Automatic Mandible Segmentation Using Convolutional Neural Networks and Feature Sharing

Lucas Mendonça de Morais Cavalcante*
Aura Conci
Leandro A. F. Fernandes
lucasmmc@id.uff.br
aconci@ic.uff.br
laffernandes@ic.uff.br
Universidade Federal Fluminense
Niterói, Rio de Janeiro, Brazil

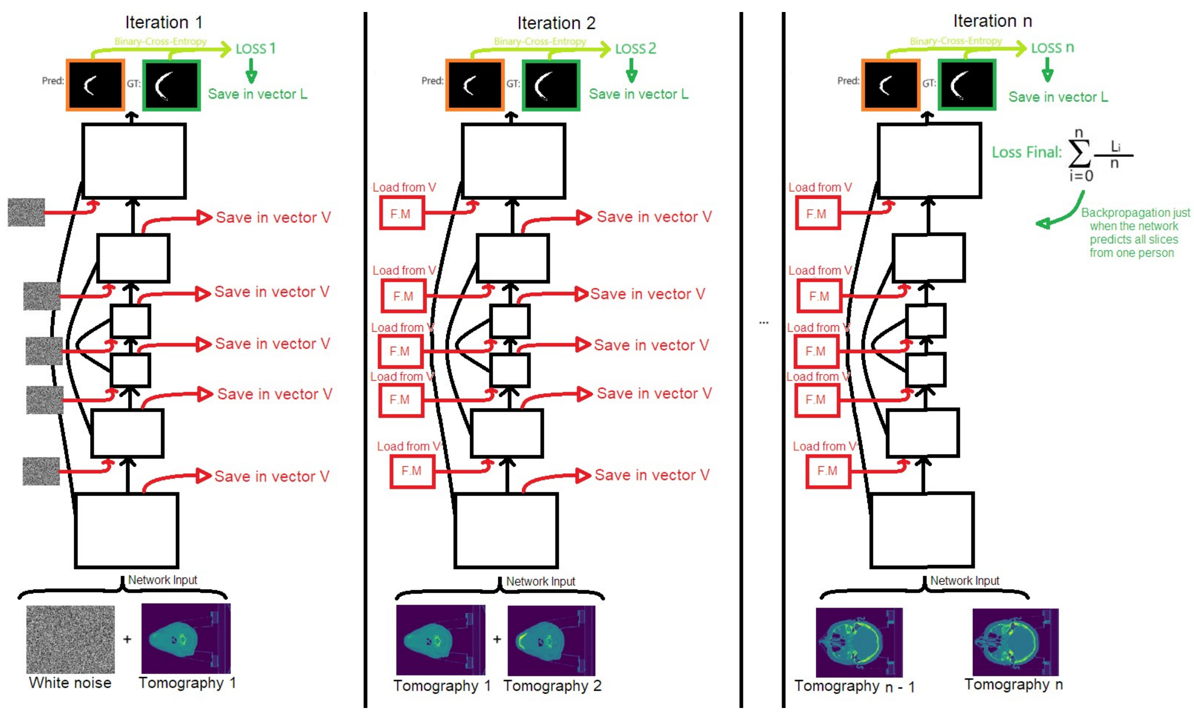

Figure 1: The new UUUUUUUUUUNet Architecture.

Abstract

Mandible segmentation is a crucial step in various dental procedures and is the focus of this work. It uses a new neural network for it. The initial results were promising, achieving 98% accuracy. However, due to class imbalance, these results did not fully reflect reality. The customized architecture is still in the testing phase and has not been fully evaluated, so no conclusive results have been obtained thus far. Nevertheless, it is undergoing extensive validation and demonstrates strong potential to significantly advance automation in dental imaging.

*This work is part of the dissertation of the first author. The second and third authors are respectively my supervisor and co-supervisor.

