



# An Explainable Deep Learning Framework for Mandibular Canal Segmentation from Cone Beam Computed Tomography Volumes

Konstantinos Barzas<sup>1</sup>, Shereen Fouad<sup>1</sup>(✉)(ID), Gainer Jasa<sup>2</sup>(ID),  
and Gabriel Landini<sup>3</sup>(ID)

<sup>1</sup> School of Computer Science and Digital Technologies, Aston University,  
Birmingham, UK

s.fouad@aston.ac.uk

<sup>2</sup> Division of Oral Radiology, Faculty of Dentistry, Republic University,  
Montevideo, Uruguay

<sup>3</sup> Oral Pathology Unit, Dentistry, School of Health Sciences, University of  
Birmingham, Birmingham, UK

**Abstract.** Cone Beam Computed Tomography (CBCT) is an indispensable imaging modality in oral radiology, offering comprehensive dental anatomical information. Accurate detection of the mandibular canal (MC), a crucial anatomical structure in the lower jaw, within CBCT volumes is essential to support clinical dentistry workflows, including diagnosis, preoperative treatment planning, and postoperative evaluation. In this study, we present a deep learning-based (DL) approach for MC segmentation using 3D U-Net and 3D Attention U-Net networks. We collected a unique dataset of CBCT scans from 20 anonymous hemisected mandibular bones, which were further processed for analysis. The samples were scanned using a CBCT scanner after inserting a wire through the whole length of the MC to identify its location in space (as a gold standard). Our experimental results demonstrate that the 3D Attention U-Net outperforms the standard 3D U-Net in detecting the MC's location, with Dice similarity score, Precision, and Recall values of 0.65, 0.75, and 0.60, respectively. Unlike current DL-enabled methods for MC segmentation, which face deployment and trust challenges due to their “black-box” nature, our approach incorporates a post-hoc visual explainability feature through the Grad-CAM++ (Gradient-weighted Class Activation Mapping) algorithm. This tool highlights important regions within the CBCT volumes that influence the model's predictions, providing valuable insights into the segmentation process, and bridging the gap between cutting-edge DL technology and clinical practice.

**Keywords:** Dental Cone Beam Computed Tomography · mandibular canal segmentation · U-Net deep learning model · explainable artificial intelligence · Grad-CAM

---

S. Fouad—Equal Contribution.

© The Author(s) 2025

M. S. Bansal et al. (Eds.): ICCABS 2023, LNBI 14548, pp. 1–13, 2025.

[https://doi.org/10.1007/978-3-031-82768-6\\_1](https://doi.org/10.1007/978-3-031-82768-6_1)

# 1 Introduction and Background

The mandible is the largest and strongest bone in the human orofacial region, giving hard tissue to the lower jaw and part of the mouth. From an anatomical perspective, the mandibular canal (MC) runs along part of the mandible, containing the inferior alveolar nerve and blood vessels. In dental surgery, the accurate detection of MC is crucial to a wide range of dental procedures including dental extractions, implant placement, thirds molar surgeries and jaw alignment surgery to avoid damage to the inferior alveolar nerve and vessels [2]. Advancements in imaging technologies, like Cone Beam Computed Tomography (CBCT), have enhanced the ability to precisely identify the MC. CBCT is a three-dimensional (3D) imaging modality that captures images from various angles with a cone-shaped beam to reconstruct volumetric information as a series of axial images. MC segmentation of volumetric CBCT images is a challenging task due to the complexity of dental anatomical structures that feature variability in shape and position. Segmenting the MC in CBCT images manually is a time-intensive process and susceptible to both intra- and inter-observer variability. Earlier segmentation approaches have used traditional segmentation methods, such as image thresholding, mathematical morphology and statistical approaches [1, 11]. However, those methods suffered often resulted in low segmentation accuracy, long processing time and poor reliability.

