IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 22, 2021, accepted December 31, 2021, date of publication January 19, 2022, date of current version February 1, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144840

Deep Segmentation of the Mandibular Canal: A New 3D Annotated Dataset of CBCT Volumes

MARCO CIPRIANO[1], STEFANO ALLEGRETTI[1], FEDERICO BOLELLI[1], MATTIA DI BARTOLOMEO[2], FEDERICO POLLASTRI[1], ARRIGO PELLACANI[2], PAOLO MINAFRA[3], ALEXANDRE ANESI[4], AND COSTANTINO GRANA[1], (Member, IEEE)

[1]Department of Engineering "Enzo Ferrari," University of Modena and Reggio Emilia, 41121 Modena, Italy
[2]Surgery, Dentistry, Maternity and Infant Department, Unit of Dentistry and Maxillo-Facial Surgery, University of Verona, 37129 Verona, Italy
[3]Affidea Modena Medica Srl, 41100 Modena, Italy
[4]Department of Medical and Surgical Sciences for Children and Adults, Cranio-Maxillo-Facial Surgery, University of Modena and Reggio Emilia, 41121 Modena, Italy

Corresponding author: Federico Bolelli (ederico.bolelli@unimore.it)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program under Grant 825111, and in part by the DeepHealth Project.

Ethical approval for this study was obtained from Comitato Etico dell'Area Vasta Emilia Nord (Approval Number 1374/2020/OSS/ESTMO SIRER ID 1275 - NAI-CBCT-D).

ABSTRACT Inferior Alveolar Nerve (IAN) canal detection has been the focus of multiple recent works in dentistry and maxillofacial imaging. Deep learning-based techniques have reached interesting results in this research field, although the small size of 3D maxillofacial datasets has strongly limited the performance of these algorithms. Researchers have been forced to build their own private datasets, thus precluding any opportunity for reproducing results and fairly comparing proposals. This work describes a novel, large, and publicly available mandibular Cone Beam Computed Tomography (CBCT) dataset, with 2D and 3D manual annotations, provided by expert clinicians. Leveraging this dataset and employing deep learning techniques, we are able to improve the state of the art on the 3D mandibular canal segmentation. The source code which allows to exactly reproduce all the reported experiments is released as an open-source project, along with this article.

INDEX TERMS 3D imaging, CBCT, image dataset, medical imaging, inferior alveolar nerve.

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have provided amazing results for a multitude of computer vision tasks [1]–[7]. One of these, image segmentation, is an important processing step in medical image analysis, for a multitude of research fields including image-guided interventions, radiotherapy, or improved radiological diagnostics. Segmentation plays an important role before surgical operations. In fact, avoiding contact with the Inferior Alveolar Nerve (IAN) is a primary concern during dental implant placement within the mandible. These kinds of operations are routinely executed, but may become very complex due to the local presence of the IAN. The nerve might be in close relation to the roots of impacted teeth (especially the molars) and a detailed description of its position must be comprehended before the surgical removal (Fig. 1).

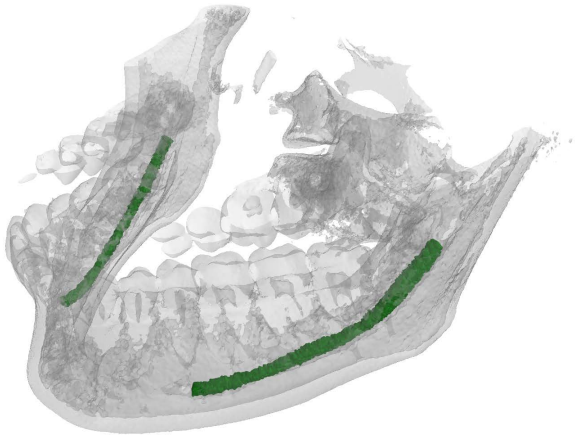


FIGURE 1. Example of a 3D annotated mandibular canal in a CBCT volume.

Nowadays, perfect anatomical annotation accuracy is usually not achieved, in favour of a fast execution time. A sparse annotation, performed on a 2D image, can be realized in

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

relatively small time, and has become the *de facto* standard in radiologic medical centers for dentistry and maxillofacial purposes (a specialty also known as dento-maxillofacial radiology). However, 2D annotations lack a considerable amount of inner information about the bone structure and the IAN position. Surgeons rely on a partial and incomplete idea of the nerve positioning, which is generally sufficient for a positive outcome of surgical intervention, but does not represent an accurate anatomical representation. Switching to 3D annotations solves the issue, but raises the costs in terms of working time: the manual segmentation of a single scan volume, depending on the available software, can take hours. Therefore, although dento-maxillofacial radiology would largely benefit from 3D annotated patients, any manual method currently appears unsuitable for practical applications.

On the other hand, a well-trained CNN can accurately segment the 3D structures of a scan in a matter of seconds. Accordingly, CNNs have the potential to supplement traditional medical imaging workflows, to reduce the associated costs. Unfortunately, the full potential of CNNs in this field is still untapped due to the lack of valid training sets. Indeed, despite the significant amount of raw information available, the supervised learning paradigm requires the presence of 3D densely labeled data in order to achieve its full potential. As already stated, acquiring large amounts of those high quality annotations is extremely expensive.

An additional constraint when dealing with the medical branch of deep learning regards the extra consideration for sensitive data, which must always be taken into account. Medical data contain a huge amount of personal information which must be protected through data anonymization. Beside the sensible details about the patient, generally stored in the DICOM format, when it comes to medical imaging, personal features of one subject can be inherently carried by the images. Therefore, in order to evaluate the correctness of the anonymization steps and the security of the information, every medical dataset must undergo the evaluation of specific committees. Several bureaucratic steps generally stand before the official release to the scientific community. This circumstance, combined with the laws of individual countries, creates a serious burden for the spread of medical datasets and make any contribution valuable. For those reasons, deep learning works applied to medical imaging — and in particular, to the maxillofacial field— are often based on private internal datasets [8]. Despite any alleged intention, datasets are rarely released after publication, creating a vast information gap for the research community: researchers are unable to replicate experiments, and lack valid benchmarks for their novelties.

In this work, we introduce and publicly release a novel maxillofacial dataset, annotated by medical practitioners at voxel level as shown in Fig. 1. We call this *3D annotation* in contrast to the traditional one, which is performed on a 2D ''panoramic view'' obtained from the volume (Fig. 2). Our dataset is specifically devised for the segmentation of the mandibular canal, and paves the way for future works which will finally benefit from a public reference in the field. We also prove the effectiveness of our proposal by showing that the current state of the art results in IAN segmentation can be improved by training a neural network with the proposed dataset.