ACM IMX Workshops, June 3 - 6, 2025.

https://doi.org/10.5753/imxw.2025.7130

CCS Concepts

• **Applied computing** → *Health care information systems.*

Keywords

Mandible; Segmentation; Neural; Networks; U-net; Features; Sharing

How to cite this paper:
Lucas Mendonça de Morais Cavalcante, Aura Conci, and Leandro A. F. Fernandes. 2025. Automatic Mandible Segmentation Using Convolutional Neural Networks and Feature Sharing. In *Proceedings of ACM IMX Workshops, June 3 - 6, 2025*. SBC, Porto Alegre/RS, Brazil,6 pages. https://doi.org/10.5753/imxw.2025.7130

1 Introduction

Automation of medical treatments is already employed in various fields of medicine, including dentistry. One area of interest in both research and industry is the identification of the mandibular integrity, due to its great exposition to accidents. Mandibular observation supports several processes such as clinical planning and prosthesis fabrication. Many treatments begin with a computed tomography (CT) scan of the skull, followed by the segmentation, usually manual, of the area to be analyzed in isolation. From this, several operations become possible, such as 3D reconstruction, computerized planning, analysis, and disease diagnosis. This segmentation is an essential tool for oral and maxillo facial surgeries [7].

The manual segmentation of the mandibular bone requires the presence of a qualified and usually experienced specialist. This area presents several issues that make the process very complex. So manual segmentation takes significant time. Moreover, the quality of segmentation usually relies on the professional's experience, and a considerable margin for errors can appear. Due to a certain degree of subjectivity and inaccuracy in the task, the quality of the outcome is primarily linked to the experience of the professional involved and their commitment to the task [9]. This manual method often consumes professionals' time with mechanical and exhausting work, taking away the opportunity to focus on more valuable and hard tasks. If automation were possible, it would not only speed up this process but also ensure uniformity and consistency in the results. This would lead to fewer discrepancies between different segmentations and greater reliability in disease progression and treatment. Taking these factors into account, there is significant interest in the dental field in automatic mandible segmentation.

To address these limitations and develop a well-structured and efficient pipeline, this work proposes the automatic segmentation of the mandibular region. Initially, we focused exclusively on the main orientation of the CT (the axial view), and on the use of Convolutional Neural Networks (CNN). We firstly used the U-net base architecture [8], it performs well in segmenting the mandibular body but struggles to delineate the *ramus region*, especially the named *condyle* (observe Figure 2). We also designed a customized version based on it (Figure 1 presents it), leveraging feature map sharing between tomographic slices, giving the network a spacial notion, helping in segmenting difficult areas using information from adjacent slices.

2 Literature Review

Automating mandibular segmentation is a broad and extensively studied topic. The diversity of approaches and solutions demonstrates that the problem can be addressed in many ways. However, there is consistency in certain articles.

In the study by Qiu et al. (2021), a combination of recurrent and convolutional neural networks was used, allowing different

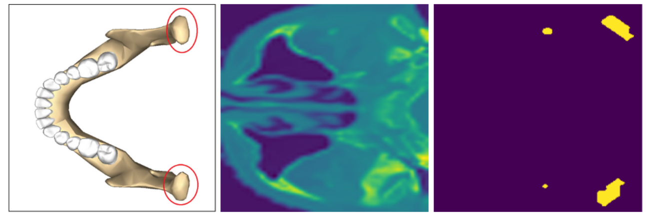

Figure 2: On the left: Mandible region with the condyle circulated in red. [Public domain], via Wikimedia Commons.(https://commons.wikimedia.org/wiki/File: Mandible_close-up_superior.png). In the middle: Axial tomography of the cranium from the PDDCA dataset containing the condyle region. In the right: Condyle Segmentation of the image in the middle. The last two images use cropped images to better show the cited region.

slices of the same scan to share features and useful information for segmentation [6]. They results were evaluated over the three dimensional mandibular area, as the input data consisted of 3D CT scans. The training and tested data included two sets of CT scans, one being a proprietary dataset in which the technique achieved a Dice Similarity Coefficient (DSC) of 0.97, an Average Symmetric Surface Distance (ASSD) of 0.21 mm, and a 95% of the maximum of the measured Hausdorff Distance, named as 95HD, of 2.65 mm [2].

