Machine learning methods as an aid in planning orthodontic treatment on the example of Cone-Beam Computed Tomography analysis: a literature review

Szymon Plotka[1], Tomasz Wlodarczyk [1], Ryszard Szczera [1], Anna Chomiak [2], Oskar Komisarek [2]

1 Institute of Informatics, Warsaw University of Technology, Warsaw, Poland
2 Department of Maxillofacial Orthopaedics and Orthodontics, Poznan University of Medical Sciences, ul. Bukowska 70, 60–812 Poznań, Poland

Abstract

Convolutional neural networks (CNNs) are used in many areas of computer vision, such as object tracking and recognition, security, military, and biomedical image analysis. In this work, we describe the current methods, the architectures of deep convolutional neural networks used in CBCT. Literature from 2000-2020 from the PubMed database, Google Scholar, was analyzed. Account has been taken of publications in English that describe architectures of deep convolutional neural networks used in CBCT. The results of the reviewed studies indicate that deep learning methods employed in orthodontics can be far superior in comparison to other high-performing algorithms.

Key words: CBCT; Cone-Beam Computed Tomography; deep learning; machine learning; orthodontics

Introduction

Cone beam computed tomography (CBCT) was introduced to dentistry in 1998. These images are created by flat panel sensors and projection beam to capture 2D images. Then they are mathematically converted to a 3D image. It is possible to achieve 3D model of the patient's skull and teeth by using CBCT. Because of this CBCT is increasingly utilized in orthodontics...
diagnosis and treatment planning. The current indications for CBCT in orthodontics are
diagnosis of impacted teeth, root resorption, supernumerary teeth, cleft lip and/or palate,
skeletal malocclusions, congenital defects of the facial part of the skull, TMJ disorders,
asymmetries, assessment of the thickness of the bone, planning the location of orthodontic
implants and micro-implants [1].

Convolutional neural networks (CNNs) are used in many areas of computer vision, such as
object tracking and recognition, security, military, and biomedical image analysis. Advances in
medical imaging technologies and methods allow CNNs to be used in orthodontics to shorten
the planning time of orthodontic treatment, including an automatic search of landmarks on
cephalometric X-ray images, tooth segmentation on Cone-Beam Computed Tomography
(CBCT) images or digital models, and classification of defects on X-Ray panoramic images
[2]. This type of network works very well for pattern analysis and recognition [3], object
tracking [4], and medical image analysis [5], [6], [7]. In this work, we describe the current
methods, the architectures of deep convolutional neural networks used in CBCT.

Material and methods

Literature from 2000-2020 from the PubMed database, Google Scholar, was analyzed, entering
the terms: CBCT, Cone-Beam Computed Tomography, deep learning, machine learning,
orthodontics. Account has been taken of publications in English that describe architectures of
deep convolutional neural networks used in CBCT. Special attention has been paid to accuracy
of the obtained results [Table 1].

Results

These methods can handle a wider range of CBCT scans, an example of which is CNN. CBCT
works can be divided into two groups: classification and segmentation.

Pavaloiu et al. [8] created a 2 layer network that detects edge of teeth on a CBCT axial slice but
there are no quantitative results for comparison with future methods. Miki et al. [9] utilized
AlexNet to classify a ROI from a CBCT axial slice into 7 tooth types with skip ROIs of the 3rd
molar or metal artifacts. The same authors upgraded AlexNet and output a heatmap of a CBCT
axial slice for tooth detection and use bounding boxes on the result if there is over a 95%
certainty that it is a tooth. Kim et al. [20] introduce a method which classifies CBCT scans into
three types of malocclusions. Classification consists in comparison of pre-trained VGG16 with
Inception-V3-based architecture. The Inception-V3 model achieve the best outcome when the
output was based on a two-step learning method show in the article. VGG-16 based architecture
achieve the best outcome when the results from 3D images based on a voting scheme. Lee et
al. [25] applied Inception-V3 with transfer learning for classification in order to present
classification of cystic lesions from CBCT axial scans and panoramic x-rays. CBCT scans
proved to be more useful because they are easier to precisely classify than panoramic images.
Segmentation 2D

