper treatment. The generation of accurate 3D models relies on precise bone segmentation [6].

Due to the small dimensions of the scanning machine, CBCT technology is particularly feasible, not only for the scanning of the head and maxillofacial district, but also for other extremity anatomical regions such as foot bones, hands and wrists, limbs, and joints.

However, bone segmentation in these anatomical regions is challenging. Unlike long bones, extremities and joints exhibit weak bone borders, fluctuating densities of cancellous tissue, and small inter-bone spacing. Furthermore, extremities consist of various tiny, asymmetrically formed structures with different densities, such as those in the hand, wrist, and foot. Additionally, unlike conventional CT, CBCT grayscale values are directly associated with X-ray attenuation and lack the standardization provided by Hounsfield Unit (HU) calibration, introducing nuances and variations in image intensity characteristics, which complicate the segmentation process.

The existing literature extensively covers bone segmentation in CBCT, primarily in dental and maxillofacial surgery contexts [7]. Examples of CBCT segmentation for extremities mainly focus on weight-bearing scans of the foot and ankle. However, there is a deficiency of studies related to extremities and joint segmentation from CBCT scans.

The U-Net architecture [8] is a widely used convolutional neural network (CNN) architecture for medical imaging segmentation. U-Net has been applied to conventional CT segmentation of various bony structures [9–15]. In the context of CBCT data, applications of U-Net have been demonstrated in the maxillofacial region. For instance, Lin et al. [16] used U-Net to accurately segment the mandibular canal in CBCT data, and Zhang et al. employed DBA-U-Net for maxillary sinus segmentation [17].

The first CNNs developed to segment three-dimensional images acquired using magnetic resonance imaging (MRI) or computed tomography (CT) were trained on two-dimensional image slices [18]. The majority of these CNNs used axial slices as the input [19,20], due to the high in-plane resolution compared to the slice thickness [21]. CBCT instead offers volumetric high resolution since 3D CBCT images have isotropic voxels, which means that the voxel dimensions are the same in all three spatial dimensions. The axial training does not take into account any 3D information. Zhou et al. [22] proposed to train separate CNNs for each of the three orthogonal slices to classify the voxel through a majority voting scheme.

This paper proposes a deep learning-based bone segmentation tool for CBCT orthopedic imaging. The aim is to provide an easy-to-use segmentation and 3D modeling workflow for intricate anatomical districts such as ankles, wrists, and joints. This workflow consists of three main steps: bone segmentation, separation, and 3D modeling, designed to be highly intuitive and user-friendly. The main innovation of our work lies in the tool's ability to deliver high-quality results with minimal user interaction. This work utilizes a U-Net architecture trained with a strategy particularly feasible for cone beam CT, to leverage the isotropic nature of CBCT data. In a previous work [23], the authors introduced the workflow focusing on segmentation results. This paper extends its findings by providing a detailed comparison between the proposed and state-of-the-art methods, along with an evaluation of the 3D models and the user interface.

2. Materials and Methods

This paper proposes a deep learning-based workflow using the U-Net architecture to accurately segment extremities and joints in high-resolution cone beam CT scans. As illustrated in Figure 1, bones in images acquired through a commercial CBCT scan undergo segmentation and are separated from the surrounding soft tissues. Binary segmentation is performed using U-Net. Then, the bones are separated by the application of the watershed algorithm. Finally, the user through the user interface can select the bone to be modeled in three dimensions. A simple and intuitive user interface was developed for the bone segmentation and modeling tool. The workflow was optimized to require minimal user intervention, relying primarily on deep learning algorithms, which automate most of the process.

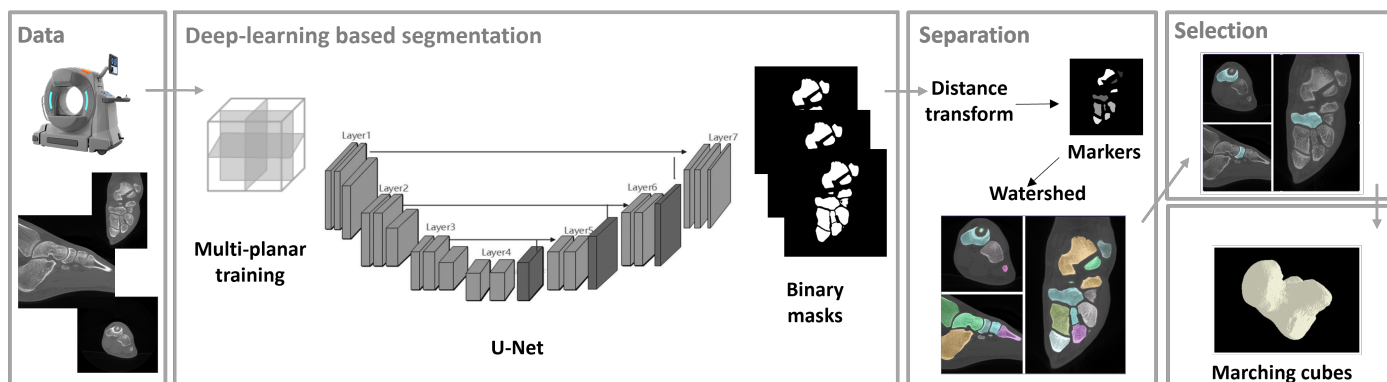


Figure 1. Workflow for the deep learning-based segmentation: The process begins with acquiring data from a CT scanner, and using a U-Net architecture trained with a multi-planar approach, binary masks are generated. These masks undergo separation and selection to isolate and identify individual structures, resulting in detailed segmented models.