In Pankert et al. (2023) multiple 3D U-Nets were used in a two-step process [5]. Two separated 3D U-Nets were trained using different configurations and loss functions to produce a high-resolution segmentation of the mandible. These networks were used in a cascaded format, with the first generating a low-resolution segmentation defined by a bounding box and the second refining the result based on the detected region of interest. A proprietary dataset was used for training and validation, where the method achieved a DSC of 0.94 after applying the aforementioned bounding box.

Irshad, Talal Bin, et al. (2024) [4] conducted a comparative study of multiple software tools used for mandible segmentation, examining both fully automated and manually assisted approaches. One notable example is Relu [3] (Adicionar site da Relu), which uses a 3D-U-net architecture to segment the mandible region from DICOM files. With a great user interface, the Relu project aims to be effective and user friendly at the same time. Due to the absence of a ground truth, the study compared the results obtained from manual-assisted tools with the only automated method. Relu achieves an average DSC of 0.92, indicating competitive performance. However, the study also points out that, while Relu was among the simplest and most intuitive tools, it produced most variation in results compared to other software. In addition, it offers limited customization options during the segmentation process. Although Relu is behind a paywall, it presents strong automatic segmentation results and an acessible interface, but its restricted availability and lack of flexibility may limit its adoption in more diverse clinical or research settings.

3 Fundamentals

Computed tomography (CT) is an X-ray based technique that allows for clearer and more contrasted visualization of internal organs. It is performed using a machine with multiple X-ray sources that are activated while moving in a circular motion around the body. The detectors, positioned on the opposite side, capture the radiation that is not absorbed by different body tissues. This information is then represented as a grayscale digital image, which can be measured in Hounsfield units (HU) [1].

3.1 Region Of Interest (ROI)

In image segmentation tasks, a large portion of the image is often occupied with background or irrelevant information. This can negatively impact the performance of a model or technique, as it may highlight patterns that should be ignored or de-emphasized, interfering with the decision making process and leading to potential errors and misleading metrics.

To address this issue, we crop the image to include only the region containing the target class. Maximizing the presence of relevant information in the input, helping the model to focus on the features that are truly representative of the class of interest.

3.2 Standard U-net

The original U-Net model was proposed in 2015. This architecture uses an Encoder, which learns to extract and retain only the relevant information and features from the image. The image reconstruction process is performed by a Decoder, which utilizes information from previous layers through skip connections, allowing for the preservation of finer details in the image [8].

4 Materials and Methods

The Medical Image Computing and Computer Assisted Interventions (MICCAI) Society is a professional organization for scientists in the areas of human medical images founding in 2004 https://miccai.org/.

4.1 Database

The most notable and widely available dataset was the Public Domain Database for Computational Anatomy (PDDCA), provided by the MICCAI to be used in the Head and Neck Auto Segmentation Challenge 2015 (promoted by its International Conference held in Munich, Germany, in October 2015). This dataset includes CT scans from 48 patients, with manual segmentation of the mandible and other anatomical structures. The data is organized by patient, with each entry representing a person's exam covering the area from the lower mandible to the top of the skull.

Upon analyzing the CT slices, only those containing the mandible were selected, as the majority did not include it. This selection was necessary because an imbalance in the dataset could bias the model's training toward the predominant class on the initial tests, the absence of the mandible potentially leading to predictions where images are mostly filled with background.

Of these 48 CT scans, 8 did not contain segmented mandibles and were therefore excluded. On the remaining 40, we used one function from the *Scikit-learn library* to randomly divide the samples, with 32 assigned for training and 8 for testing. Throughout all processes, full CT scans were used without applying cur outside the Region of Interest (ROI). The original dataset can be found at the URL: https://www.imagenglab.com/newsite/pddca/, and the pre-processed tomographies at: https://github.com/ColdmaterL/PDDCA_Mandible_pre_processed.

4.2 Pre-processing

In the initial experiments, we applied a windowing configuration, we used a level value of 1000, because bone structures, such as the mandible, are usually represented on this scale in Hounsfield Units. Regarding the window width, we used a value of 200, as it provided a good range of values in our experiments, effectively differentiating the tissues in this part of the body. However, the network achieved better performance when no filtering was applied, using only the raw data from the dataset.