Egger et al. [11] use CNNs for mandible segmentation. VGG-16 classification network allows to check if the mandible is or not in the image. If the mandible is present then it goes through three VGG-16-based nets in order to segmentation. Authors reach a Dice score of 0.8964 and affirm that a greater dataset would achieve better result. Minnema et al. [15] present dense CNN - a mixed-scale for metal artifact reduction. Their method achieved better outcomes than the clinical benchmark, snake evolution. They receive comparable results to U-Net and ResNet with 300x and 700x less trainable parameters accordingly. Consecutive papers are based on one of the most popular neural network architectures used for segmentation - U-Net. Torosdagli et al. [16] present three-step method. This method consists of Tiramisu network, a U-Net-based learning algorithm and a LSTM network. This combination allows automatic segmentation and annotation of 9 anatomical landmarks of the mandible. In order to compare they use their private dataset and the public available MICCAI Head-Neck Challenge (2015) [21]. Kwak et al. [22] compare different methods based on 2D U-Net, 3D U-Net [23], and 2D SegNet [24] in effect of what they present segmenting the mandibular canal in a CBCT scan. The interesting thing is that SegNet shows better results than the widely-used U-Net. The 3D U-Net has the best results but on the other hand the image is downsized to 132 x 132 x 132, which results in a loss of accuracy. Setzer et al. [27] are able to detect periapical lesions in CBCT scans by using U-Net architecture. Output of the network has 5 classes: bone, tooth, lesion, restorative materials and background. Lee et al. [26] similarly to previous authors choose U-Net architecture as a method for tooth segmentation from a CBCT scan. They elaborated a multi-phase training method with each phase increasing the area around the teeth. Then it is passed as the training dataset. Thanks to the above mentioned model, the model converges faster and empty voxels are handled. Minnema et al. [15] and Lee et al. [26] obtained better results (0.917 vs 0.87 DSC) segmenting teeth from CBCT scan, with more training data and more scans (precisely 73) with metal artifacts. Nevertheless the MS-D network [15] achieved the same outcomes with less training parameters.

Segmentation-3D

Chen et al. [17] and Ezhov et al. [13] present methods using a V-Net [19]. The Chen et al. [17] applied a V-Net to output a tooth probability map and a tooth surface map. Marker-controlled watershed transform is used in order to tooth segmentation. According to the authors for their segmentation method a patch size of 64 is optimal. Because of the time required to make voxel-
level masks authors have a small set of data. Ezhov et al. method [13] showed outcomes for the segmentation of CBCT voxels into 33 classes (32 teeth + background). The method uses sequences of two models. First model is coarse and outputs 33 classes. Second, the fine model outputs 2 classes: data is about of a given tooth type or not. Authors use 2 types of datasets: coarse and precise. The coarse dataset is formed by linearly interpreted bounding boxes of CBCT axial slices. The fine dataset is a per voxel mask executed manually. Average Symmetric Surface Distance was used in this method as one of the specifications to estimate the model. Average Symmetric Surface Distance aka ASSD or ASD is the medium distance from all points on the boundary predicted mask to ground and the other way around [12]. Because of time required to have an accurate set of data there were only 120 precies scans. In this paper the author shows that using a coarse dataset in comparison with only a fine dataset allows to achieve better outcome. Cui et al. [14] present the ToothNet architecture created on the grounds of Mask R-CNN. The method is two-step: an edge detection of the CBCT and then a 3D region proposal module. This method is applicable for 3D data both quantitatively and qualitatively. Chung et al. [18] created a method which includes metal artifacts. This was achieved by an improvement segmentation method on ToothNet. Initially authors use a 2D CNN for pose regression, as the second for tooth detection they use a Faster R-CNN based network and after all a 3D U-Net based network is applied to tooth segmentation.

Compared to ToothNet, they obtain a 0.93 F1 score while ToothNet scores 0.88. F1-score is a harmonic mean of recall and precision combined (see Eq. 1) in the following way:

$$F1 \text{ score} = \frac{2 \cdot \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

It is used to compress the performance of a model into a single metric. Wu et al. [28] also tried to solve the problem of tooth segmentation. Authors show two-step approach. In their method the first stage is a heatmap regression U-Net for tooth ROIs. Then in the second stage they use a network which is based on U-Net and Spatial Pyramid Pooling for segmentation. Because of the time required to do voxel-level masks they use small set of data. In the last paper authors present approach which allow to obtain better results with a DSC of 0.962 than either of the previous papers. It is possible because using heatmap regression and Pyramid Polling [29] can retain the resolution of ROI. Example of a diagram (see Fig. 1) that shows Cone-Beam Computed Tomography (CBCT) processing to segment teeth.
Table 1. Comparison of papers based on Cone-Beam Computed Tomography (CBCT) images by year

