Università degli Studi di Modena e Reggio Emilia

DIPARTIMENTO DI INGEGNERIA "ENZO FERRARI"

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

Inferior Alveolar Canal Segmentation using Deep Neural Networks

Relatore:
Prof. Costantino Grana
Correlatore:
Prof. Federico Bolelli

Candidato:
Luca Lumetti

ANNO ACCADEMICO 2021-2022

Abstract

Inferior Alveolar Canal Segmentation using Deep Neural Networks

Keywords: Medical Imaging, Cone Bean Computed Tomography, Inferior Alveolar Canal, Image Segmentation, Deep Learning.

Abstract in Italian

Utilizzo di Reti Neurali per la Segmentazione del Canale Alveolare Inferiore

Parole Chiave: Medical Imaging, Cone Bean Computed Tomography, Inferior Alveolar Canal, Image Segmentation, Deep Learning.

Summary in Italian

Il regolamento della Facoltà di Ingegneria di Modena prevede che le tesi scritte in lingua inglese debbano contenere un *abstract* ed un'ampia sintesi dei contenuti in lingua italiana: in accordo a questa regola, proponiamo un sunto degli argomenti, delle tecniche e dei risultati che verranno delineati nell'elaborato. Si noti che non si tratta di un sunto esaustivo, e che non è possibile valutare la tesi dalla semplice lettura di queste righe. Per una descrizione più dettagliata e più rigorosa, e per i risultati sperimentali ottenuti, si rimanda al testo in inglese.

Questa tesi tratta del... Durante il lavoro svolto è inoltre stato pubblicato un *paper* attualmente in fase di revisione (si veda l'Appendice per ulteriori dettagli).

To ...

Acknowledgements

Foremost, I would like to express my sincere gratitude to ...

Contents

Abstract in English i

Abstract in Italian ii

Summary in Italian iii

Acknowledgements 2

Contents 3

List of Figures 5

List of Tables 6

List of Listings 7

1 Segmentation of the Inferior Alveolar Canal: an overview 8

1.1 Introduction . . . 8
1.2 Inferior Alveolar Canal . . . 9
1.3 Cone Beam Computed Tomography . . . 9
1.4 DICOM file format . . . 10
1.5 Image Segmentation . . . 10

2 Segmentation Neural Networks 12

2.1 Segmentation using Deep Neural Networks . . . 12
2.1.1 Fully Convolutional Networks . . . 12
2.1.2 Convolutional Neural Networks . . . 13
2.1.3 Fully Convolutional Networks . . . 13
2.1.4 Encoder-Decoder Networks . . . 14
2.1.5 Regional Convolutional Neural Networks . . . 14

2.1.6 DeepLab 15
2.2 Dealing with 3D Medical Images 15

Bibliography **16**

List of Figures

1.1 Example of a multiclass semantic segmentation, different colors represent different classes 11

2.1 Groups of a segmentation network architecture. 13

List of Tables

List of Listings

Chapter 1

Segmentation of the Inferior Alveolar Canal: an overview

1.1 Introduction

Dental implant placement within the jawbone is a routine surgical procedure that can become complicated due to the presence of the Inferior Alveolar Nerve (IAN) nearby. The nerve, in particular, is frequently in close proximity to the roots of molars, and its position must thus be meticulously detailed prior to surgical removal. Avoiding contact with the IAN is a primary concern during these operations, thus its segmentation is crucial in surgical planning. Today the standard de-facto is to take a CBCT scan of the jawbone and a 2D panomaric view is extracted. This view allow medical experts to depict the IANs position with line. We refer to this type of annotation as *sparse* or *2D* annotation. A 3D annotation of the IAC is often avoided as it would require a huge amount of time, but this type of segmentation would offer a much precise knowledge of the position of the IAN and IAC and could allow dentists to plan a more detailed surgical approach. For this reason a lot of research about automatic segmentation of the IAC has been carried out and is still active today.

In this chapter we first describe in details the role of the IAN and the IAC, what a CBCT is and the definitions and carachteristics of different types of segmentations. In the following chapter we will take a look on how segmentation is performed today using neural networks paying more attention field of medical images. Next in chapter 3 we will detail the dataset, the network, the metrics, and other tools used as baseline to define the current state of the art, which is essential to understand the importance and goodness of the work that i've carried out. Last, in chapter 4, all the different techniques tried will be described in details, to conclude with a description of which would possible be the future work for this specific tasks.