We also applied a ROI on the tomographies in the tomographic images for some experiments. Initially, to assess the feasibility of this approach,we used ROI coordinates derived from the mask data provided by the PDDCA dataset. For each subject, we determined the minimum and maximum pixel coordinates of the target class, and cropped the image accordingly. This ensures that no relevant information is lost. By significantly reducing background content, this strategy provides more informative input to the U-net, potentially enhancing its performance.

4.3 Model Structure

In the experiments, two types of U-net were used, one containing a standard model and the other modifying it with a feature map sharing technique. In this mechanism, we stack feature maps generated by adjacent CT slices. From this strategy, we hope to provide a sense of positioning and refinement of the segmented region. We will refer to this modified U-net as UUUUUUUUUU-net, due to the organization of the model in the training process. All models were trained with the Tensorflow framework.

4.3.1 Evaluating the results. In training and validation, the binary-cross-entropy (BCE) is used for result evaluation. BCE is defined by the formula 1.

$$\mathrm{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{1}$$

where N is the total number of samples, y_i the true value of the i-th sample, and $\hat{y}_i$ is the model's prediction of the i-th sample.

It measures the difference between the probability distribution predicted by the model and the real distribution. Generally used in binary classification problems, it was initially chosen due to the nature of ground truth (GT) segmentations being binary images. The objective was for the network to generate images with a high degree of similarity to the original segmentations, helping to distinguish between pixels belonging to the mandible and background pixels.

A characteristic of binary classifications is that we are interested in classifying each pixel in the image as belonging to only two classes. This is essential for highlighting anatomical structures. U-nets return the probability that each pixel belongs to a class, this probability is between 0 and 1. BCE assigns higher penalties for errors in pixels where the model is very confident and wrong. For instance, if the model predicts that a pixel belongs to the mandible class with a value of 0.99, but the class value is 0. It penalizes less than when the model is uncertain, for example, when it assigns a pixel a value of 0.5, which is not completely wrong, but there is room for improvement.

In mandibular segmentation, BCE is calculated for each pixel in the image and then averaged. This allows us to evaluate the overall performance of the model in segmenting the image as a whole.

4.3.2 UUUUUUUUUU-net. One of the major difficulties encountered in the experiments with the standard U-net was the correct segmentation of the upper regions of the mandible, namely the ramus region and especially the condyle (Figure 2 show them).

Their different shape and location in relation to the upper parts make it difficult for the network to adapt to the delimitation of the mandibular area.

It is possible to observe, by analyzing the adjacent tomographic slices, that the mandible region follows a well-defined structure and a characteristic continuity 4. If the network could incorporate this structural knowledge, it would probably avoid segmenting tissues with intensities similar to the mandible.

The UUUUUUUUUU-net consists of sharing information about the area of interest in adjacent slices, with the stacking of feature maps generated during training. We hope that in the convolution process, the network will learn about the continuity in the bone region, making the segmentation of the mandible in the most superior areas, the ramus of the mandible and the condyle region, more precise and more consistent with reality. The architecture can be seen in Figure 1.

5 Results

No meaningful results were generated with the new architecture proposed here in Figure 1. We are still working on its implementation. All presented results are by using the standard U-net.

5.1 Qualitative Results

Analyzing standard U-net results without applying the ROI, it was possible to observe that the network had great difficulty distinguishing the mandible from the bones and other tissues of the skull region, since all have very similar intensities in these slices. That is the condyle, due to its small size as seen in Figure 2 and 3, is very complex to be segmented. In contrast, there are promising results of the mandible body region for slices of the lower region of the skull. In this region, precise segmentations occur, very close to the GT images, as can be seen in Figure 4, because there are a great contrast among the organ of interest and the tissues on it.

When applying ROI, the results became noticeably more optimistic. By restricting the attention of the network to the regions containing mandible structures, high-quality segmentations were consistently achieved throughout the tomographies, including the challenging ramus region, as illustrated in Figure 4. Even in the least accurate cases, such as the upper portion of the condyle, the model still produced coherent, though faulty, boundary delineations. We believe this technique significantly improved the overall segmentation performance.