From a medical point of view, many investigators are redefining anatomic landmarks fundamental in clinical practice (IAN injection, third molars removal, endodontics, placement of dental implants, IAN injury with osteotomic device, forensic dentistry) through CBCT 3D data [9]–[15]. The proposed dataset can improve the actual knowledge regarding the anatomy of the mandibular canal and its variations. Moreover, the open access granted to our 3D dataset is the basis for guaranteeing worldwide high-quality 3D data for investigators interested in further anatomic studies.
In summary, the novelties introduced by this paper are:

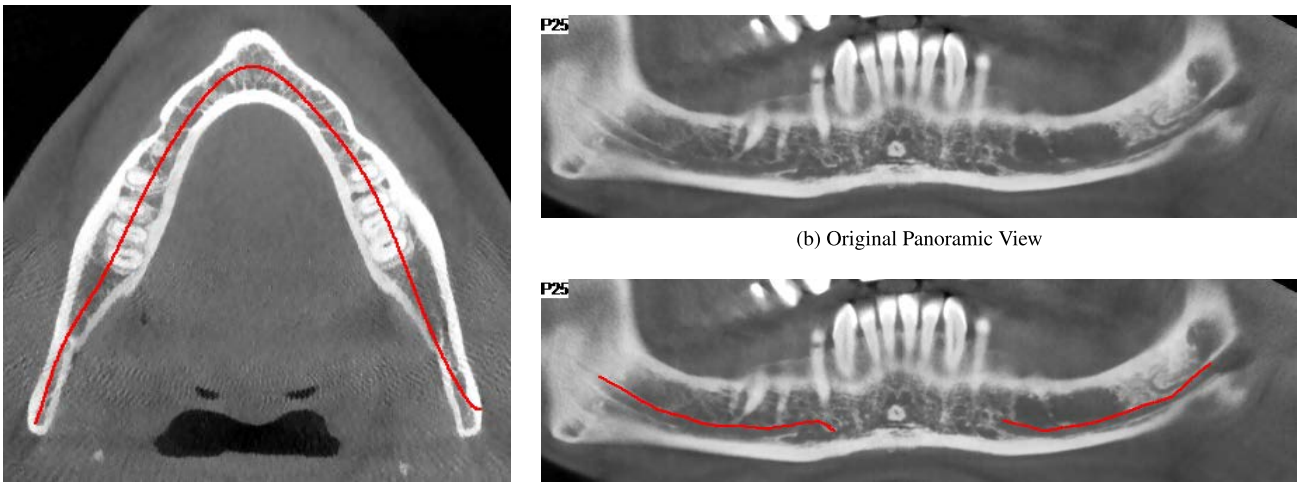
1) First publicly released CBCT (Cone Beam Computed Tomography) 3D dataset with professionally produced 3D annotations;
2) Improved state-of-the-art mandibular canal segmentation accuracy, using a newly trained deep learning architecture;
3) The source code which allows to exactly reproduce all the reported experiments is released as an open source project.

The paper is organized as follows: in Section II the state-of-the-art approaches for automatic detection of the inferior alveolar nerve are detailed. Section III describes the collected dataset, providing a comparison with the current training set-up available in the literature. Section IV includes an explanation of the processing methods used to refine the input data quality and introduces the deep learning model employed. Experiments are reported in Section V, and conclusions are drawn in Section VI.

II. RELATED WORKS

Three-dimensional imaging has long been a fundamental technology for diagnosis in dentistry and maxillofacial surgery [16]. Earlier studies have been focused on CT scans [17], [18]; then, after the diffusion of CBCT [19] in the early 2000s, a substantial amount of research has been dedicated to the development of automatic systems for the segmentation of the IAN in CBCT scans, both using classical computer vision methods [10], [20]–[25] and, more recently, machine learning and deep learning [26]–[28].

For what regards classical computer vision, one of the first fully automatic proposals by Kainmueller *et al.* [20] dates back to 2009. Their method is based on a combined Statistical Shape Model (SSM) of the bone surface and the nerve course, and the nerve position reconstruction is further improved with a Dijkstra-based tracing algorithm. A similar solution was also proposed by Abdolali *et al.* [23], who added a pre-processing step based on low-rank decomposition, and replaced the tracing algorithm with fast marching, to find the optimal path between mandibular and mental



(a) Axial Slice

(b) Original Panoramic View

(c) Annotated Panoramic View

FIGURE 2. CT annotation based on 2D panoramic views. (a) is an axial slice extracted from the CT volume. The red curve, called panoramic base curve, identifies the jawbone. (b) is the panoramic view obtained from the CT-volume displaying voxels of the curved plane generated by the base curve and orthogonal to the axial view. (c) is the same view of (b) showing a manual annotation of the IAN performed by an expert technician.

foramen. These methods, however, are limited by the need for segmented mandible bone in the training annotation, which requires additional manual work.

Following a different approach, in the work by Moris *et al.* [24], the volume is considered from three perspectives to produce a collection of slices, and the canal is extracted by searching several candidate positions in subsequent slices between the mental and mandibular foramen. These candidate positions are selected between those with the highest similarity with an ideal model of the canal. A weakness of the method is the use of predefined thresholds to separate the canal from other tissue, which often results in the exclusion of part of the canal, because of the low contrast of CBCT scans.

Recently, Wei and Wang [25] published a method based on the curved Multi-Planar Reconstruction (MPR) image set. The MPR and total average intensity projection panoramic image are reconstructed with the sampling distance of 1 pixel, and then texture features of the gray level-gradient co-occurrence matrix of the region of interest are clustered with K-means, to improve the image contrast of the IAN canal. Finally, the mandibular canal edges are segmented using 2D line-tracking, and the so obtained results are fitted by the fourth-order polynomial.

One of the first applications of deep learning to the segmentation of the mandibular canal is represented by the work by Jaskari *et al.* [28], who trained a fully convolutional network on a dataset of coarsely annotated 3D scans. On average, each canal is annotated using 10 manually assigned control points, which are then interpolated in a spline, and finally turned into a volume applying a static 3.0 mm diameter. Their approach achieves better results than previous attempts based on SSM, but is still hindered by the lack of handmade voxel-level annotations, and the suboptimal quality of segmentation masks automatically generated from coarse annotations.

Another recent work based on convolutional neural networks is due to Kwak *et al.* [27], who trained 2D and 3D models based on SegNet [29] and U-Net [1], [30] on a private annotated dataset. Because neither the dataset nor the code of the experiments is publicly available, it is not possible to directly compare our work to theirs.

To the best of our knowledge, none of the proposals related to the mandibular canal segmentation includes a link to the data or the source code required to reproduce the experiments. Few papers mention data availability on explicit request. We tried, but the answer (if any) has always been negative. This paper aims at breaking this objectively bad practice. The full dataset with annotations and subdivision in training and testing splits is available at https://ditto.ing.unimore.it/maxillo/.

III. DATASET

This section introduces and describes the maxillofacial dataset released with this paper, detailing the entire procedure carried out for data acquisition and the subsequent annotation steps. The 3D CBCT volumes composing our dataset have been acquired by the Affidea center located in Modena, Italy. Affidea is a leading pan-European healthcare group specialized in the provision of advanced diagnostics, specialist outpatient services, laboratory analyses, physiotherapy and rehabilitation, cancer diagnosis and treatment. It counts 312 different centers in 15 different countries, with about 11 000 professionals.