1.2 Inferior Alveolar Canal

The Inferior Alveolar Canal (IAC) is a small passageway shaped as a tube that runs through the lower jawbone. It houses the Inferior Alveolar Nerve (IAN), which is responsible for transmitting sensory information from the teeth, gums and lips to the brain. It also provides motor innervation to the muscles of mastication (i.e. the muscles responsible for chewing). Dentists need to be able to accurately locate the IAC before performing certain surgical operations, such as tooth extractions or placement of dental implants. This is because the IAC is located very close to the roots of the teeth, and damage to the IAC during surgery can result in permanent nerve damage which would cause numbness, tingling, and pain in the affected area. In severe cases, it can also lead to muscle weakness and paralysis.

1.3 Cone Beam Computed Tomography

Cone beam computed tomography (CBCT) is a medical imaging technique consisting of X-ray computed tomography where rays are divergent, forming a cone. This type of computed tomography is well suited for imaging the craniofacial area as it provides clear images of highly contrasted structures, very helpfull to evaluate bones. It has become common in dentistry such as oral surgery, endodontics and orthodontics. The main reasons and advantages of CBCT with respect to other CTs are:

1. **X-ray beam limitation:** reducing the size of the irradiated area by collimating the primary x-ray beam to the are of interest minimizes the radiation dose. Most CBCT units can be adjusted to scan small regions for specific diagnostic task. They are also able to scan the whole craniofacial structure if needed.

2. **Image accuracy:** We created a novel, large, and publicly available maxillo-facial CBCT (Cone Beam Computed Tomography) dataset, with 2D and 3D manual annotations, provided by expert clinicians. All CBCT units provide voxel resolutions that are isotropic (i.e. equals in all the 3 dimensions) while in conventional CT, voxel are anisotripic (i.e. rectangular cubes).

3. **Rapid scan time:** CBCT acquires all the basis images in a single rotation, thus scan time goes from 10s to 70s. Although faster scanning time usually means fewer basis images from which to reconstruct the volumetric dataset, motion artifacts due to subject movement are reduced.

These advantages come with some drawbacks: Hounsfield units (HU) is the metric used to determine the radiodensity of tissue analized. In the Hounsfield scale, numbers go from values of -1000 for air to values of 1600 for dense bones. In CBCT scans, the radiodensity is inaccurate because different areas in the scan appear with different greyscale values depending on their relative positions in the organ being scanned, despite possessing identical densities, because the image value of a voxel of an organ depends on the position in the image volume. HU measured from the same anatomical area with both CBCT and medical-grade CT scanners are not identical and are thus unreliable for determination of site-specific, radiographically-identified bone density for purposes such as the placement of dental implants, as there is ”no good data to relate the CBCT HU values to bone quality” [1].
The images resulting from a CBCT scans are usually exported as DICOM (Digital Imaging and Communications in Medicine) which is the standard used worldwide to store, exchange, and transmit medical images.

1.4 DICOM file format

TODO?

1.5 Image Segmentation

Image segmentation is a well known topic in computer and image processing with a wide range of application, such as medical imaging, robotics, video surveillance, etc. It involves partitioning images into one or more objects and can also includes classify these objects.

Many traditional algorithms have been developed in the literature but, in the most recent years, they have all been dominated by deep neural networks. Since the 2015 a huge amount of different types of networks that aim to perform image segmentation has been proposed for each of the field where it's needed. Before presenting how nowday segmentation is performed, we must state which are the different type of segmentation that have been classified:

- **Semantic segmentation:** Semantic Segmentation perform a pixel-by-pixel classification with a predefined set of objects categories for all the pixels of the images. In pratice, given a RGB image (`height` × `width` × 3) we output a segmentation map of size (`height` × `width` × `classes`) where each value correspond to which class the same pixel in the original images belongs.
- **Instance segmentation:** One possible issue of semantic segmentation is that it doesn't allow to distinguish two or more object of the same class when they overlap in the image. Instance segmentation overcome this problem by outputting a different number of channels based on the number of instances present in the image.
- **Panoptic segmentation:** The latter type of segmentation is called Panoptic segmentation and is the result of the previously presented method joined together. The difference with the instance segmentation is that in this case instances are not allowed to overlap then for to a single pixel, a single instance must be assigned.