2.1. Binary Semantic Segmentation

2.1.1. Data

Anatomical preparations were scanned using a commercial CBCT device, the See Factor CT3 (Imaginallis, Florence, Italy). The Feldkamp, Davis, and Kress (FDK) algorithm was employed to reconstruct the scans [24]. The volumetric data include $512 \times 512 \times 512$ pixels and boast an isotropic resolution of 0.2 mm. Scans were performed under varying acquisition parameters (kV and mA).

To train and evaluate the proposed deep learning models, an in-house annotated dataset of CBCT scans was created. A total of fourteen CBCT scans, acquired with different parameters, were considered. These scans focus on extremities, where bone segmentation is particularly challenging due to the high number of adjacent bones and the similar gray levels between spongy tissues and soft tissues.

To generate ground truth labels, the scans were masked using 3D Slicer (version 5.6), an open-source software developed by the Surgical Planning Laboratory at Brigham . Initially, a custom threshold was applied to each scan to isolate cancellous tissues. Subsequently, manual segmentation refinements were performed using the 3D Brush tool in 3D Slicer to ensure accuracy.

2.1.2. Network Architectures

The first stage of the proposed workflow involves binary semantic segmentation. A simple network architecture with fewer parameters was implemented to achieve faster inference times, which is critical for ensuring a good user experience in clinical applications. This decision was based on the need for a practical tool that can be easily adopted in routine workflows without requiring extensive computational resources. Consequently, the approach was compared with other methods that share a similar focus on simplicity and efficiency. To achieve this task, a neural network based on the well-known U-Net architecture was implemented, and its performances were compared with another encoder–decoder architecture designed for accurate binary segmentation tasks: SegNet.

The U-Net architecture [8] is a CNN designed for semantic segmentation tasks, where the goal is to assign a label to each pixel in an input image. It has an encoder–decoder path. The encoder path consists of convolutional and pooling layers, which progressively downsample the input image. Each convolutional layer is followed by a rectified linear unit (ReLU) activation function. Max pooling operations are applied to reduce the spatial dimensions of the feature maps while increasing the receptive field. The decoder path consists of upsampling and convolutional layers, which gradually upsample the feature maps to the original input resolution. Skip connections are introduced between corresponding layers in the encoder and decoder paths to preserve spatial information. These

connections directly link layers at the same spatial resolution in the encoder and decoder paths. By concatenating feature maps from the encoder with those in the decoder, skip connections enable the decoder to access high-resolution features from earlier stages of the network. This helps the decoder refine the segmentation masks by incorporating detailed spatial information that may have been lost during downsampling.

SegNet has an encoder–decoder architecture with a symmetrical contracting and expanding structure. Unlike U-Net, which uses skip connections, SegNet uses pooling indices to perform upsampling. These indices store the locations of max pooling during the downsampling phase and are used to retain fine-grained details during the upsampling. SegNet was implemented as described by Badrinarayanan et al. [25].

2.1.3. Training Strategies

Three distinct training strategies were assessed and compared for U-Net and SegNet. The first strategy, referred to as *axial* training, is the traditional 2D training method. Axial slices of 512×512 pixels were employed to train the developed networks. This strategy does not take into account any 3D information, and to overcome this, the so-called 2.5D training strategies were considered.

The second strategy is majority voting (MV), in which separate convolutional neural networks (CNNs) are trained for each of the three orthogonal slices. The prediction for a voxel is determined by aggregating predictions from all three CNNs, and the voxel is considered part of the foreground if at least two of the CNNs predict it as the foreground.

Lastly, an augmented 2D training strategy was evaluated, referred to as *multi-planar* training (MPT). In this approach, the network is trained with a dataset composed of slices from all three orthogonal planes (axial, sagittal, and frontal). Each batch, during the training process, contains images from the three views. The networks under evaluation include six variations: U-Net with axial training, MPT, and MV and SegNet with axial training, MPT, and MV training, as shown in Figure 2.

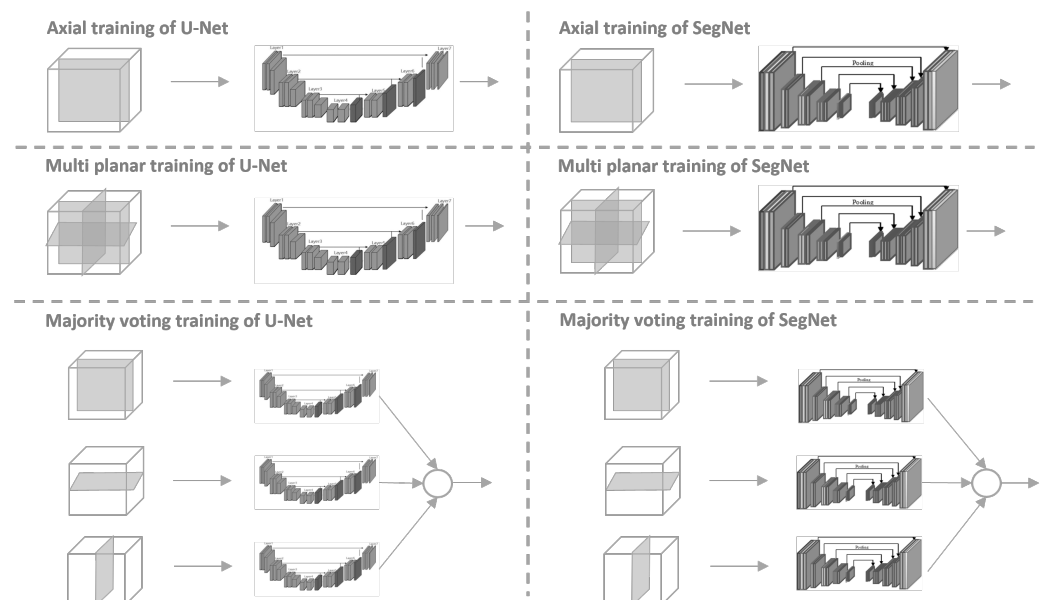


Figure 2. Comparison of training strategies for U-Net and SegNet: the diagram illustrates axial, multi-planar, and majority voting training methods for both U-Net and SegNet.