The dataset counts 347 dental scans obtained by means of Cone Beam Computed Tomography (*NewTom/NTVGiMK4*, 3 mA, 110 kV, 0.3 mm cubic voxels). Pixel spacing and intra-slice distance are always 0.3 millimeters. Data volumes are already converted to the Hounsfield Unit (HU) and their values range between −1 000 and 5 264. During the conversion to HU we also took care of the proper processing for the

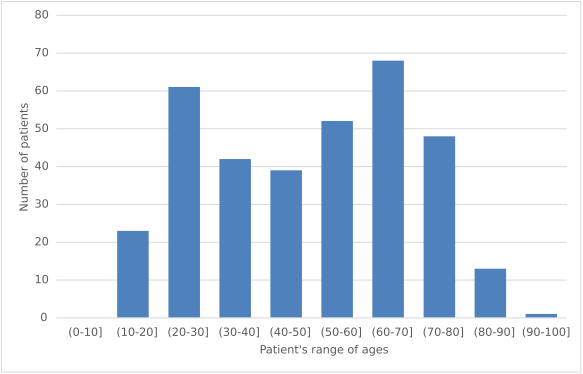


FIGURE 3. Overview of patients by age. Ranges of age on the X-axis, amount of patients on the Y-axis.

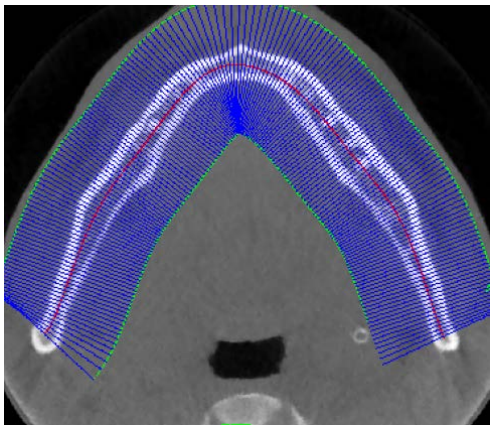


FIGURE 4. Panoramic base curve (red curve) computed on the axial slice, and series of axial cuts (blue lines) orthogonal to the panoramic base curve. Best viewed in color.

window width and center according to the DICOM format. Volume shapes range from (148, 265, 312) to (178, 423, 463) for the Z, Y and X axes respectively. Every patient was anonymized, hence we were only able to access few personal details - namely gender, age and year of the scan. Specifically, 59% of the patients are female, all the scans were performed between 2019 and 2020, and volumes belong to patients with ages in range (10-100] with the highest frequencies in ranges (20-30] and (60-70], see Fig. 3 as a reference.

After user registration, the dataset is available at https://ditto.ing.unimore.it/maxillo/.

A. 2D ANNOTATIONS

Technicians involved in the diagnostic exam were also responsible for the original *sparse* annotation of the mandibular canal. This annotation, performed on 2D panoramic views of the jawbone, is employed in everyday surgical practice to measure the height and depth of the sites where the implant must be placed, avoiding inferior alveolar nerve injuries. In these labels, the upper bound of the canal is marked along the entire dental arch, providing a useful sparse approximation trace of nerve position. Particularly, the annotation process starts from an axial slice of the original volume (Fig. 2a). Upon this slice, a spline is manually drawn to fit the central part of the jawbone (red curve in the image). This spline, called panoramic base curve, is then employed for generating the panoramic view (Fig. 2b), composed by the voxels of the curved plane identified by the base curve and orthogonal to the axial slice. From this view, the inferior alveolar nerve canal should be clearly identifiable. Fig. 2c depicts an annotation example provided by an expert technician.

Unfortunately, due to the poor amount of actual canal covered in the three-dimensional space, this type of annotation is still unsuitable for the training of a deep neural network. As a matter of fact, the lack of densely labeled volumes has long posed the problem of what kind of data should be used for training the neural network, and drove researchers to engineer new ways to infer 3D annotations for the training set. Jaskari *et al.* [28] based their method on the assumption that mandibular canals have a pipe-shaped structure. Starting from sparsely annotated volumes, they expanded the labels to produce a tubular shape. In order to create a 3D synthetic dataset out of sparsely annotated patients, we implemented the method described by Jaskari et al. from scratch as follows:

- For each point in the sparse annotation, the direction of the canal is firstly determined using the coordinates of the next point.
- A 1.6 voxel-long radius is computed to be orthogonal to the direction of the canal in that point, and a circle is drawn.
- The radius length is set to ensure a circle diameter of 3 millimeters in real-world measurements. This unit can differ due to the diverse voxel spacing specified in the patient DICOM files (0.3 millimeters for each dimension, in our data).
- The previous step produces a hollow pipe-shaped 3D structure, that is finally filled with traditional computer vision algorithms.

For the rest of the paper, this type of annotation will be referred to as *Circle Expansion* or *synthetic*, in opposition to the new densely annotated dataset introduced with the present work. We refer to Jaskari et al. as a competitor, and in Section V we prove how our new manual-annotated dataset overcomes the limitations of a training set generated in a semi-supervised manner.

B. 3D ANNOTATIONS

In order to obtain finely-grained annotations, a team of doctors with years of experience in maxillofacial surgery elaborated 91 volumes to produce *dense* voxel-level annotations of the canal. The proposed dataset of 347 scans is therefore divided into two partitions: the *primary* dataset, composed of the 91 volumes for which both dense and sparse annotations are available, and the *secondary* dataset, only sparsely annotated.

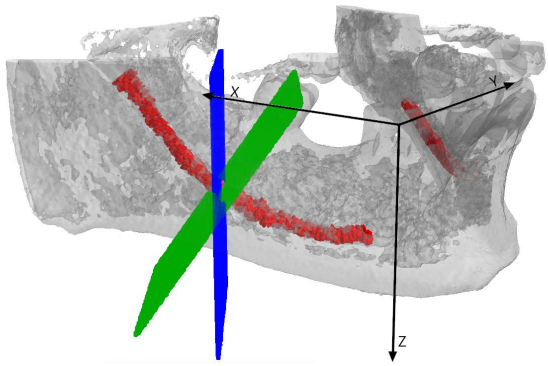


FIGURE 5. Visualization of the cross-sectional plane (blue) and its rotation around the CSL (green). The inferior alveolar nerve is depicted in red. Best viewed in color.

The entire voxel-level annotation procedure has been performed by means of the tool described in [31]. The annotation steps can be summarized as follows:

- After loading the input data, the arch approximation that better describes the canal course is identified inside one of the axial images and manually adjusted. The output is a one-pixel thick curve crossing the dental arch which is approximated with a polynomial. The result is similar to the one of Fig. 2a, but this time it is automatically generated and manually adjusted only if needed.
- Sampling the polynomial, the tool thus generates a Catmull-Rom spline. For each point of the spline, a perpendicular line (lying on the axial plane) is computed (Fig. 4). These lines are called *Cross-Sectional Lines* or CSLs in short. Here, a different resolution of the spline generates more or fewer CSLs. This represents a crucial step for having a complete annotation of the mandibular canal. Indeed, with a short spline, some regions of the jawbone —especially near the mandibular foramen— would be excluded from the next stages.
- CSLs are the base of *Multi Planar Reformations* (MPRs) called *Cross-Sectional Views* (CSVs). These views are 2D images obtained interpolating the values of the respective base lines (CSL) across the whole volume height. As an example, the blue plane reported in Fig. 5 is a *Cross-Sectional Plane* (CSP) generating a Cross-Sectional View. CSPs are additionally rotated around the CSLs, to have CSVs orthogonal to the canal slope (green plane of Fig. 5).
- For each CSV, a closed Catmull-Rom spline is finally drawn to annotate the position of the IAC (green lines of Fig. 6).
- The splines are saved as the coordinates of their control points. The final smooth and precise ground-truth volume constituting the dataset is generated from this set of points by means of the α-shape algorithm, which is described in details in Section IV-A.