More recently, Deep Learning-based (DL) (a subset of machine learning) has shown remarkable performance in segmenting MC from CBCT volumes. In the context of image segmentation, machine learning-based algorithms aim at labelling each image pixel with a specific category [7, 8]. Many attempts to perform MC segmentation in CBCT using DL have primarily employed 3D convolutional neural networks (CNN). For example, Jaskari et al. [9] demonstrated that 3D CNN models produced better MC segmentation quality than statistical shape models, but the segmentation results were able to achieve a Dice similarity score of just 0.57. Kwak et al. [12] used 2D U-Net network [6], 2D SegNet and 3D U-Net [5], to automatically localise the MC in 2D and 3D CBCT images. A thresholding-based tooth segmentation technique was applied initially to remove non-mandibular bone areas from the scans. 3D U-Net outperformed the 2D approach, however, it was unable to detect the canal when the surrounding cortical bone layers were not clearly defined and the pre-processing thresholding step was difficult to initialise. Lahoud et al. [13] proposed a two step segmentation framework using 3D CNN networks to firstly perform voxel-wise segmentations to adjust for variations in MC shape and width and then train another 3D CNN model to produce a full-resolution segmentation output. They achieved a Dice similarity score of 0.77 using a dataset of 235 CBCT scans. Jeoun et al. [10] introduced the Canal-Net approach, based on the 3D U-Net architecture, enhanced with bidirectional convolutional long short-term memory units within a multi-task learning framework. This approach outperformed 2D U-Net, SegNet and 3D U-Net, but has large memory requirements. Recently, a two-stage 3D U-Net method for MC segmentation was proposed in [14] which yielded a Dice

similarity score of 0.95 on a different dataset that is significantly larger than our studied images.

Despite remarkable results achieved by DL-methods in MC segmentation tasks, these methods suffer from the “black-box” nature, i.e. while they can produce accurate results, it is often challenging to understand and interpret why the model arrived to a particular decision. This lack of transparency can be problematic in clinical settings where trust and interpretability are paramount. This lack of explainability, common in DL-based segmentation tools, hinders the their adoption and use in critical clinical settings. Additionally, the “black-box” nature of DL-based models can easily pose challenges in meeting the requirements of regulatory bodies. Recently, the concept of Explainable AI (XAI) has emerged, introducing a range of techniques that aim to strike a balance between explainability and robust detection and prediction performance [18]. Post-hoc visual XAI in medical image segmentation provides visual explanations for the decisions made by DL models after they have performed the image segmentation. This is achieved by highlighting image regions that exert the most significant influence on the model’s decision. This is a crucial step for understanding and validating the segmentation results and ensuring that its trustworthy and accurate. Various post-hoc visual XAI methods have been proposed in the literature, however, the application of XAI has not been well explored in the context of MC segmentation in 3D CBCT.

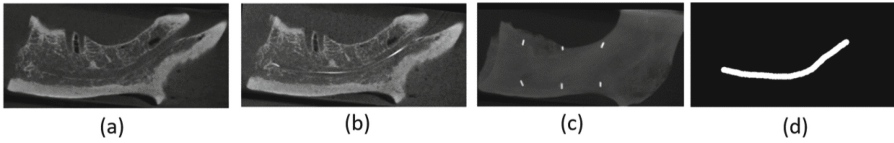
To close this gap, we propose an explainable DL framework for MC segmentation from CBCT volumes using U-Net architectures and Grad-CAM++ (Gradient-weighted Class Activation Mapping) algorithms [4]. In particular, we study the performance of 3D U-Net [5] and 3D Attention U-Net [15] networks (a modified version of the U-Net). 3D U-Net has shown remarkable results in segmenting volumetric biomedical images [10, 12, 14], however, 3D Attention U-Net has not been fully explored in the context of detecting MC in 3D CBCT images. The ‘attention’ mechanism seems particularly effective when handling medical images characterized by high levels of noise or when the region of interest occupies only a small portion of the complete image, which applies to our CBCT images. Furthermore, the proposed framework addresses explainability by incorporating Grad-CAM++ [4] (an extended version of the original Grad-Cam [17]) into the segmentation pipeline, thereby providing clinicians with an additional and deeper understanding of the decision-making process.