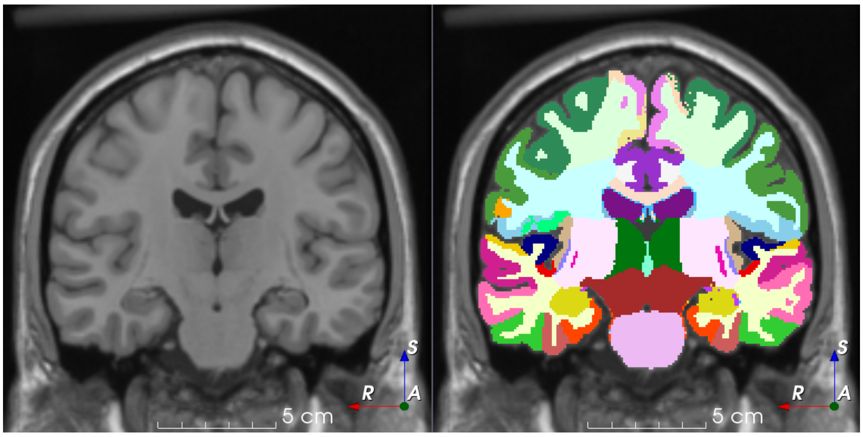

FIGURE 1.1: Example of a multiclass semantic segmentation, different colors represent different classes

Chapter 2

Segmentation Neural Networks

2.1 Segmentation using Deep Neural Networks

Today, deep neural netoworks are the state-of-the-art in many fields, including image segmentation. In this section, we will briefly review the most common approaches to segmentation using deep neural networks, and we will discuss the advantages and disadvantages of each approach.

2.1.1 Fully Convolutional Networks

Fully convolutional networks (FCNs) are a class of deep neural networks that are designed to perform pixel-wise classification. The main idea behind FCNs is to use a convolutional neural network (CNN) to extract features from the input image and return an output of the same size as the input image, where each pixel is assigned a class label. The main advantage of FCNs is that they can be trained end-to-end, which means that the network can be trained to perform the classification of each pixel without the need of any post-processing step. On the other hand, Deep Neural Networks lacks for explainability, which is a major drawback for medical applications. The architecture of such networks can be grouped as shown in Figure 2.1.

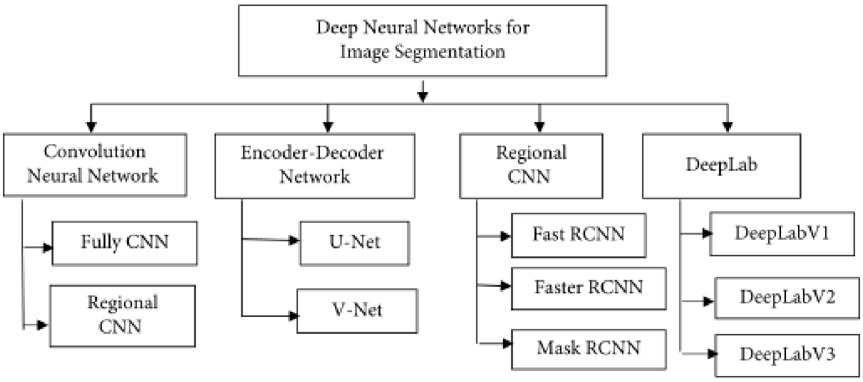

FIGURE 2.1: Groups of a segmentation network architecture.

2.1.2 Convolutional Neural Networks

A convolutional neural network or CNN consists of a stack of three main neural layers: convolutional layer, pooling layer, and fully connected layer. Each layer has its own role. The convolution layer detects distinct features like edges or other visual elements in an image. Convolution layer performs mathematical operation of multiplication of local neighbours of an image pixel with kernels. CNN uses different kernels for convolving the given image for generating its feature maps. Pooling layer reduces the spatial `(width, height)` dimensions of the input data for the next layers of neural network. It does not change the depth of the data. This operation is called as subsampling. This size reduction decreases the computational requirements for upcoming layers. The fully connected layers perform high-level reasoning in NN. These layers integrate the various feature responses from the given input image so as to provide the final results.
Different CNN models have been reported in the literature, including AlexNet, GoogleNet, VGG, Inception, SequeezeNet, and DenseNet. This type of networks were mostly used for classification, but they have been easily adapted to perform segmentation.