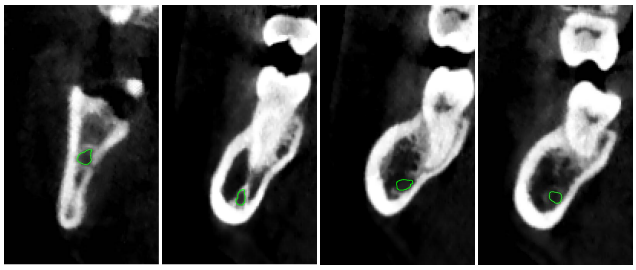


FIGURE 6. Examples of the Cross-Sectional Views (CSVs). The green closed spline is the annotated canal on the specific view. Combining the annotations from each CSV we obtain the final voxel-level dense annotation. Best viewed in color.

Comparing the two procedures, it is possible to see that while sparse annotations are quick and easy to obtain, creating dense labels from 3D volumes is a very tedious and time-consuming process. For this reason, researchers typically have few dense annotations available and preserve them as test set. Indeed, densely annotated volumes must always be used in test, to ensure a real medical feedback on model performance. A comparison between the 3D dense annotations provided by expert doctors and available in the new dataset and the synthetic generated ones is reported in Fig. 7, where the typical errors of the semi-supervised approach are clearly visible.

IV. METHOD

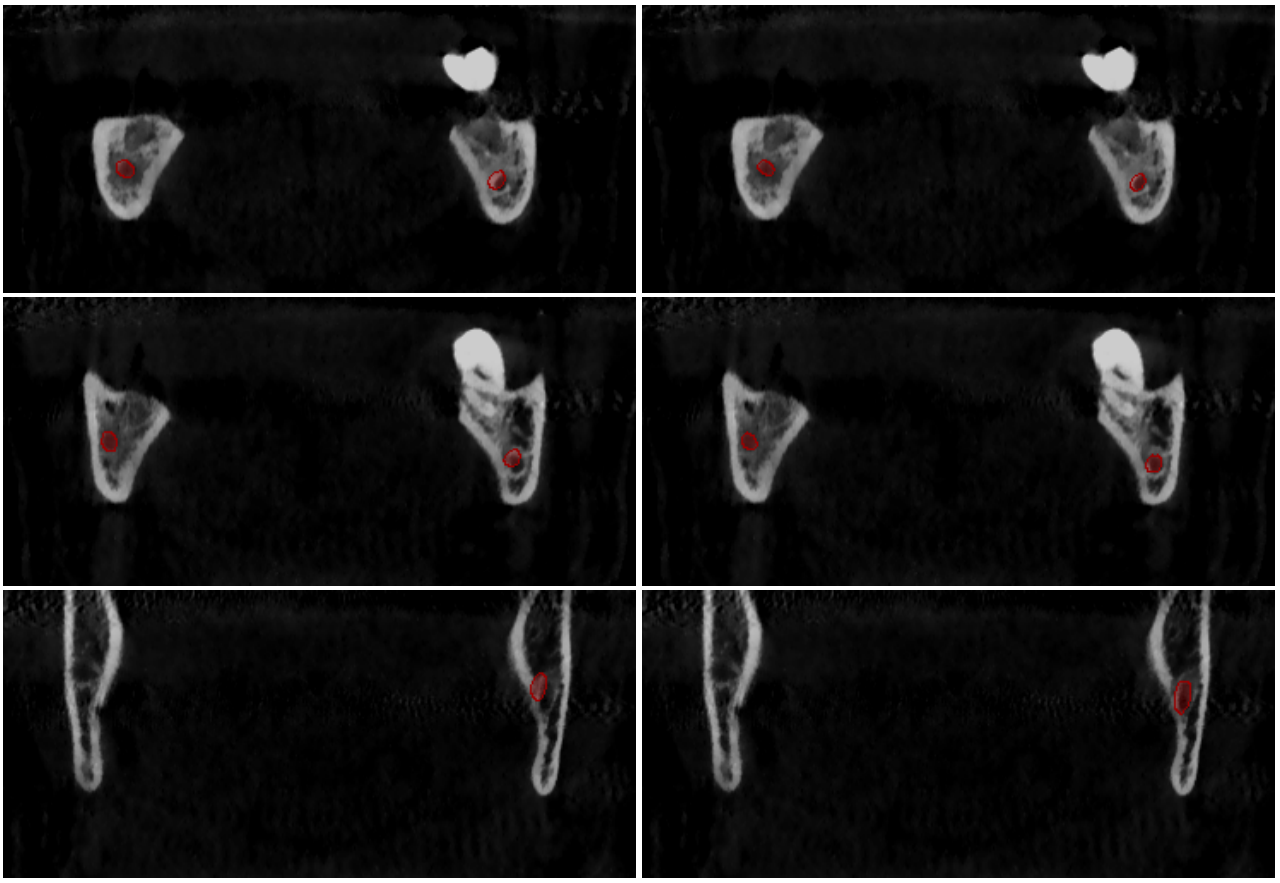
A. PRE-PROCESSING: ALPHA-SHAPE

The hand-drawn ground truth annotations produced with the tool described in [31] and summarized in the previous section result in a dense and jagged point set; an example is depicted in Fig. 8a. Starting from this point set, we reconstruct a smoother polygon mesh in the form of the α-*shape*.

The α-shape, defined by Edelsbrunner *et al.* [32], is a generalization of the concept of convex hull, useful to capture the intuitive notion of shape of a point set. The definition refers to the 2-dimensional case, but the extension to point sets in k-dimensions is straightforward. The α-shape is parameterized over $\alpha \in \mathbb{R}$, which determines the "crudeness" of the result. First, a *generalized disk of radius* $\frac{1}{\alpha}$, D_α, is defined in Equation 1.

$$D_\alpha = \begin{cases} \text{The complement of a disc of radius } -\frac{1}{\alpha}, & \text{if } \alpha < 0 \\ \text{A halfplane,} & \text{if } \alpha = 0 \\ \text{A disc of radius } \frac{1}{\alpha}, & \text{if } \alpha > 0 \end{cases} \quad (1)$$

Then, given a point set S and a specific value for α, the α-shape graph is constructed in the following way: an edge is created between two points p_i and p_j whenever there exists a D_α containing the entire S, and which has the property that p_i and p_j lie on its boundary. It is straightforward to notice that, when $\alpha = 0$, this process constructs the convex hull. Instead, positive or negative values of α allow to build



(a) Circle Expansion (b) Ours

FIGURE 7. Cross-Sectional Views of an annotated volume at different depth using (a) Circle Expansion (synthetic) and (b) α-shape labels contained in the proposed dataset. From this perspective, the canal can be spotted for its circular shape. This figure highlights the typical errors of the semi-supervised approach. First, the labels result inaccurate near the borders due to the natural variation of the canal along the jawbone. Furthermore, the ground truth mask is not centered —mostly near the mandibular foramen— because the sparse annotations used to generate these labels are generally marked on the upper area of the canal. To cope with this problem it is not sufficient to translate down the mask by a few voxels, because the width —hence the center— of the canal varies with its depth.