## 2 Dataset Description and Preprocessing

The sample set consisted of 20 anonymous dry hemisected mandibular bones, obtained from the Anatomy Museum collection at the Faculty of Dentistry, Universidad de la República, Uruguay. The ethical approvals for collecting and using the data were obtained prior to conducting the study. To create gold standard images, each half mandible was scanned with a wire inside the MC to precisely locate it. The wire was then removed without moving the sample, and the half mandibles were scanned using the same field of view and the same exposure

time, but at varying tube voltages and currents, using a combination of kilovoltage (kV) and milliamperage (mA) values. The bones were scanned, submerged in water to simulate soft tissue absorption and positioned so that the MC was parallel to the axial plane. Data volumes captured with and without the inserted wire were spatially registered (see Fig. 1). Digital files were exported in the digital imaging and communication in medicine (DICOM) format for further analysis with Python software. Notably, the dimensions of the CBCT voxels varied within the range of (543, 543, 80) and (543, 543, 190) across the  $X$ ,  $Y$ , and  $Z$  axes respectively, with the associated Hounsfield units spanning from  $-1000$  to  $2000$ .

The ground truth segmentation masks of the MC were hand traced guided by the inserted wire (see Fig. 1) using Slicer imaging software (version 5.2.2). Volumes were cropped to a resolution of (120, 240, 512), denoting the designated Region of Interest (ROI). Thereafter, a 3D spherical paint brush instrument with a diameter of 2% of the ROI was used to mark the location of the wire and thus create the binary mask (ground truth).



**Fig. 1.** (a) mandibular canal without wire inserted, and (b) with wire. (c) original image after applying maximum projection, and (d) binary mask image

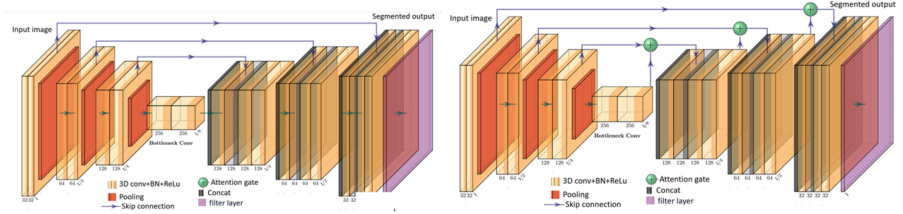
Data augmentation is a key data pre-processing technique that is commonly used to improve the robustness and generalization of DL models, particularly where annotated training data is small in size, which applies to this study. It involves applying various geometric transformations to the original image to create additional training data, including random rotation, scaling, and flipping. The addition of Gaussian noise to the original images also helps the model to become more robust to real-life image artifacts. In the context of rotational transformations, an initial large angle was selected, followed by gradual reductions until the model exhibited convergence, which was assessed by monitoring the decline of the Dice loss over increasing training epochs. The model was trained for a total of 100 epochs and where divergence was noted, the angle of rotation was decreased incrementally, specifically by 5% with each divergence occurrence. Subsequently, once convergence was achieved, a further reduction of the angle by 30% was implemented to establish the final range for that particular parameter. This strategy was also applied for the range of the added random Gaussian noise. After experimentation, the values of mean and standard deviation of Gaussian noise are 0 and 0.1, respectively. While the angle range for the rotational transformation is 0.2 radians in  $x$ ,  $y$ , and  $z$  coordinates. The data augmentation was performed by the MONAI library transforms [3].

### 3 The Segmentation Model

In this study, we propose a DL-based approach for MC segmentation. We investigated the performance of 3D U-Net [5] and 3D Attention U-Net [15] networks. Both methods have demonstrated exceptional performance in similar semantic segmentation tasks [10, 12, 14].

#### 3.1 3D U-Net Architecture

3D U-Net [5] is an adapted version of the conventional 2D U-Net [16] configured to process 3D images. It preserves the core structure of the initial 2D U-Net model, but it substitutes all 2D operations with their 3D counterparts, such as 3D convolutions, 3D max pooling, and 3D up-convolutions, resulting in a segmented image in three dimensions. A graphical illustration of the applied 3D U-Net model is presented in Fig. 2.