Table 1: U-net Results Metrics

	U-net without ROI	U-net with ROI
Accuracy	0.98	0.90
Precision	0.07	0.05
F1-Score	0.11	0.05

5.2 Quantitative Results

In the training process, before using ROI, we utilized accuracy to evaluate the adequacy of the network. However, it was not completely representative of the reality of results. The average accuracy of the test dataset, it is 98% as can be seen in Table 1, which in theory would represent that the model produces segmentations very close to the GT, even in the upper slices of the mandible, where we had major problems (representing a significant part of the image). At same time, the loss value is low throughout the training, showing inadequate learning rate. This indicates that the network does not learn significantly when going through the data.

The training process using ROI was the same as without it. Using the same test dataset we achieved an accuracy of 90%. Although it resulted in lower accuracy, precision, and F1-score as can be seen in 1, the values were more representative of the real output of the network. Most segmentations accurately captured the mandible regions, with only a few tomographies presenting issues, primarily in the upper portion of the ramus region.

It is important to note that the images in Figures 3 and 4 do not represent the same slices. They are just ilustrative images of the segmentation shape produced by the network in both scenarios.

6 Conclusion and Future Works

Automatic Mandibular Segmentation is an area of great interest for the dentintry field, with applications in both research and commercial areas. An efficient tool could reduce costs, increase productivity, improve the consistency of results and alleviate the workload of professionals in this mechanical and exhausting task.

In this article we show how U-net can be used. This is one of the most widely used models in mandibular segmentation, and a new architecture that incorporates the sharing of feature maps between adjacent slices is proposed, as well. This approach seeks to solve a recurring problem in the standard implementation: incorrect segmentation of the upper part of the mandible.

Although initial results were limited, probably due to class imbalance in the ground truth images, where background regions dominate, the experiments provided valuable insights. Traditional evaluation metrics such as accuracy proved to be insufficient to fully capture segmentation performance in this context. However, the introduction of ROI techniques helped mitigate these issues, revealing a clearer potential for improvement and setting a promising direction for future work.

6.1 Future Works

At this moment three aspects were visored to improve this study:

6.1.1 Use of Different Tomographic Perspectives: In the current study, only axial CT scans were used. However, the PDDCA dataset