(a) Dense point set (b) α-shape polygonal mesh (c) Voxelized raster volume

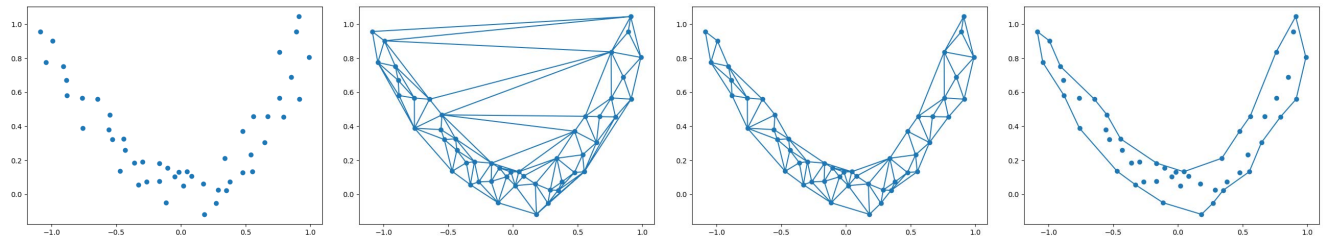
FIGURE 8. Ground truth preprocessing. The annotation tool outputs a dense and jagged point set (a), which shape is given by the concave α-shape (b). Finally, the so obtained polygonal mesh undergoes voxelization, resulting in a binary raster volume (c) that is used as ground truth for the training process.

cruder or finer shapes respectively, with the latters possibly including concave angles. Because of the geometrical nature of the alveolar nerve, we are indeed exclusively interested in concave α-shapes, *i.e.*, with $\alpha < 0$.

When $\alpha < 0$, the α-shape can be computed starting from the Delaunay triangulation: the set of triangles of the Delaunay triangulation whose circumradius is at most $-\frac{1}{\alpha}$ form a simplicial subcomplex, called α-*complex*, and its border coincides with the α-shape. The process is exemplified in Fig. 9.

The generalization of the above notions to 3-dimensions is just a matter of substituting disks and triangles with spheres and tetrahedra. An example of α-shape constructed from the volumetric annotations of the alveolar nerve is depicted in Fig. 8b.

The α-shape is a good representation of the annotated volume, but because it is a polygonal mesh, it cannot directly be used as a ground truth segmentation mask for training a neural network. Therefore, the next mandatory step is the voxelization, through which the α-shape is transformed into



(a) Point Set (b) Delaunay Triangulation (c) α-complex (d) α-shape

FIGURE 9. α-shape construction process for the point set of (a), with $\alpha = -2.5$. First the Delaunay triangulation is built (b), then only triangles whose circumradius is at most $-\frac{1}{\alpha}$ are kept, which form a subcomplex called α-complex (c). Finally, the border of the α-complex is the α-shape (d).

a binary raster volume. The voxelization of a polygonal mesh consists of finding which cubes (voxels) of a 3-dimensional grid intersect any triangle composing the mesh: the specific method used to check triangle-cube intersection is described in [33]. The final result of the ground truth preprocessing is illustrated in Fig. 8c.

B. MODEL

The main goal of this paper is to provide the research community with a new and valid dataset for the automatic segmentation of the alveolar canal. Hence, our focus went on proving that our data can be proficiently exploited for achieving this goal. For this reason, we choose to preserve competitor's pipeline —including the base model—, highlighting the differences between two identical training experiments carried out with the synthetic generated data (competitor) and with the proposed dataset (ours).

The neural network employed for the experiments is built from scratch, following the description provided in [28]. This 3D segmentation method is based on the U-Net 3D [30] fully convolutional neural network. Feature maps are down-sampled with stride 2 convolutions in the contractive path of the network, while they are up-sampled with stride 2 transpose convolutions in the expanding path. All the convolutions (transpose included) have a size of $3 \times 3 \times 3$ and are followed by batch normalization and Rectified Linear Unit non-linearity (ReLU). The only exception is for the last layer, which has a convolution with a $1 \times 1 \times 1$ size kernel and employs the logistic sigmoid as link function. Long skip connections concatenate the hidden layers of the contracting and expanding path, along the channel dimension. Moreover, each of the down-sampling and up-sampling blocks make use of residual connections. The PyTorch-based [34] implementation of this model can be found at https://ditto.ing.unimore.it/maxillo.

The training set is prepared offline: input values, stored in Hounsfield units, need to be cropped to avoid peak values due to metal artifacts in the patient mouth or acquisition noise. In their work, Jaskari *et al.* employed the range $[0, 3\,095]$, but our experiments identify 2 100 as the best upper bound value for our data. We then re-sampled all of our volumes to have a 0.4 millimeters voxel spacing from the original 0.3 millimeters. Finally, we executed a grid sampling [35] in which 32^3-sized patches were extracted using a stride of 22. Training patches without any mandibular canal voxel are discarded to reduce class imbalance. Data patches are shuffled on every training epoch and augmented with random flips in all the spatial dimensions, with a probability of 0.8.

The primary dataset —available both with dense and sparse annotations— is splitted to have 68, 15 and 8 volumes for training, test and validation set respectively. The grid-sampling process produced 4 831 and 3 897 training patches from the synthetic and dense datasets respectively.

Test and validation volumes were not subject to any sampling. On inference, a grid sampling is applied at run-time, without discarding any patch. Model outputs are then aggregated to rebuild the original volume shape and metrics are computed by using the initial ground truth volume directly.

Our secondary dataset —consisting of 256 sparsely annotated volumes— produced 19 466 patches and is used only for the training stage.

Numbers of patches resulting from the sampling process are conveniently summarized in Table 1. In each configuration setting, the model is trained for 100 epochs using a batch size of 24, and employing the Adam optimizer with an initial learning rate of 0.0001.

C. POST-PROCESSING

The 3D network outputs were further refined by the post-processing technique described in [28]. We implement their method by applying a 3D connected components labeling (CCL) [36] algorithm on the output masks, preserving only the two largest objects that should represent the left and right branch of the inferior alveolar nerve canal. Unlike Jaskari *et al.*, there is no need to verify at run-time how many connected components we have in our ground truth volumes since these are always two in the proposed dataset, namely the two branches of the canal. This filtering approach aims at removing the high amount of false-positive voxels, probably related to the small dimension of the employed patches.

By visually inspecting the network outputs, the background noise can be classified into two main categories:

1) False positives scattered over the mental foramen, but mostly not connected to the alveolar canal;

(a) Ground Truth (b) Competitor Prediction (c) Competitor Post-Processed Prediction (d) Our Prediction (e) Our Post-Processed Prediction

FIGURE 10. Examples of the post-processing algorithm when applied to our and competitor predictions. Specifically, (a) is the ground truth selected for this showcase, (b) is the competitor original prediction which produces (c) after noise filtering. Finally, (e) depicts our post-processed prediction when applying filtering on (d).

TABLE 1. Summary of the datasets employed for the experiments reported in section V.