**Fig. 2.** 3D U-Net (left) and 3D Attention U-Net architectures (right)

The 3D U-Net architecture contains two paths, the contracting path (or encoder) captures the high-level semantic features of the image, while the expanding path (or decoder) restores the high-level semantic feature map back to the original image resolution. The layers in the encoder part are skip connected and concatenated with layers in the decoder part (highlighted in Fig. 2). This enables 3D U-Nets to utilize the finely-detailed information acquired during the encoder phase to generate an image within the decoder phase.

In the contracting path, the network starts with a 3D convolutional layer featuring 32 filters, followed by instance normalization and dropout layers and a ReLU activation function to introduce non-linearity. This phase follows a consistent pattern of convolution layers, doubling the number of filters at each subsequent stage, ultimately reaching 256 filters. This incremental approach enhances feature representation while progressively reducing the output spatial dimensions, resulting in a systematic down-sampling. In the expanding path, the network uses up-convolution layers to gradually increase the spatial dimensions of the feature maps. It starts with a layer containing 128 filters and incorporates skip connections from corresponding layers in the contracting phase. Feature

channels are halved at each step, aiming to reduce feature channels during up-sampling. This process includes two 3D convolutions followed by a ReLU activation function to refine features. Finally, a single-filter layer is used to convert feature vectors into the segmented output.

### 3.2 3D Attention U-Net Architecture

Skip connection in original 3D U-Net combines spatial information from the contracting path with the expanding path to retain good spatial information. But this process brings along some irrelevant features from the initial layers. Attention in U-Net is a recent extension of the basic U-Net method, and it aims at highlighting only important/relevant parts of the image while ignoring unnecessary areas during training. The 3D Attention U-Net architecture, illustrated in Fig. 2, utilises an attention mechanism by adding attention gates at each level of the expansive path. These gates re-calibrate the feature channels by taking input feature maps from both the contracting and expanding path and merges them to generate a set of attention coefficients which are then used to adjust the feature maps from the contracting path before they are concatenated with the corresponding feature maps in the expanding path. The resultant attention-gated feature maps are subsequently passed through up-convolution layers and other successive layers in the expanding path, mirroring the original 3D U-Net structure. 3D Attention U-Net also adopts the ReLU activation function, and it maintains the use of instance normalization and dropout layers to reduce over-fitting and ensuring a stable training process.

## 4 Post-hoc Visual Explainability - Grad-CAM++

Post-hoc visual explainability refers to the process of providing visual explanation and interpretation of the decisions made by the DL models [18]. It helps DL users understand why a model has generated a specific result, which seems particularly important in medical imaging applications, where trust and interpretability are crucial.

Gradient-weighted Class Activation Mapping (Grad-CAM) [17] is a XAI technique for generating visual explanations of decisions made by CNNs. It generates a ‘heatmap’ that highlights the regions of the input image most responsible for a particular output. The heatmap is constructed by leveraging the gradient information from the last convolutional layer in the network, which is activated for different channels with respect to the class. Specifically, for each feature map  $A_{ij}^k$  at this convolutional layer, an importance weight  $\alpha_k^c$  is computed with respect to the  $k$ th feature map and class  $c$ . This is done through global-average-pooling over pixel location  $(i, j)$  for the gradient  $\frac{\partial y^c}{\partial A_{ij}^k}$  of respective classification score  $y^c$  of class  $c$ . Mathematically, the weights can be estimated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where  $Z$  is the number of pixels in the activation map. Once these importance weights are obtained for each feature map, the next step is to calculate the heatmap, denoted here as  $L^c$ . This is done by taking a weighted sum of the feature maps and applying a ReLU function to keep only the positive contributions:  $L^c = \text{ReLU}(\sum_k \alpha_k^c A^k)$ . The resulting heatmap  $L^c$  therefore shows the areas of the input image that were most influential in producing the class  $c$ , thus providing a visualization of the model's focus and decision-making process.