A workstation with a GeForce RTX 2070 SUPER (NVIDIA Corporation, Santa Clara, CA, USA) was used for the training. We employed an Adam Optimizer, with an initial learning rate of 10^{-3} . The networks were trained for 250 epochs using a batch size of 16. Early stopping and learning rate decay were integrated into the network as callbacks. Before training, a normalization of the pixel intensity values to standardize the input images was performed. Extensive data augmentation techniques were applied during

training. These techniques included random rotations, flips, shifts, and zooms, as well as the addition of noise and adjustments to brightness and contrast. Moreover, the dataset was arranged according to volumes to prevent data leakage, making sure that every slice from a given volume was either in the training set or the test set. In this way, the training and test sets' data independence was preserved, which is helpful for the generalization of the model's performance and an accurate model evaluation. Four volumes were used for testing and eight volumes for training. Moreover, two volumes distinct from the training and test sets were considered to be the validation set.

2.1.4. Metrics

To quantitatively assess the performance of our training strategies and facilitate comparison, the predictions of our networks were compared to the ground truth by calculating the following well-established segmentation metrics: the Jaccard index (JI) and the Dice coefficient (DC). The Jaccard index evaluates the similarity between two sets by evaluating the ratio of the size of their intersection to the size of their union.

$$JI(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

The Dice coefficient serves as a metric for quantifying the similarity between two sets. Specifically, it is calculated by dividing the size of the intersection of the sets by the average size of the sets.

$$DC(A, B) = \frac{2 \cdot |A \cap B|}{|A| + |B|} \quad (2)$$

In the given context, set A comprises pixels labeled as positive (e.g., bone pixels) in the ground truth, while set B encompasses pixels predicted as bone by the CNN being evaluated. Both coefficients have a range from 0 to 1, with 1 being the optimal value.

2.2. Instance Segmentation

The results of binary segmentation were processed to label each bone separately. Initially, a binary hole-filling operation was applied, effectively closing the little void space within the segmented regions. Subsequently, a binary erosion operation was employed to refine the boundaries of the filled mask. Following the pre-processing stages, a distance transform was computed based on the filled binary mask. The resulting distance map encodes the Euclidean distances from each foreground pixel to the nearest background pixel. Using the computed distance transform, markers for the watershed algorithm are identified through a thresholding process.

The watershed algorithm was then applied, utilizing the negative of the distance transform as a gradient image and the identified markers as seeds for region segmentation.

2.3. Three-Dimensional Model

To accurately model the bone in 3D in this study, a function implementing the Lewiner algorithm [26] for extracting iso-surfaces from 3D volumetric data was implemented. This algorithm is an enhanced version of the original marching cubes algorithm [27], providing faster performance and resolving ambiguities to ensure topologically correct results. Specifically, it uses a refined set of lookup tables to handle all possible configurations of surface intersections within a cube. This approach not only improves the accuracy of surface reconstruction, but also ensures the robustness of the generated meshes, making it particularly suitable for complex and high-resolution datasets. The algorithm's implementation allows for efficient processing and accurate depiction of iso-surfaces, contributing to the visualization and analysis of the 3D data generated by the segmentation process in our research.

To evaluate the results in terms of modeling, a surface-based metric was used to calculate the distance between the obtained mesh and a reference mesh derived from the manual segmentation of bone. The Hausdorff distance is a measure of the extent to

which two subsets of a metric space are close to each other. More formally, given two non-empty subsets A and B of a metric space with a distance function d , the Hausdorff distance $d_H(A, B)$ is defined as:

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(b, a) \right\} \quad (3)$$

where the following applies:

- $\inf_{b \in B} d(a, b)$ measures the shortest distance from a point a in set A to any point in set B .
- $\sup_{a \in A} \inf_{b \in B} d(a, b)$ then considers the farthest distance of these shortest distances for all points in A . This ensures that every point in A is close to some point in B .
- Similarly, $\sup_{b \in B} \inf_{a \in A} d(b, a)$ ensures that every point in B is close to some point in A .

The maximum of these two quantities is taken to make the distance symmetric and to reflect the greatest extent to which one set can be far from the other.

In brief, the Hausdorff distance $d_H(A, B)$ is the greatest of all the distances from a point in one set to the closest point in the other set, ensuring that both sets are close to each other in a symmetrical sense. Smaller values of the Hausdorff distance indicate better performance.

2.4. User Interface

The user interface (UI) of the bone segmentation and modeling tool has been designed with a focus on simplicity and usability, and a schematic representation is given in Figure 3. The UI presents a clean and organized layout, with essential tools and options easily accessible. The main screen includes a central viewing area for the CBCT images and a sidebar with segmentation tools and options.