Dataset Name	Split	Sparse	Dense	n. Volumes	n. Sparse Patches	n. Dense Patches
Primary	Training	✓	✓	68	4 831	3 897
	Validation	✓	✓	8	-	-
	Testing	✓	✓	15	-	-
Secondary	Training	✓	✗	256	19 466	-

2) Some additional noise sprinkled in very remote areas of the volume.

The post-processing technique appeared to be particularly effective to filter out both types of artifacts. See Fig. 10 as a reference.

V. RESULTS

This section reports the experimental results obtained by employing the dataset described in Section III and the methods depicted in Section IV (Jaskari *et al.*). Three separate tests are run, as displayed in Table 2.

To ensure experiment fairness and avoid producing fabricated results [37], [38], all of the hyperparameters are tuned using the randomly-selected validation set, which is also employed to apply the early-stopping technique and select the length of the training process. Once the training process is completed, inference is run over the official test set, and the different methodologies are compared over two different metrics.

Although both Intersection over Union and Sørensen–Dice similarity [39] (Dice score) are presented, their values are very strictly correlated, and thus only the latter is discussed in the following.

In the first experiment the synthetic data obtained from sparse annotations is compared against dense voxel-level annotations, by using solely the primary dataset. Although the number of volumes used for sampling is 68 in both cases (synthetic and dense), the amount of final patches for training have different values, as reported in Table 1. This difference is due to a diverse number of empty samples, which are discarded during the patch generation process. The tests result in 0.56 and 0.67 Dice score for the synthetic and dense labels respectively. The model trained on densely annotated data achieves a 20% improvement on the Dice score, proving how this new dataset is indeed suitable and effective for deep learning purposes.

In the second experiment we test the effectiveness of the whole synthetic dataset, first by training a CNN with only the secondary dataset, and then merging the primary and secondary datasets in a single training process. In this case, the sample patches amount to almost 20k, extracted from 256 volumes (Table 1). The model reaches a Dice score of 0.60, which rises to 0.62 when using the entire dataset. This test demonstrates that the best results obtained with a large, sparsely annotated dataset are unable to reach the performance achieved with a much smaller voxel-level labeled one.

Overall, the circle expansion technique can be useful for deep learning algorithms, despite being based on the wrong assumption that the canal preserves a regular shape on its entire body. Several clinical studies [40] have highlighted how the alveolar nerve diameter is subject to huge variations: larger in the mandibular foramen areas, shorter near the mental foramen. Circle expanded volumes exploit the direction of the canal over its entire body, but they provide no information about the actual thickness and size in the orthogonal directions to its axis.

Nonetheless, the second experiment suggests that the secondary sparsely annotated dataset might be useful for our segmentation task, as we further explore in the third test. In this regard, the first attempt consists in merging the primary and secondary subsets into a single dataset, despite the different quality of available labels. As a matter of fact, merging a large amount of low-quality labels with a much smaller amount of high-quality labels (experiment III, first line) can lower performance w.r.t. only using the smaller high-quality set [41], [42] (experiment I, second line). Unfortunately, due to the different number in samples —19 466 synthetic over 3 897 dense patches— the primary dataset fails to improve the performance with respect to the experiment conducted with synthetic labels only (0.62 Dice similarity). The second and most successful attempt, on the other hand, regards the adoption of the secondary dataset in a pre-training step before

TABLE 2. Experimental results using the different datasets and techniques.

Experiment	Primary	Secondary	Dataset Merging Technique	Jaskari et al.		Med3D	
				IoU	Dice	IoU	Dice
I	Circle Expansion	-	-	0.39	0.56	0.23	0.37
	Dense Annotation	-	-	0.52	0.67	0.43	0.59
II	-	Circle Expansion	-	0.43	0.60	0.35	0.51
	Circle Expansion	Circle Expansion	Union	0.45	0.62	0.34	0.51
III	Dense Annotation	Circle Expansion	Union	0.45	0.62	0.36	0.53
	Dense Annotation	Circle Expansion	Pretraining	0.54	0.69	0.45	0.62

using the primary dataset with dense labels for the fine-tuning of the network. This method is responsible for an additional 0.02 increment in the Dice similarity on the test set (0.69). To the best of our knowledge, this is the highest Dice score ever reached on the segmentation of the inferior alveolar canal.

To further consolidate the proposed benchmark, the performance of the described network is compared with Med3D (the last two columns of Table 2), a 3D segmentation CNN that exploits residual blocks and transfer learning to tackle several medical imaging tasks [43]. The performance of this architecture confirms all of the insights pointed out so far, albeit with lower overall accuracy *w.r.t.* the originally selected model.

We finally conclude this section with some evaluation of the post-processing technique. Experiments that solely rely on the circle expansion labels gain, on average, an additional 0.02 on both IoU and Dice metrics after the post-processing. These values are already considered in the scores presented in Table 2. As already stated, this is due to a large number of false positives in the prediction masks. Experiments that use our new densely labeled dataset appear to mitigate this issue, since no measurable benefits are achieved when the CCL filter is applied. However, the post-processing is always employed, due to the better qualitative result of the final output, clearly visible in Fig. 10.

VI. CONCLUSION

This paper presented a novel dataset of 3D annotated mandibular images. To the best of our knowledge, this is the first 3D mandibular dataset with voxel-level annotations of the Inferior Alveolar Nerve canal which is publicly available for the scientific community. Therefore, our work stands as a milestone for researchers who want to apply deep learning for the segmentation of the IAN canal. By training a well-known deep learning model with this new dataset we established a new state-of-the-art result for the canal segmentation. We strongly believe that our 3D dataset can boost the research in the field of mandibular canal anatomy, the knowledge of which is fundamental for clinicians who operate in the oral cavity.

In order to ensure experiments reproducibility, the pipelines described in the manuscript are publicly accessible to the scientific community as an open-source project alongside the proposed dataset. The code is wrapped in a *plug&play* set-up that easily allows researchers to replicate the reported experiments. The code includes several utilities to recreate the Circle Expansion dataset, the training patches and the α-shape outcomes. All the aforementioned resources can be reached by visiting https://ditto.ing.unimore.it/maxillo/.

ACKNOWLEDGMENT

The authors would like to thank Affidea's radiology health technician Marco Buzzolani for his valuable effort in collecting and anonymizing data.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017.

[3] L. Canalini, F. Pollastri, F. Bolelli, M. Cancilla, S. Allegretti, and C. Grana, "Skin lesion segmentation ensemble with diverse training strategies," in *Computer Analysis of Images and Patterns*, vol. 11678. Springer, Sep. 2019, pp. 89–101.

[4] F. Pollastri, F. Bolelli, R. Paredes, and C. Grana, "Augmenting data with GANs to segment melanoma skin lesions," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 15575–15592, 2019.

[5] G. Ligabue, F. Pollastri, F. Fontana, M. Leonelli, L. Furci, S. Giovanella, G. Alfano, G. Cappelli, F. Testa, F. Bolelli, C. Grana, and R. Magistroni, "Evaluation of the classification accuracy of the kidney biopsy direct immunofluorescence through convolutional neural networks," *Clin. J. Amer. Soc. Nephrol.*, vol. 15, no. 10, pp. 1445–1454, Oct. 2020.