Grad-CAM++ [4] is an extension of the original Grad-CAM technique, which uses second order gradients. It offers a more fine-grained explanation of model decisions by accounting for the importance of individual pixels within the feature maps. On the other hand, the pixel gradients that have no impact on the prediction will be scaled down. In particular, while Grad-CAM computes a global importance weight  $\alpha_k^c$  for each feature map  $A_k$  through average-pooling, Grad-CAM++ goes a step further and calculates the following weight

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{\partial (A_{ij}^k)^2}}{2 \frac{\partial^2 y^c}{\partial (A_{ij}^k)^2} + \sum_a \sum_b A_{ab}^k \frac{\partial^3 y^c}{\partial (A_{ij}^k)^3}}$$

where  $\alpha_{ij}^{kc}$  is the value of  $\alpha$  at pixel location  $(i, j)$  for the  $k$ -th feature map corresponding to the output class  $c$ . Here,  $(a, b)$  are iterators over the same activation map  $A_k$  and are used to avoid confusion. The heatmap  $L^c$  can be expressed as follows:  $L^c = \sum_i \sum_j \alpha_{ij}^{kc} \cdot \text{ReLU}(\frac{\partial y^c}{\partial A_{ij}^k})$ . The equation considers both the positive and negative contributions of each pixel in determining the final output. This allows Grad-CAM++ to produce more nuanced heatmaps compared to its predecessor Grad-CAM.

## 5 Model Training and Hyperparameters Fine-Tuning

The dataset was partitioned into training, validation, and testing subsets, adhering to a ratio of 60:20:20, respectively. The Hyperparameters of the U-Net model were fine-tuned utilizing the three parameters,

(a) **Dice Loss**, and it is estimated as,  $\frac{2 \times \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + g_i^2}$ . Where  $p_i$  and  $g_i$  represent pairs of corresponding pixel values of predicted binary segmentation volume  $P$  and ground truth binary volume  $G$ ,

(b) **Mean Intersection over Union (mIoU)**, and it is estimated as,  $\frac{1}{C} \sum_c \text{IoU}_c$ , where  $C$  is the number of classes,  $\text{IoU}_c = \frac{TP_c}{TP_c + FP_c + FN_c}$ , and  $TP_c$ ,  $FP_c$ , and  $FN_c$  denotes true positives, false positives, and false negatives for class  $c$ , respectively.

(c) **Generalized Dice Focal Loss (GDFL)**, and it is estimated as

$-\frac{\sum_{i=1}^N 2 \cdot TP_i \cdot (1-p_i)^\gamma}{\sum_{i=1}^N TP_i + \sum_{i=1}^N (1-p_i)^\gamma \cdot (TP_i + FN_i) + \epsilon}$ , where  $N$  is the number of classes,  $p_i$  is the predicted probability for class  $i$ ,  $TP_i$  is the True Positives,  $FN_i$  is the False Negatives,  $\gamma$  is the Focal Loss hyperparameter that modulates the focusing effect, and  $\epsilon$  is a constant value.

As shown in Table 1, a range of Hyperparameters have been adjusted to optimize both the computational efficiency and the predictive performance of the two studied models. A random search was conducted with these parameters; the search was run for 60 iterations for 3D U-Net and 20 for 3D Attention U-Net. Both models were trained with patches from randomly rotated and translated CBCT volumes using the loss objectives.

**Table 1.** Hyperparameters to be tuned

Hyperparameters	Model	
	<b>3D U-Net</b>	<b>3D Attention U-Net</b>
Loss function	Dice Loss, GDFL, Mean IoU	Dice Loss, GDFL
Batch Size	1, 2, 3, 4	1, 2
Channels	[8, 16, 32, 64], [16, 32, 64, 128], [32, 64, 128, 256], [64, 128, 256, 512]	[8, 16, 32, 64], [16, 32, 64, 128]
Dropout	0, 0.2, 0.3, 0.4, 0.6	0, 0.1, 0.2
Learning Rate	3e-3, 5e-4, 5e-5	3e-3, 5e-4, 2e-5
Early Stopping	25	25
Max Epochs	150	150
Val Interval	5	5