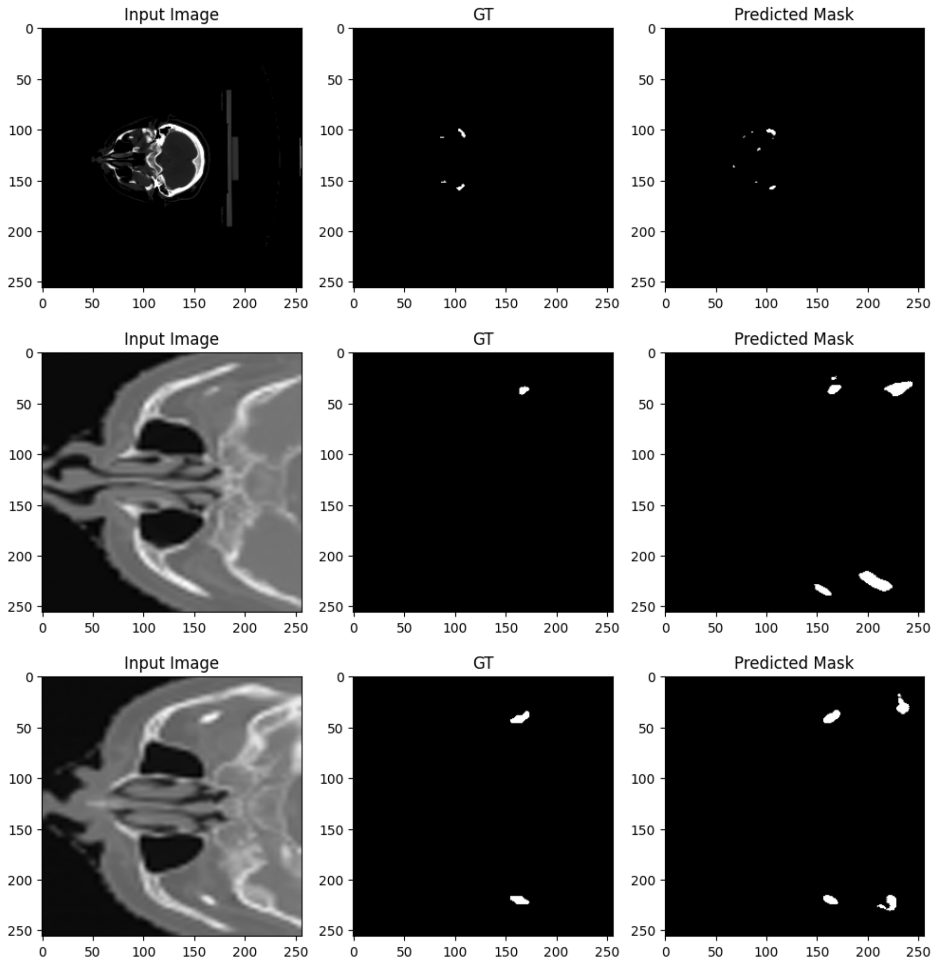

Figure 3: Examples of poor mandible segmentation. Different slices of different patients. First Image: Before applying ROI. Second and Third Image: After Applying ROI.

includes sagittal and coronal slices. Incorporating these views into the pipeline can provide complementary information, especially in anatomical regions where the mandible presents a more complex shape. Combining these sllices can improve segmentation in critical areas and reduce errors.

6.1.2 Most Suitable Loss Functions: The loss BCE, which we use in U-nets, is not suitable for images where there is a large class imbalance. The use of loss functions such as Dice Loss, Focal Loss, Weighted Cross-Entropy Loss, and Tversky Loss are more recommended. They are more robust in imbalance scenarios and prioritize the region of interest, better reflecting the quality of the segmentation, without the need for a qualitative evaluation.

6.1.3 Automatic ROI delimitation: The initial ROI experiments were promising; however, we used predefined coordinates to perform the cropping. The next step is to automate this process employing a technique that, given a tomography, can determine the optimal coordinates for the region of interest. We plan to leverage windowing information to achieve this.

Acknowledgments

I would like to thank UFF professor Flavio Seixas and the Rebone 3D study group for introducing me to the field of mandibular segmentation and for all the support during the writing of this article. Their guidance and collaboration made a huge difference along the way.

A.C. thanks to CNPq and FAPERJ Brazilian Agencies for the support on this research.

References

[1] Edson Amaro Júnior and Helio Yamashita. 2001. Aspectos básicos de tomografia computadorizada e ressonância magnética. *Brazilian Journal of Psychiatry* 23 (2001), 2–3.
[2] Aura Conci, Stephenson S.L. Galvão, Giomar O. Sequeiros, Débora C.M. Saade, and Trueman MacHenry. 2015. A new measure for comparing biomedical regions of interest in segmentation of digital images. *Discrete Applied Mathematics* 197 (2015), 103–113. doi:10.1016/j.dam.2015.06.002 Distance Geometry and Applications.
[3] Robert R. Ileșan, Michel Beyer, Christoph Kunz, and Florian M. Thieringer. [n. d.]. Comparison of Artificial Intelligence-Based Applications for Mandible Segmentation: From Established Platforms to In-House-Developed Software. *Bioengineering* 10, 5 ([n. d.]). https://www.mdpi.com/2306-5354/10/5/604