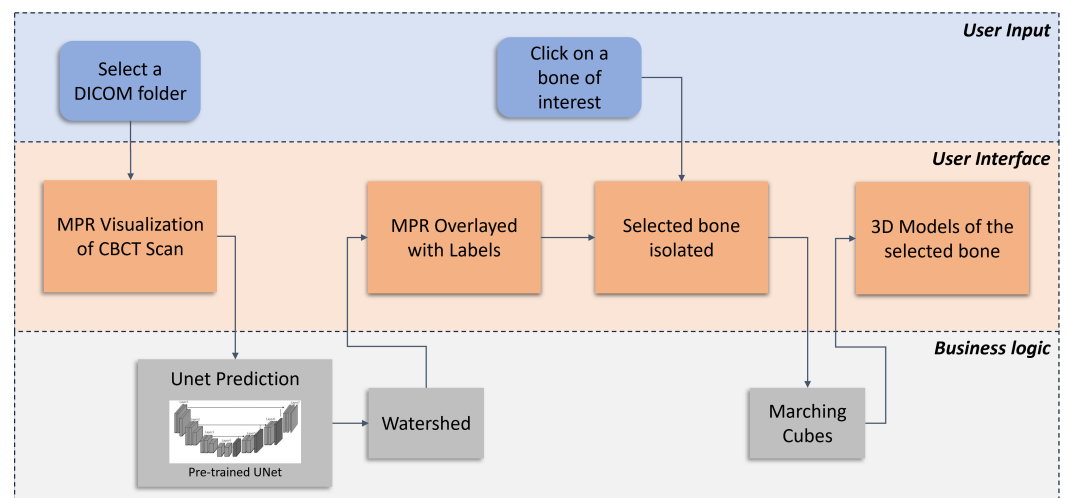


Figure 3. This diagram illustrates the user interface workflow for bone segmentation and 3D modeling from CBCT scans. The user selects a DICOM folder, and the interface allows visualizing the volume in multiplanar reformation (MPR). Users can click on a bone of interest, isolate the selected bone, and generate 3D models. The underlying business logic involves U-Net prediction for initial segmentation, the watershed for labeling, and modified marching cubes for the final 3D model construction.

The UI provides real-time feedback, allowing users to see the immediate effects of their actions. Additionally, the UI guides users through the segmentation and modeling process with tooltips, minimizing the learning curve for new users.

The interface allows the user to load a DICOM folder and visualize volume in multiplanar reformation (MPR) mode. The user can scroll through all the views. Labels are generated fully automatically, and the bone segmentation is superimposed over the MPR views with different colors. By clicking on a specific anatomical part, the user can visualize only the label related to the bone of interest, easily switching between different bones. Once the user selects the bone to be exported, he/she can save it as a triangular mesh. The UI is part of the Multimodal Biomedical Imaging Platform All-in-One software 2022 [28] developed by Imaginalis S.r.l. (Sesto Fiorentino, Italy), and this software was tested using a reliable usability testing protocol [29,30]. The UI design follows Nielsen's usability heuristics, focusing on simplicity, consistency, and error prevention [31].

3. Experiments

While the study primarily focuses on bone segmentation and modeling in specific anatomical regions (limbs, joints, extremities) using CBCT images, the underlying methodology has the potential to be generalized to other body parts and different types of medical imaging. The same deep learning framework and segmentation techniques were applied to assess the versatility of the tool. To evaluate the generalizability of the method, preliminary experiments were conducted on additional anatomical regions, including the spine and pelvis, as well as veterinary CBCT scans. The tool was used for the pre-planning phase of a canine acetabular cup insertion, in particular to generate the anatomical models of the femurs and pelvis to be printed.

4. Results

4.1. Binary Segmentation

To evaluate the first stage of this workflow, the developed U-Net was compared with the benchmark SegNet. To identify the most effective training strategies for the task, networks trained with various approaches on the same test subset were evaluated, using four volumes. The networks under evaluation included six variations: U-Net with axial training, MPT, and MV and SegNet with axial training, MPT, and MV training. Each network was tested individually using axial slices, sagittal slices, and frontal slices. For majority voting, three separate training sessions were conducted for axial, sagittal, and frontal orientations and then combining the results using a majority voting scheme. This comprehensive evaluation allowed us to rigorously compare the different training strategies and architectures, providing insight into the most effective methods for CBCT segmentation of bones.

In Figure 4, segmented images of the human foot are presented, featuring one axial, one sagittal, and one frontal view, alongside binary masks obtained from the six networks under evaluation. The proposed networks performed well, with challenges emerging mainly in segmenting frontal and sagittal views using axial-trained networks.

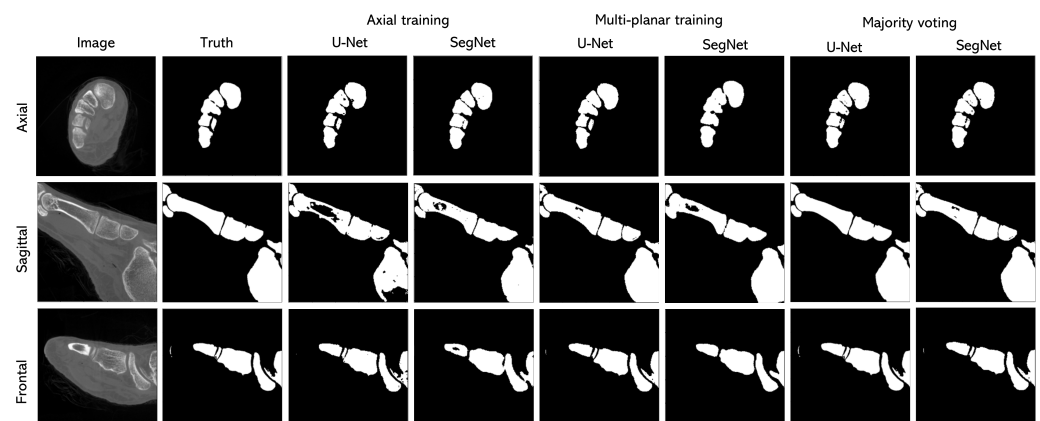


Figure 4. The figure shows the qualitative results of the six evaluated networks over the anatomical preparation of a human foot.