## 6 Results

### 6.1 Optimization Results

Figure 3 reports the hyperparameter optimization results from the random search for the 3D U-Net and 3D Attention U-Net models. For the 3D U-Net model, the optimal values for dropout rate, channel sequence and batch size were identified as 0, [32, 64, 128, 256], and 1, respectively. However, further investigation of the learning rates along with loss functions, revealed that the learning rates of  $2 \times 10^{-4}$ , along with GDFL as loss functions, yielded the best results on the validation set. For the 3D Attention U-Net model, the optimal settings for the loss function, learning rate, dropout rate, and batch size, were GDFL, 0.003, 0.1, and 1, respectively. Further investigation to determine the optimal values for the channel sequence (between [8,16,32,64] and [16,32,64,128]) revealed that the lower number channel sequence converged faster, whereas the higher channel sequence converged slower but provided slightly more accurate results on the evaluation set. Eventually, the lower sequence, [8,16,32,64], was selected as the optimal value as it required less number of trainable parameters, making the model more accessible for practical use.



## 6.2 Evaluation Metrics

To evaluate the optimal models generated through the random search the following metrics were used:

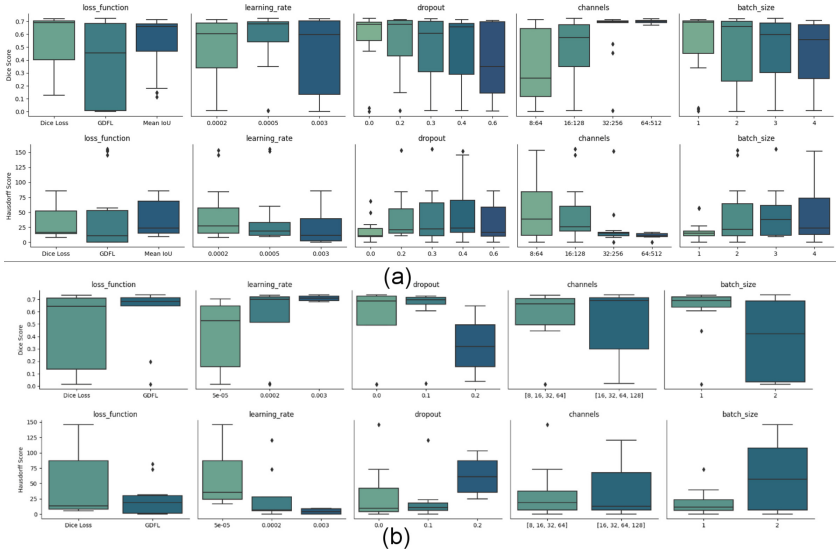
(a) **Dice similarity score**, measures the similarity between two image samples, and it is calculated as:  $\frac{2 \times TP}{2 \times TP + FP + FN}$ ,

(b) **Precision**, measures the proportion of true positive predictions among all the positive predictions, and it is calculated as:  $\frac{TP}{TP + FP}$

(c) **Recall**, measures the model's ability to identify all relevant instances, and it is calculated as:  $\frac{TP}{TP + FN}$

(d) **Hausdorff Distance**, measures the maximum perpendicular distance between the automatic and manual segmentation, and it is defined as:

$\max(\max_{a \in A}(\min_{b \in B} d(a, b)), \max_{b \in B}(\min_{a \in A} d(a, b)))$ , where  $d(\dots)$  represents Euclidean distance between two sets of points denoted as  $a$  and  $b$ .