[6] F. Pollastri, J. Maronas, F. Bolelli, G. Ligabue, R. Paredes, R. Magistroni, and C. Grana, "Confidence calibration for deep renal biopsy immunofluorescence image classification," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 1298–1305.

[7] S. Allegretti, F. Bolelli, F. Pollastri, S. Longhitano, G. Pellacani, and C. Grana, "Supporting skin lesion diagnosis with content-based image retrieval," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 8053–8060.

[8] P. Lahoud, S. Diels, L. Niclaes, S. Van Aelst, H. Willems, A. Van Gerven, M. Quirynen, and R. Jacobs, "Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT," *J. Dentistry*, vol. 116, Jan. 2022, Art. no. 103891.

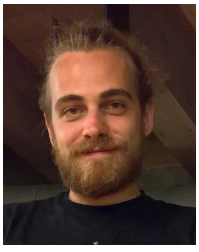
[9] A. Komal, R. S. Bedi, P. Wadhwani, J. K. Aurora, and H. Chauhan, "Study of normal anatomy of mandibular canal and its variations in Indian population using CBCT," *J. Maxillofacial Oral Surg.*, vol. 19, no. 1, pp. 98–105, Mar. 2020.

[10] J. Blacher, S. Van DaHuvel, V. Parashar, and J. C. Mitchell, "Variation in location of the mandibular foramen/inferior alveolar nerve complex given anatomic landmarks using cone-beam computed tomographic scans," *J. Endodontics*, vol. 42, no. 3, pp. 393–396, Mar. 2016.

[11] J. S. Angel, H. H. Mincer, J. Chaudhry, and M. Scarbecz, "Cone-beam computed tomography for analyzing variations in inferior alveolar canal location in adults in relation to age and sex," *J. Forensic Sci.*, vol. 56, no. 1, pp. 216–219, 2011.
[12] U. Nair, M. Yazdi, G. Nayar, H. Parry, M. Nair, and R. Katkar, "Configuration of the inferior alveolar canal as detected by cone beam computed tomography," *J. Conservative Dentistry*, vol. 16, no. 6, p. 518, 2013.
[13] S. Negrello, A. Pellacani, M. di Bartolomeo, G. Bernardelli, R. Nocini, M. Pinelli, L. Chiarini, and A. Anesi, "Primary intraosseous squamous cell carcinoma of the anterior mandible arising in an odontogenic cyst in 34-year-old male," *Reports*, vol. 3, no. 2, p. 12, 2020.
[14] A. Anesi, M. Ferretti, F. Cavani, R. Salvatori, M. Bianchi, A. Russo, L. Chiarini, and C. Palumbo, "Structural and ultrastructural analyses of bone regeneration in rabbit cranial osteotomy: Piezosurgery versus traditional osteotomes," *J. Cranio-Maxillofacial Surg.*, vol. 46, no. 1, pp. 107–118, Jan. 2018.
[15] A. Anesi, M. Di Bartolomeo, A. Pellacani, M. Ferretti, F. Cavani, R. Salvatori, R. Nocini, C. Palumbo, and L. Chiarini, "Bone healing evaluation following different osteotomic techniques in animal models: A suitable method for clinical insights," *Appl. Sci.*, vol. 10, no. 20, p. 7165, 2020.
[16] A. Schramm, M. Rücker, N. Sakkas, R. Schön, J. Düker, and N.-C. Gellrich, "The use of cone beam CT in cranio-maxillofacial surgery," *Int. Congr.*, vol. 1281, pp. 1200–1204, May 2005.
[17] T. Kondo, S. Ong, and K. Foong, "Computer-based extraction of the inferior alveolar nerve canal in 3-D space," *Comput. Methods Programs Biomed.*, vol. 76, pp. 181–191, Dec. 2005.
[18] S. Rueda, J. A. Gil, R. Pichery, and M. Alcañiz, "Automatic segmentation of jaw tissues in CT using active appearance models and semi-automatic landmarking," in *Medical Image Computing and Computer-Assisted Intervention–(MICCAI)*, R. Larsen, M. Nielsen, and J. Sporring, Eds. Berlin, Germany: Springer, 2006, pp. 167–174.
[19] W. C. Scarfe, A. G. Farman, and P. Sukovic, "Clinical applications of cone-beam computed tomography in dental practice," *J.-Can. Dental Assoc.*, vol. 72, no. 1, pp. 75–80, 2006.
[20] D. Kainmueller, H. Lamecker, H. Seim, M. Zinser, and S. Zachow, "Automatic extraction of mandibular nerve and bone from cone-beam CT data," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2009, pp. 76–83.
[21] D.-J. Kroon, "Segmentation of the mandibular canal in cone-beam CT data," Ph.D. dissertation, Univ. Twente, Enschede, Netherlands, Dec. 2011, doi: 10.3990/1.9789036532808.
[22] F. Abdolali and R. A. Zoroofi, "Mandibular canal segmentation using 3D active appearance models and shape context registration," in *Proc. 21th Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2014, pp. 7–11.
[23] F. Abdolali, R. A. Zoroofi, M. Abdolali, F. Yokota, Y. Otake, and Y. Sato, "Automatic segmentation of mandibular canal in cone beam CT images using conditional statistical shape model and fast marching," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 12, no. 4, pp. 581–593, Apr. 2017.
[24] B. Moris, L. Claesen, Y. Sun, and C. Politis, "Automated tracking of the mandibular canal in CBCT images using matching and multiple hypotheses methods," in *Proc. 4th Int. Conf. Commun. Electron. (ICCE)*, Aug. 2012, pp. 327–332.
[25] X. Wei and Y. Wang, "Inferior alveolar canal segmentation based on cone-beam computed tomography," *Med. Phys.*, vol. 48, no. 11, pp. 7047–7088, 2021.
[26] J.-J. Hwang, Y.-H. Jung, B.-H. Cho, and M.-S. Heo, "An overview of deep learning in the field of dentistry," *Imag. Sci. Dentistry*, vol. 49, no. 1, pp. 1–7, 2019.
[27] G. H. Kwak, E.-J. Kwak, J. M. Song, H. R. Park, Y.-H. Jung, B.-H. Cho, P. Hui, and J. J. Hwang, "Automatic mandibular canal detection using a deep convolutional neural network," *Sci. Rep.*, vol. 10, no. 1, p. 5711, Dec. 2020.
[28] J. Jaskari, J. Sahlsten, J. Järnstedt, H. Mehtonen, K. Karhu, O. Sundqvist, A. Hietanen, V. Varjonen, V. Mattila, and K. Kaski, "Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes," *Sci. Rep.*, vol. 10, no. 1, p. 5842, 2020.
[29] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Jan. 2017.
[30] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 424–432.
[31] C. Mercadante, M. Cipriano, F. Bolelli, F. Pollastri, M. Di Bartolomeo, A. Anesi, and C. Grana, "A cone beam computed tomography annotation tool for automatic detection of the inferior alveolar nerve canal," in *Proc. 16th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2021, pp. 724–731.
[32] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 551–559, Jul. 1983.
[33] D. Voorhies, "Triangle-cube intersection," in *Graphics Gems III*, D. Kirk, Ed. New York, NY, USA: Academic, 1992, pp. 236–239.
[34] A. Paszke, S. Gross, F. Massa, and A. Lerer, "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
[35] F. Pérez-García, R. Sparks, and S. Ourselin, "TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," *Comput. Methods Programs Biomed.*, vol. 208, Sep. 2021, Art. no. 106236.
[36] F. Bolelli, S. Allegretti, and C. Grana, "One DAG to rule them all," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
[37] A. E. Kavur, N. S. Gezer, M. Baris, S. Aslan, P. H. Conze, V. Groza, D. D. Pham, S. Chatterjee, and P. Ernst, "CHAOS challenge–combined (CT-MR) healthy abdominal organ segmentation," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101950.
[38] M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. Aviles-Rivero, C. Etmann, C. McCague, and L. Beer, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Mach. Intell.*, vol. 3, pp. 199–217, Mar. 2021.
[39] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons," *Biologiske Skrifter*, vol. 5, pp. 1–34, Jun. 1948.
[40] J. C.-H. Chen, L.-M. Lin, J. R. Geist, J.-Y. Chen, C.-H. Chen, and Y.-K. Chen, "A retrospective comparison of the location and diameter of the inferior alveolar canal at the mental foramen and length of the anterior loop between American and Taiwanese cohorts using CBCT," *Surgical Radiol. Anatomy*, vol. 35, no. 1, pp. 11–18, Jan. 2013.
[41] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May/Jun. 2017, pp. 746–753.
[42] M. Fabbri, G. Brasó, G. Maugeri, O. Cetintas, R. Gasparini, A. Ošep, S. Calderara, L. Leal-Taixe, and R. Cucchiara, "MOTSynth: How can synthetic data help pedestrian detection and tracking?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 10849–10859.
[43] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer learning for 3D medical image analysis," 2019, *arXiv:1904.00625*.