The quantitative evaluation of segmentation performance on experimental CBCT images utilized two metrics: the Jaccard index and Dice coefficient. These metrics were computed separately for each volume, and then, the mean and standard deviation of the results across the three test volumes were computed. Table 1 displays the results in terms of the JI and DC. The MPT networks exhibited the highest metrics ($JI = 0.939 \pm 0.010$, $DC = 0.969 \pm 0.005$), while *axial* training during sagittal and frontal tests yielded the lowest metrics ($JI = 0.861 \pm 0.014$, $DC = 0.899 \pm 0.003$). In terms of training and segmentation time, MPT requires more time than *axial* training, while MV requires three times the time requested for *axial* training due to the need to train three distinct networks to perform the voting scheme. *Axial* training took nearly 12 h, MPT 24 h, and MV training, involving three networks for the voting scheme, took almost 36 h. The segmentation of a volume ($512 \times 512 \times 512$ pixels) took 70 s for the *axial* and MPT and 220 s for the MV training. Although *majority voting* training achieved results as good as MPT, it required more computational time, as three predictions have to be made to complete the majority voting scheme.

Table 1. Results in terms of Jaccard index and Dice coefficient.

		Axial Training		Multiplanar Training		Majority Voting	
		U-Net	SegNet	U-Net	SegNet	U-Net	SegNet
Axial	JJ	0.904 ± 0.019	0.917 ± 0.029	0.939 ± 0.010	0.939 ± 0.014		
	DC	0.949 ± 0.011	0.957 ± 0.015	0.969 ± 0.005	0.968 ± 0.008		
Sagittal	JJ	0.861 ± 0.014	0.901 ± 0.021	0.936 ± 0.002	0.931 ± 0.020	0.927 ± 0.010 ¹	0.933 ± 0.018 ¹
	DC	0.899 ± 0.03	0.948 ± 0.011	0.967 ± 0.001	0.964 ± 0.012	0.962 ± 0.004 ¹	0.965 ± 0.009 ¹
Frontal	JJ	0.899 ± 0.006	0.897 ± 0.011	0.933 ± 0.010	0.937 ± 0.013		
	DC	0.946 ± 0.012	0.946 ± 0.006	0.965 ± 0.007	0.968 ± 0.009		

¹ Results obtained through a majority voting scheme, which uses axial, sagittal, and frontal views.

The results validated the proposed U-Net trained using MPT's superior performance in handling complicated anatomical structures, underscoring its practical utility in extremity binary segmentation. It is worth noting that SegNet also provides good results in terms of the evaluated metrics, but with a significantly higher number of parameters. This indicates that, while both architectures are effective, U-Net, particularly with MPT, offers a more computationally efficient solution without compromising the quantitative or quality results.

4.2. Three-Dimensional Model

To evaluate the quality of the 3D model derived from the proposed workflow, the models were compared using two pixel-based metrics, the DC and JJ, and one distance-based metric, the Hausdorff distance. Specifically, the mesh derived from U-Net with MPT segmentation was compared to those obtained using other methods, including SegNet with MPT, thresholded bone segmentation, and with a graph cut algorithm proposed by Boykov et al. [32] and implemented as described by Tiribilli et al. [33]

The results in term of the Jaccard index and Dice coefficient are presented in Table 2.

Table 2. Jaccard index and Dice coefficient of the 3D model of a human talus, obtained with the proposed workflow.

	Threshold	Graph Cut	SegNet	U-Net
JJ	0.74	0.77	0.83	0.85
DC	0.85	0.87	0.91	0.92

The Hausdorff distance was computed between the mesh produced by each segmentation method and the corresponding ground truth mesh. The Hausdorff distance provides

a measure of the maximum discrepancy between two point sets on the surfaces of the meshes, thus allowing assessing the accuracy and precision of the segmentation methods.

The results of this comparative analysis are presented in Table 3. This table highlights the max, mean, and standard deviation of the Hausdorff distances for each segmentation method on the target bone, thereby allowing us to determine which method produces the most accurate and reliable 3D models. By analyzing these distances, the effectiveness of each segmentation technique can be objectively evaluated. The proposed method, U-Net with MPT segmentation, demonstrated the lowest mean in terms of the Hausdorff distance. Figure 5 depicts a visualization of the Hausdorff distances for the target bone across the evaluated methods. In this context, blue areas on the models indicate minimal differences from the ground truth, while red areas signify substantial discrepancies. The model obtained via threshold-based segmentation shows high Hausdorff distances across the entire surface, indicating poor accuracy. The graph cut segmentation model has high distances in specific regions, reflecting localized segmentation errors. The U-Net and SegNet methods exhibit lower Hausdorff distances, with the U-Net model demonstrating the best overall performance, closely aligning with the ground truth. The color bar at the bottom provides a visual reference for the distance values, emphasizing the superiority of the U-Net segmentation approach.

Table 3. Max value, mean value, and standard deviation of the Hausdorff distance between the reference models under evaluation. The comparison includes models obtained using different segmentation techniques: threshold segmentation, graph cut segmentation, SegNet segmentation, and U-Net segmentation.

	Threshold	Graph Cut	SegNet	U Net
Max [mm]	7.9	8.53	2.41	3.81
Mean [mm]	1.48	1.00	0.10	0.067
Std [mm]	2.16	1.96	0.16	0.20