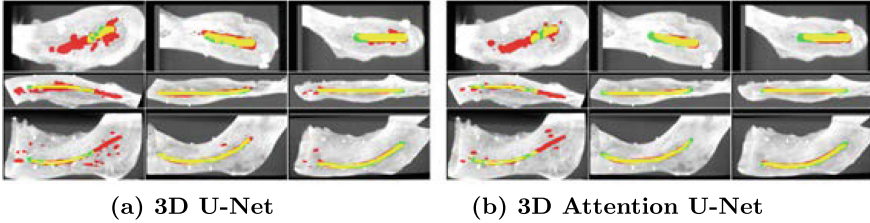
**Fig. 3.** Hyperparameter results from the random search of 3D U-Net (a) and 3D Attention U-Net (b). Each box contains the first and third quartile of data, and medians are indicated as lines inside each box.

## 6.3 Segmentation Results Obtained from the Optimised Models

Segmentation quantitative results obtained from the optimised models are shown in Table 2 and qualitative results are presented in Fig. 4. The 3D Attention U-Net outperformed the standard 3D U-Net, specially in handling high-noise scans and in detecting the central portion of the canal. On the other hand, the 3D U-Net demonstrated better performance in identifying the extremes of the canal, as denoted by a more pronounced presence of red markers at the canal ends.

**Table 2.** Segmentation results obtained from the optimised U-net and Attention U-Net models.

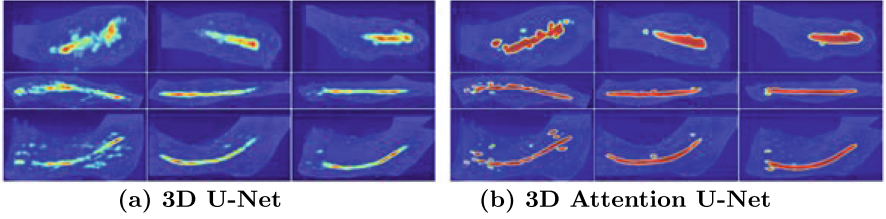
Dataset	Model Name	Dice similarity score	Hausdorff	Precision	Recall
Train	3D U-Net	0.76	15.52	0.81	0.72
	3D Attention U-Net	<b>0.79</b>	<b>6.06</b>	<b>0.82</b>	<b>0.76</b>
Validation	3D U-Net	<b>0.72</b>	16.67	0.77	<b>0.67</b>
	3D Attention U-Net	0.72	<b>15.81</b>	<b>0.79</b>	0.66
Test	3D U-Net	0.62	47.54	0.74	0.57
	3D Attention U-Net	<b>0.65</b>	<b>35.04</b>	<b>0.75</b>	<b>0.60</b>

**Fig. 4.** The segmentation results on Coronal, Axial and Sagittal view of three example CBCT test scans. Green: manual segmentation, red: automatic segmentation, yellow: overlap between automatic and manual segmentation. (Color figure online)

#### 6.4 Post-hoc Visual Explanation Results via Grad-CAM++

After the segmentation output of 3D U-Net and 3D Attention U-Net, was obtained, the Grad-CAM++ method was applied. The M3d-CAM library was used to streamline the process of visualizing attention maps utilizing the Grad-CAM++. The library provides a simple method for incorporating the required components into the pre-existing PyTorch model. Contrary to the common practice of focusing on the last convolutional layer for attention visualization, exploratory analyses revealed that the first convolutional layer provided more insightful attention maps for both 3D Attention U-Net and 3D U-Net. This notable divergence from standard practices offers an insight into the networks' focus during the image segmentation process. The M3d-CAM library streamlined this exploratory process, enabling to load the best-performing models post-training and swiftly visualize attention at various layers. Based on this, the first convolutional layer was ultimately selected for in-depth attention map analyses. Features in this layer contained some high-level semantics and simultaneously should preserve some spatial information.

The output of the Grad-CAM++ for 3D U-Net and 3D Attention U-Net is shown in Fig. 5. The 3D Attention U-Net heatmap showed better region localizing ability when compared to the 3D U-Net model, which was acting as indicator of the model's confidence rather than providing a coherent heatmap representation. The Grad-CAM++ heatmap highlights the features (ROI) that the model considered before making the segmentation decision, enhancing the transparency



**Fig. 5.** Grad-CAM++ output/heatmap on Coronal, Axial and Sagittal view of three example CBCT test scans.

and trustworthiness of the segmentation model. This could help clinicians to verify and validate the model’s decisions, ultimately improving trust and deployment of DL-based techniques in dental surgical procedures.