2.1.3 Fully Convolutional Networks

In fully convolutional network (FCN), only convolutional layers exist. The different existing in CNN architectures can be modified into FCN by converting the last fully connected layer of CNN into a fully convolutional layer. This type of networks can output spatial segmentation map and can have dense pixel-wise prediction from the input image of full size instead of performing patch-wise predictions. They can also uses skip connections, when performing upsampling on feature maps from final layer, these skip connections fuses it with the feature map of previous layers. The model thus produces a detailed segmentation in just one go but as drawback they do not have a global context of the image and the output can be fuzzy close to the boundaries of segmented objects.

2.1.4 Encoder-Decoder Networks

Encoder-decoder based models employ two-stage model to map data points from the input domain to the output domain. The encoder stage compresses the given input to a latent space representation, while the decoder predicts the output from this representation. This latent space representation is a compressed version of the input image, which is used to generate the output, it have smaller spatial dimension than the input image but an increased depth. In order to upsample this latent space representation to the size of the input image, transposed convolutional layers are used. One of the most popular encoder-decoder networks is the U-Net [**?**], for which we can find in literature many variants. The U-Net architecture is shown in Figure **??**. U-Net model has a downsampling and upsampling part. The downsampling section with FCN like architecture extracts features using 3×3 convolutions to capture context. The upsampling part performs deconvolution to decrease the number of computed feature maps. The feature maps generated by downsampling or contracting part are fed as input to upsampling part so as to avoid any loss of information. The symmetric upsampling part provides precise localization. The model generates a segmentation map which categorizes each pixel present in the image. This type of architecture proposed in 2015 is still widely used today as it can obtain state-of-the-art resuts.

2.1.5 Regional Convolutional Neural Networks

Regional convolutional network has been utilized for object detection and segmentation task. The R-CNN architecture presented in [69] generates region proposal network for bounding boxes using selective search process. These region proposals are then warped to standard squares and are forwarded to a CNN so as to generate feature vector map as output. The output dense layer consists of features extracted from the image and these features are then fed to classification algorithm so as to classify the objects lying within the region proposal network. The algorithm also predicts the offset values for increasing the precision level of the region proposal or bounding box. The processes performed in R-CNN architecture are shown in Figure 4. The use of basic RCN model is restricted due to the following:

2.1.6 DeepLab

DeepLab model employs pretrained CNN model ResNet-101/VGG-16 with atrous convolution to extract the features from an image. The use of atrous convolutions gives the following benefits:

- It controls the resolution of feature responses in CNNs
- It converts image classification network into a dense feature extractor without the requirement of learning of any more parameters employs conditional random field (CRF) to produce fine segmented output

The various variants of DeepLab have been proposed in the literature including DeepLabv1, DeepLabv2, DeepLabv3, and DeepLabv3+. In DeepLabv1, the input image is passed through deep CNN layer with one or two atrous convolution layers. This generates a coarse feature map. The feature map is then upsampled to the size of original image by using bilinear interpolation process. The interpolated data is applied to fully connect conditional random field to obtain the final segmented image.

2.2 Dealing with 3D Medical Images

It is not uncommon in the medical field to deal with 3D images that comes from CT scans or MRI scans. In this case, a 3D image can be represented as a stack of 2D images and fed them to the network. Each output is then stacked back to produce the final 3D output. Another approach is to use 3D convolutional networks. The 3D convolutional network is a generalization of the 2D convolutional network to 3D data. In the first case, each 2D layer is indipendend from the others, while in the second case, the 3D convolutional network is able to learn the spatial relationships between the different slices of the 3D image.

Bibliography

[1] Dale Miles and Robert Danforth. A clinician's guide to understanding cone beam volumetric imaging (cbvi). *Acad Dent Ther Stomatol*, pages 1–13, 01 2007.