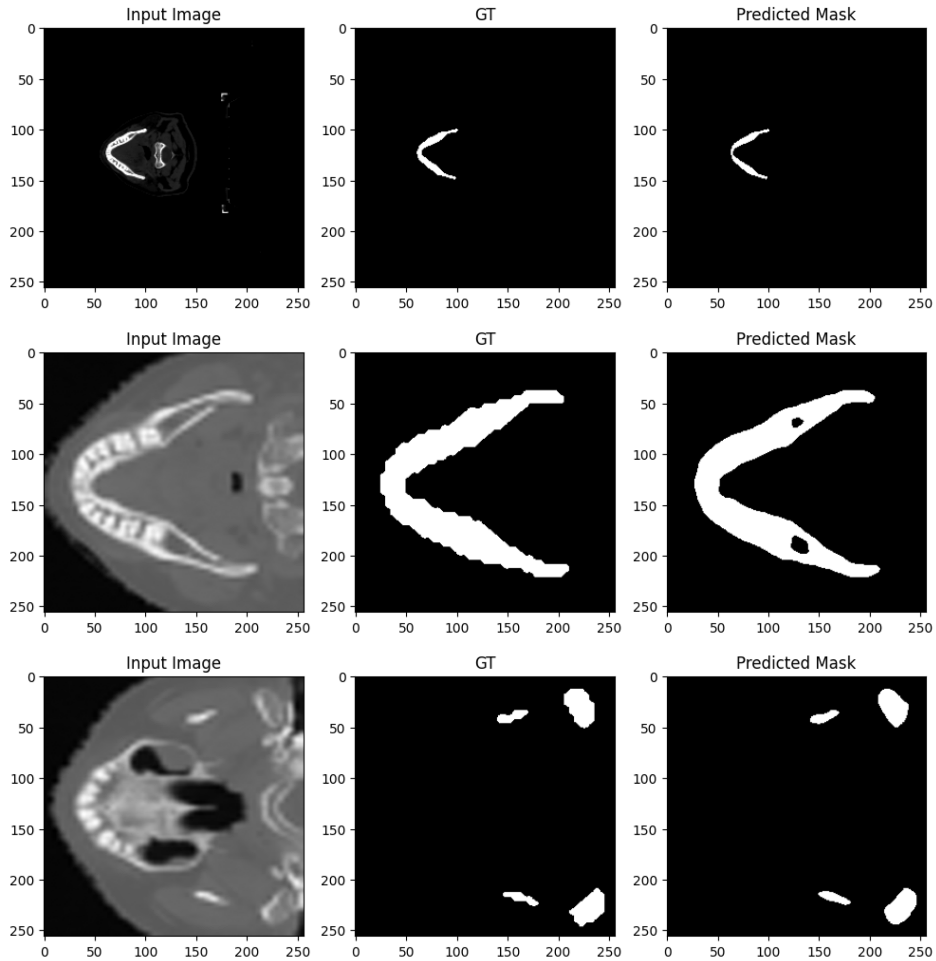

Figure 4: Examples of good mandible segmentation. Different slices of different patients. First Image: Before applying ROI. Second and Third Image: After Applying ROI.

[4] Talal Bin Irshad, Giulia Pascoletti, Francesco Bianconi, and Elisabetta M Zanetti. 2024. Mandibular bone segmentation from CT scans: Quantitative and qualitative comparison among software. *Dental Materials* 40, 8 (2024), e11–e22.

[5] Tobias Pankert, Hyun Lee, Florian Peters, Frank Hölzle, Ali Modabber, and Stefan Raith. 2023. Mandible segmentation from CT data for virtual surgical planning using an augmented two-stepped convolutional neural network. *International Journal of Computer Assisted Radiology and Surgery* 18, 8 (2023), 1479–1488.

[6] Bingjiang Qiu, Jiapan Guo, Joep Kraeima, Haye Hendrik Glas, Weichuan Zhang, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter MA van Ooijen. 2021. Recurrent convolutional neural networks for 3D mandible segmentation in computed tomography. *Journal of personalized medicine* (2021).

[7] Bingjiang Qiu, Hylke van der Wel, Joep Kraeima, Haye Hendrik Glas, Jiapan Guo, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter MA van Ooijen. 2021. Automatic segmentation of mandible from conventional methods to deep learning—a review. *Journal of personalized medicine* 11, 7 (2021), 629.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.* Springer, 234–241.

[9] Muhammad Usman, Azka Rehman, Amal Muhammad Saleem, Rabeea Jawaid, Shi-Sub Byon, Sung-Hyun Kim, Byoung-Dai Lee, Min-Suk Heo, and Yeong-Gil Shin. 2022. Dual-stage deeply supervised attention-based convolutional neural networks for mandibular canal segmentation in CBCT scans. *Sensors* 22, 24 (2022), 9877.