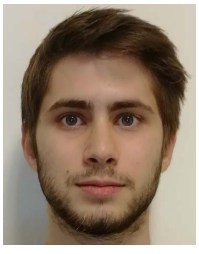
MARCO CIPRIANO received the B.Sc. and M.Sc. degrees in computer engineering from the Università degli Studi di Modena e Reggio Emilia, Italy. He is currently a Research Fellow with the AImageLab Laboratory, Dipartimento di Ingegneria "Enzo Ferrari," Università degli Studi di Modena e Reggio Emilia. His research interests include machine learning, deep learning, and medical imaging.

STEFANO ALLEGRETTI received the B.Sc. and M.Sc. degrees in computer engineering from the Università degli Studi di Modena e Reggio Emilia, Italy. He is currently pursuing the Ph.D. degree with the AImageLab Laboratory, Dipartimento di Ingegneria ''Enzo Ferrari,'' Università degli Studi di Modena e Reggio Emilia. His research interests include image processing, algorithm optimization, and medical imaging.

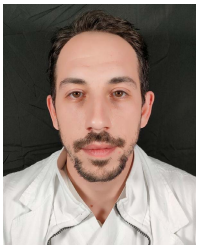


FEDERICO BOLELLI received the B.Sc. and M.Sc. degrees in computer engineering from the Università degli Studi di Modena e Reggio Emilia, Italy, where he is currently pursuing the Ph.D. degree. He is currently working as a Postdoctoral Researcher with the AImageLab Group, Dipartimento di Ingegneria ''Enzo Ferrari,'' Università degli Studi di Modena e Reggio Emilia. He is currently involved in two different H2020 European Projects. His research interests include image processing, algorithms and optimization, medical imaging, deep learning, and historical document analysis.

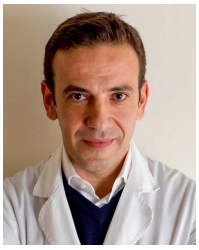
MATTIA DI BARTOLOMEO received the graduate degree *(cum laude)* in medicine and surgery from the Catholic University of the Sacred Heart, Rome, with an international master thesis in collaboration with the Plastic, Reconstructive and Hand Surgery Department, Maastricht University Medical Centre. He is a third-year resident in maxillofacial surgery at the University of Verona. He is currently attending the University Hospital of Modena. He has been involved in different research projects regarding head and neck surgical oncology, bone healing mechanisms, deep learning, vascular anomalies, aesthetic surgery, and temporomandibular joint diseases.



FEDERICO POLLASTRI received the B.Sc. and M.Sc. degrees in computer engineering from the Università degli studi di Modena e Reggio Emilia, Italy, where he is currently pursuing the Ph.D. degree with the AImagelab Laboratory, Dipartimento di Ingegneria ''Enzo Ferrari.'' His research interests include deep learning, pattern recognition, and medical imaging.



ARRIGO PELLACANI received the degree *(cum laude)* in medicine and surgery from the University of Modena and Reggio Emilia, in 2017. From 2017 to 2019, he worked as a freelance doctor. Since 2019, he has been a Resident Doctor in maxillofacial surgery with the University of Verona, and is currently attending the University Hospital of Modena. His research interests include oral surgery, maxillofacial traumatology, vascular anomalies, orthognathic, aesthetic, and oncological surgery.



PAOLO MINAFRA received the degree in medicine and surgery from the University of Palermo and achieved medical specialization in radiology from the University of Pavia. He is currently the Health Director, Chief Radiologist, and Sport Medicine Physician with Affidea, Modena Medica, Modena, Italy. He is also a Faculty Member ''Refresher Courses Small Parts Interventional Ultrasound'' with the Radiology Society of North America (RSNA). He has been the Head of Medical Staff and the Head Radiologist of several football teams. With regard to his scientific activity, he has been involved in the study of musculoskeletal ultrasound and elastosonographic evaluation for sports medicine. He is the coauthor of four books, focused on MSK imaging. He is an expert who has given several lectures, courses, and hands-on MSK ultrasound lessons worldwide.

ALEXANDRE ANESI received the degree in medicine and surgery from the University of Verona, with a specialization in maxillofacial surgery. He is currently an Assistant Professor of maxillofacial surgery with the Department of Medical and Surgical Sciences for Children and Adults, Faculty of Medicine, University of Modena and Reggio Emilia. He is also a part of the Teaching Board of the Ph.D. School in ''Materials, Mechatronics and Systems Engineering'' at the University of Trento. With regard to his scientific activity, he has been involved in the study of bioceramics-based bone grafting materials for regenerative biomedicine, bioceramics biocompatibility through in vitro/cytotoxicity assays and in vivo/implantation in animal models, bone-cutting device evaluation, deep learning, and head and neck surgical oncology.

COSTANTINO GRANA (Member, IEEE) received the degree from the Università degli Studi di Modena e Reggio Emilia, Italy, in 2000, and the Ph.D. degree in computer science and engineering, in 2004. He is currently a Full Professor with the Dipartimento di Ingegneria ''Enzo Ferrari,'' Università degli studi di Modena e Reggio Emilia, Italy. He published five book chapters, 38 articles in international peer-reviewed journals, and more than 100 papers in international conferences. His research interests include computer vision and multimedia and include medical imaging, image processing, analysis of digital images of historical manuscripts and other cultural heritage resources, multimedia image and video retrieval, and color-based applications.