## 7 Conclusion

Mandibular canal (MC) segmentation is a challenging task in maxillofacial radiology. This study introduces an explainable framework for segmentation of CBCT volumes using 3D U-Net and 3D Attention U-Net models, complemented by Grad-CAM++ visualization. We analysed a set of 20 CBCT scans, followed by a series of data preprocessing steps and optimization procedures to identify the optimal training parameters for MC segmentation. The results revealed that the 3D Attention U-Net system outperformed the standard 3D U-Net in terms of both segmentation quality and post-hoc visualization. The segmentation results were relatively lower than those of some previously DL-based methods, possibly due to the limited number of training scans available. However, the primary focus here centred in improving the explainability of MC automatic segmentation over the perceived “black box” approach of other DL segmentations. The post-hoc heatmaps using Grad-CAM++ highlighted significant regions in the images that contributed to the model’s decision therefore enhancing explainability and trust, thereby facilitating the deployment of DL-enabled segmentation tools within clinical imaging workflows.

## References

1. Abdolali, F., Zoroofi, R.A., Abdolali, M., Yokota, F., Otake, Y., Sato, Y.: Automatic segmentation of mandibular canal in cone beam ct images using conditional statistical shape model and fast marching. *Int. J. Comput. Assist. Radiol. Surg.* **12**, 581–593 (2017)
2. Batstone, M.D., Scott, B., Lowe, D., Rogers, S.N.: Marginal mandibular nerve injury during neck dissection and its impact on patient perception of appearance. *Head Neck: J. Sci. Special. Head Neck* **31**(5), 673–678 (2009)
3. Cardoso, M.J., et al.: Monai: An open-source framework for deep learning in healthcare. *arXiv preprint [arXiv:2211.02701](https://arxiv.org/abs/2211.02701)* (2022)

4. Chattopadhyay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications Of Computer Vision (WACV), pp. 839–847. IEEE (2018)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016)
6. Falk, T., et al.: U-net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**(1), 67–70 (2019)
7. Fouad, S., Landini, G., Robinson, M., Song, T.H., Mehanna, H.: Human papilloma virus detection in oropharyngeal carcinomas with in situ hybridisation using hand crafted morphological features and deep central attention residual networks. *Comput. Med. Imag. Graph.* **88**, 101853 (2021)
8. Fouad, S., Randell, D., Galton, A., Mehanna, H., Landini, G.: Unsupervised morphological segmentation of tissue compartments in histopathological images. *PLoS ONE* **12**(11), e0188717 (2017)
9. Jaskari, J., et al.: Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes. *Sci. Rep.* **10**, 5842 (2020)
10. Jeoun, B.S., et al.: Canal-net for automatic and robust 3d segmentation of mandibular canals in cbct images using a continuity-aware contextual network. *Sci. Rep.* **12**(1), 13460 (2022)
11. Kainmueller, D., Lamecker, H., Seim, H., Zinser, M., Zachow, S.: Automatic extraction of mandibular nerve and bone from cone-beam ct data. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009: 12th International Conference, London, UK, September 20–24, 2009, Proceedings, Part II 12, pp. 76–83. Springer (2009)
12. Kwak, G.H., et al.: Automatic mandibular canal detection using a deep convolutional neural network. *Sci. Rep.* **10**(1), 5711 (2020)
13. Lahoud, P., et al.: Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on cbct. *J. Dent.* **116**, 103891 (2022)
14. Lin, X., et al.: Accurate mandibular canal segmentation of dental cbct using a two-stage 3d-unet based segmentation framework. *BMC Oral Health* **23**(1), 551 (2023)
15. Oktay, O., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint [arXiv:1804.03999](https://arxiv.org/abs/1804.03999) (2018)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III, pp. 234–241. Springer International Publishing, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
17. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
18. Van der Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A.: Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Med. Image Anal.* **79**, 102470 (2022)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), 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 license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license 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.