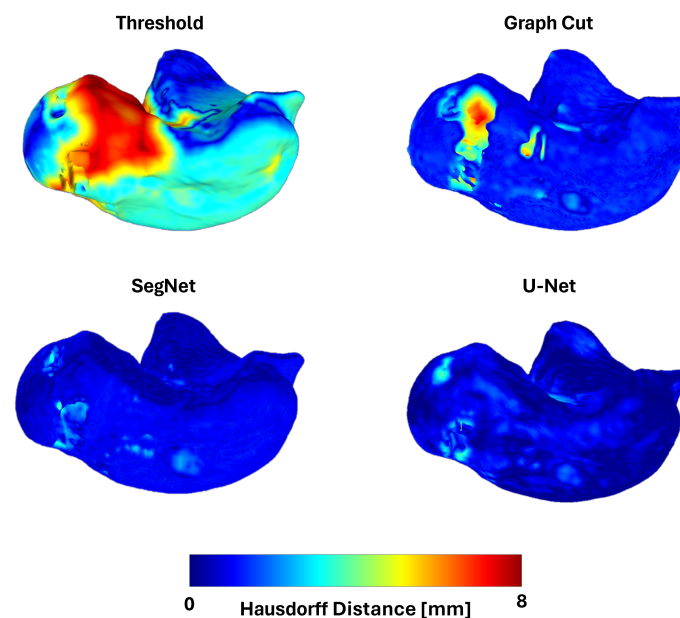


Figure 5. Visual representation of the Hausdorff distance. The comparison includes models obtained using different segmentation techniques: threshold segmentation, graph cut segmentation, SegNet segmentation, and U-Net segmentation.

4.3. User Interface

In Figures 6 and 7, the user interface is shown, specifically showcasing the segmentation and modeling of a human talus and a human hamate. These figures illustrate the effectiveness of the segmentation process and the ability to handle complex anatomical structures. Labels are generated automatically and superimposed on the MPR views. The interface provides a clear and intuitive visualization that aids in the accurate identification and segmentation of these structures. The user can isolate a single bone with a click on the bone of interest and export it as a mesh.

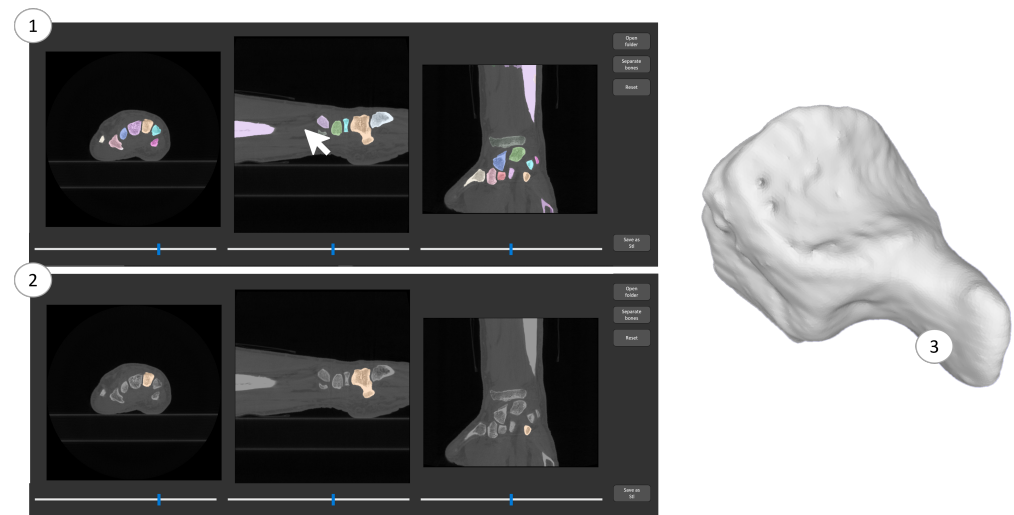


Figure 6. User interface providing segmentation and modeling of a hamate bone of a human wrist.

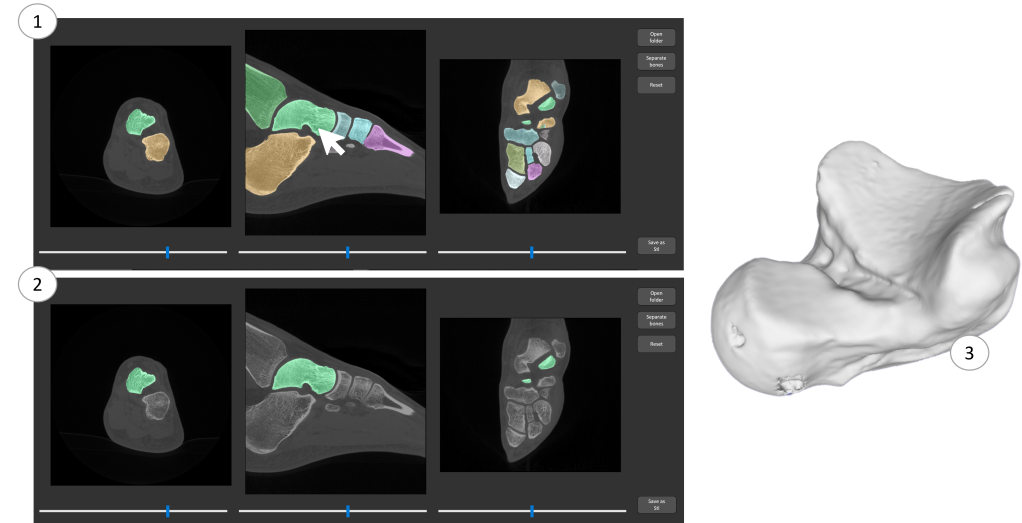


Figure 7. User interface providing segmentation and modeling of a talus of a human foot.

4.4. Experiment

The preliminary results indicate that the method can be effectively extended to other anatomical regions. For example, when applied to spine and pelvis images, our tool achieved segmentation accuracies comparable to those obtained for limbs and joints. Similarly, experiments using veterinary scans demonstrated the robustness of our approach across different types of medical imaging. Figure 8 shows the results achieved by printing the models obtained with the tool for the surgical planning of an acetabular cup insertion of a dog.



Figure 8. Patient-specific 3D model of a canine femur and pelvis obtained through the proposed workflow.

5. Discussion

The results of the presented study demonstrate the efficacy of using deep learning techniques, particularly the U-Net architecture, for the segmentation and 3D modeling of bones in CBCT orthopedic imaging.

The performances of the proposed U-Net and a benchmark SegNet architecture for CBCT segmentation of bones were compared, employing different training strategies including axial, MPT, and MV. The evaluation on three CBCT volumes revealed that both the U-Net and SegNet architectures achieve high segmentation accuracy, but with distinct differences in computational efficiency and parameter count. U-Net trained with MPT exhibited the highest performance metrics ($JI = 0.939 \pm 0.010$, $DC = 0.969 \pm 0.005$), particularly excelling in handling complex anatomical structures. This training strategy effectively leverages information from multiple planes, enhancing segmentation robustness across different orientations. However, while MV also achieved high accuracy, it required significantly more computational resources and time due to the necessity of training and combining three separate networks. Although SegNet provided competitive results regarding the JI and DC, it required a substantially higher number of parameters than U-Net. This higher parameter count translates into increased computational load and longer training times, which may be a limiting factor in resource-constrained environments. The challenges observed in segmenting frontal and sagittal views using axial-trained networks highlight the importance of considering multiplanar information during training. The superior performance of MPT underscores its potential as a preferred training strategy for enhancing the accuracy and reliability of CBCT data segmentation.

The choice to focus on a simple network with fewer parameters was driven by the need to create a tool that balances technical performance with practical usability. While more advanced models may offer incremental improvements in segmentation accuracy, they often come at the cost of increased complexity and longer inference times. This approach demonstrates that a well-designed, simple architecture can provide excellent results with significant advantages in speed and user experience, making it highly suitable for real-world clinical applications.

By comparing the Hausdorff distances and the JI and DC between the meshes generated by various segmentation methods, it is evident that the U-Net-MPT-based approach offers superior accuracy. The Hausdorff distance indicates that the U-Net MPT segmentation method yields models with closer alignment to the ground truth, which is critical for applications requiring high precision, such as pre-surgical planning and customized implant design. The workflow's user interface has also proven to be effective in facilitating

the segmentation and visualization process. The intuitive design allows users to generate 3D models easily. The ability to accurately segment and model bones such as the talus, wrist, knee, elbow, and shoulder underscores the versatility of the proposed approach. It allows the generation of an accurate instance segmentation of bones in different anatomical parts without the need to train different neural networks for each specific task.

The method may struggle with highly complex or irregular anatomical structures, such as bones with extensive deformities or fractures. In such cases, the segmentation accuracy could be reduced, potentially requiring manual intervention to correct the segmentation boundaries.

6. Conclusions

In conclusion, these findings suggest that U-Net with MPT is a highly effective and computationally efficient approach for CBCT bone segmentation. While SegNet also performs well, its higher parameter requirements pose a significant drawback.

The integration of accurate segmentation with advanced 3D modeling and a user-friendly interface significantly enhances the practical utility of the entire workflow. The tool not only technically improves the bone segmentation process in CBCT images, but also represents a significant advancement in terms of usability. By reducing the need for user intervention, the process is more accessible and practical for everyday clinical use. This can lead to broader adoption of the technology in clinical settings. The positive results from the preliminary experiments suggest that this bone segmentation and modeling tool can be generalized to various anatomical regions and different types of medical imaging, including veterinary. Future work will involve further validation on larger datasets and additional anatomical regions to fully establish the generalizability of this method. Moreover, additional case studies will be investigated to apply the tool for fracture visualization, bone growth assessment, and preoperative planning for other surgical procedures.

Author Contributions: Conceptualization, E.T.; Methodology, E.T.; Software, E.T.; Data curation, E.T.; Writing—original draft, E.T.; Writing—review & editing, E.T.; Project administration, L.B.; Funding acquisition, L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The informed consent was waived because this study used ex vivo anatomical preparations.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBCT	cone beam computed tomography
MPT	multiplanar training
MV	majority voting
JI	Jaccard index
DC	Dice coefficient

References

1. Bailey, J.; Solan, M.; Moore, E. Cone-beam computed tomography in orthopedics. *Orthop. Trauma* **2022**, *36*, 194–201. [[CrossRef](#)]
2. Jacques, T.; Morel, V.; Dartus, J.; Badr, S.; Demondion, X.; Cotten, A. Impact of introducing extremity cone-beam CT in an emergency radiology department: A population-based study. *Orthop. Traumatol. Surg. Res.* **2021**, *107*, 102834. [[CrossRef](#)] [[PubMed](#)]
3. Grassi, R.; Guerra, E.; Berritto, D. Bone fractures difficult to recognize in emergency: May be cone beam computed tomography (CBCT) the solution? *Radiol. Medica* **2023**, *128*, 1–5. [[CrossRef](#)]

4. Bhat, S.H.; Hareesha, K.S.; Kamath, A.T.; Kudva, A.; Vineetha, R.; Nair, A. A Framework to Enhance the Experience of CBCT Data in Real-Time Using Immersive Virtual Reality: Impacting Dental Pre-Surgical Planning. *IEEE Access* **2024**, *12*, 45442–45455. [[CrossRef](#)]
5. LaLonde, L.; Askar, M.; Paurazas, S. A Novel Diagnostic and Treatment Approach to an Unusual Case of Dens Invaginatus in a Mandibular Lateral Incisor Using CBCT and 3D Printing Technology. *Dent. J.* **2024**, *12*, 107. [[CrossRef](#)]
6. van Eijnatten, M.; van Dijk, R.; Dobbe, J.; Streekstra, G.; Koivisto, J.; Wolff, J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng. Phys.* **2018**, *51*, 6–16. [[CrossRef](#)]
7. Vaitiekunas, M.; Jegelevicius, D.; Sakalauskas, A.; Grybauskas, S. Automatic method for bone segmentation in cone beam computed tomography data set. *Appl. Sci.* **2020**, *10*, 236. [[CrossRef](#)]
8. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci.* **2015**, *9351*, 234–241.
9. Klein, A.; Warszawski, J.; Hillengaß, J.; Maier-Hein, K.H. Automatic bone segmentation in whole-body CT images. *IJCARS* **2019**, *14*, 21–29. [[CrossRef](#)]
10. Lessmann, N.; van Ginneken, B.; de Jong, P.A.; Išgum, I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Med Image Anal.* **2019**, *53*, 142–155. [[CrossRef](#)]
11. Morita, D.; Mazen, S.; Tsujiko, S.; Otake, Y.; Sato, Y.; Numajiri, T. Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net. *Int. J. Oral Maxillofac. Surg.* **2023**, *52*, 787–792. [[CrossRef](#)] [[PubMed](#)]
12. Schnider, E.; Wolleb, J.; Huck, A.; Toranelli, M.; Rauter, G.; Müller-Gerbl, M.; Cattin, P.C. Improved distinct bone segmentation in upper-body CT through multi-resolution networks. *Int. J. Comput. Assist. Radiol. Surg.* **2023**, *18*, 2091–2099. [[CrossRef](#)]
13. Liao, D.; Shi, C.; Wang, L. A complementary integrated Transformer network for hyperspectral image classification. *CAAI Trans. Intell. Technol.* **2023**, *8*, 1288–1307. [[CrossRef](#)]
14. Gheisari, M.; Ebrahimzadeh, F.; Rahimi, M.; Moazzamigodarzi, M.; Liu, Y.; Dutta Pramanik, P.K.; Heravi, M.A.; Mehbodniya, A.; Ghaderzadeh, M.; Feylizadeh, M.R.; et al. Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. *CAAI Trans. Intell. Technol.* **2023**, *8*, 581–606. [[CrossRef](#)]
15. Zhang, Q.; Xiao, J.; Tian, C.; Chun-Wei Lin, J.; Zhang, S. A robust deformed convolutional neural network (CNN) for image denoising. *CAAI Trans. Intell. Technol.* **2023**, *8*, 331–342. [[CrossRef](#)]
16. Lin, X.; Xin, W.; Huang, J.; Jing, Y.; Liu, P.; Han, J.; Ji, J. Accurate mandibular canal segmentation of dental CBCT using a two-stage 3D-UNet based segmentation framework. *BMC Oral Health* **2023**, *23*, 551. [[CrossRef](#)]
17. Zhang, Y.; Qian, K.; Zhu, Z.; Yu, H.; Zhang, B. DBA-UNet: A double U-shaped boundary attention network for maxillary sinus anatomical structure segmentation in CBCT images. *Signal, Image Video Process.* **2023**, *17*, 2251–2257. [[CrossRef](#)]
18. Cireşan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Deep neural networks segment neuronal membranes in electron microscopy images. *Adv. Neural Inf. Process. Syst.* **2012**, *4*, 2843–2851.
19. Vivanti, R.; Ephrat, A.; Joskowicz, L.; Lev-Cohain, N.; Karaaslan, O.A.; Sosna, J. Automatic liver tumor segmentation in follow-up CT scans: Preliminary method and results. *Lect. Notes Comput. Sci.* **2015**, *9467*, 54–61.
20. Shin, H.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med Imaging* **2016**, *35*, 1285–1298. [[CrossRef](#)]
21. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
22. Zhou, X.; Takayama, R.; Wang, S.; Hara, T.; Fujita, H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med Phys.* **2017**, *44*, 5221–5233. [[CrossRef](#)] [[PubMed](#)]
23. Tiribilli, E.; Manetti, L.; Bocchi, L. Single Bone Modeler: Deep learning bone segmentation for cone-beam CT. *IEEE* **2024**, *in press*.
24. Feldkamp, L.; Davis, L.; Kress, J. Practical cone-beam algorithm. *J. Opt. Soc. Am.* **1984**, *1*, 612–619. [[CrossRef](#)]
25. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
26. Lewiner, T.; Lopes, H.; Vieira, A.W.; Tavares, G. Efficient Implementation of Marching Cubes' Cases with Topological Guarantees. *J. Graph. Tools* **2003**, *8*, 1–15. [[CrossRef](#)]
27. Lorensen, W.E.; Cline, H.E. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.* **1987**, *21*, 163–169. [[CrossRef](#)]
28. Denisova, E.; Tiribilli, E.; Manetti, L.; Bocchi, L.; Iadanza, E. Multimodal Biomedical Imaging Platform All-in-One. *Proc. IFMBE Proc.* **2022**, *in press*.
29. Denisova, E.; Tiribilli, E.; Luschi, A.; Francia, P.; Manetti, L.; Bocchi, L.; Iadanza, E. Enabling reliable usability assessment and comparative analysis of medical software: A comprehensive framework for multimodal biomedical imaging platforms. *Health Technol.* **2024**, *14*, 671–682. [[CrossRef](#)]
30. Tiribilli, E.; Denisova, E.; Luschi, A.; Manetti, L.; Bocchi, L.; Iadanza, E. In Proceedings of the IUPESM World Congress on Medical Physics and Biomedical Engineering, Singapore, 12–17 June 2022.
31. Nielsen, J.; Molich, R. Heuristic evaluation of user interfaces. In Proceedings of the CHI90: Conference on Human Factors in Computing, Seattle, WA, USA, 1–5 April 1990; pp. 249–256.
32. Boykov, Y.; Funka-Lea, G. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.* **2006**, *70*, 109–131. [[CrossRef](#)]

33. Tiribilli, E.; Manetti, L.; Bocchi, L.; Iadanza, E. Extremity Bones Segmentation in Cone Beam Computed Tomography, a Novel Approach. *Proc. IFMBE Proc.* **2024**, *93*, 278–284.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.