<table>
<thead>
<tr>
<th>Author</th>
<th>Architecture</th>
<th>Results (average accuracy if not mentioned)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavaloiu et al. [8]</td>
<td>Neural network</td>
<td>Quantitative</td>
<td>2015</td>
</tr>
<tr>
<td>Miki et al. [9]</td>
<td>CNN (AlexNet)</td>
<td>0.744 (detection)</td>
<td>2017</td>
</tr>
<tr>
<td>Miki et al. [10]</td>
<td>CNN (AlexNet)</td>
<td>0.888</td>
<td>2017</td>
</tr>
<tr>
<td>Egger et al. [11]</td>
<td>CNN (VGG-16, FCN)</td>
<td>0.8964</td>
<td>2018</td>
</tr>
<tr>
<td>Lee et al. [25]</td>
<td>Inception-V3</td>
<td>0.914</td>
<td>2020</td>
</tr>
<tr>
<td>Kim et al. [20]</td>
<td>VGG-16 or Inception-V3</td>
<td>0.9333 (VGG-16), 0.9383 (Inception-V3)</td>
<td>2020</td>
</tr>
<tr>
<td>Kwak et al. [22]</td>
<td>2D U-Net, 3D U-Net and 2D SegNet</td>
<td>0.577 (3D U-Net, mean IoU), 0.491 (2D SegNet, mean IoU), 0.459 (2D U-Net, mean IoU), 0.95 (3D U-Net accuracy), 0.902 (2D SegNet, accuracy), 0.68 (2D U-Net, accuracy)</td>
<td>2020</td>
</tr>
<tr>
<td>Setzer et al. [27]</td>
<td>U-Net</td>
<td>0.714 (Dice score)</td>
<td>2020</td>
</tr>
<tr>
<td>Chung et al. [18]</td>
<td>2D CNN, Faster R-CNN, 3D U-Net</td>
<td>0.86 (IoU)</td>
<td>2020</td>
</tr>
<tr>
<td>Minnema et al. [15]</td>
<td>MS-D CNN</td>
<td>0.87 (DSC)</td>
<td>2020</td>
</tr>
<tr>
<td>Lee et al. [26]</td>
<td>U-Net</td>
<td>0.918 (Dice score)</td>
<td>2020</td>
</tr>
<tr>
<td>Cui et al. [14]</td>
<td>3D Mask R-CNN</td>
<td>0.924 (DSC), 0.995 (detection accuracy), 0.958 (identification accuracy)</td>
<td>2020</td>
</tr>
<tr>
<td>Chen et al. [17]</td>
<td>V-Net</td>
<td>0.936 (DSC), 0.881 (IoU), 0.363 (ASSD)</td>
<td>2020</td>
</tr>
<tr>
<td>Torosdagli et al. [16]</td>
<td>CNN, LSTM</td>
<td>0.9382 (DSC)</td>
<td>2020</td>
</tr>
<tr>
<td>Ezhov et al. [13]</td>
<td>V-Net</td>
<td>0.94 (IoU)</td>
<td>2020</td>
</tr>
<tr>
<td>Wu et al. [28]</td>
<td>U-Net</td>
<td>0.962 (DSC), 0.995 (detection accuracy), 0.991 (identification accuracy), 0.122 (ASSD)</td>
<td>2020</td>
</tr>
</tbody>
</table>
Discussion

In general, initial papers focus on tooth detection to sequentially go to tooth segmentation. Then, in 2018, interest in using CNN has increased. In 2019 ToothNet was the first reports of tooth segmentation in CBCT images. Next papers was based on this approach. Authors tried to improve the proposed method by including wisdom teeth and overcoming problem with metal artifacts.

Currently in 2020, the number of published papers about this subject still growing. Authors show state-of-the-art results in tooth instance segmentation (0.962 Dice). Researchers approach new pathways using CBCT to detection of lesions in craniofacial, detection of cysts, segmentation of mandibular canal and other [38]. The greatest challenge is the morphological heterogeneity of humans and the fact that different norms of facial aesthetics are accepted in different cultures. Each patient presents various diseases, takes various medications, is often burdened with genetic disorders, injuries, and hospitalizations, which may affect the therapeutic process. As a result, the number of therapeutic combinations is unlimited [35]. After collecting the medical history, examining the patient, and performing additional tests, i.e. diagnostic model, CBCT, and cephalometric X-ray, an individual treatment plan is created. Sometimes, a
patient may not consent to a proposed treatment plan [36]. The orthodontist may then suggest an alternative treatment plan which further increases the number of therapeutic combinations. Another challenge is the patient himself. The obtained therapeutic effects depend on the cooperation of the patient who, due to insufficient oral hygiene, led to dental caries, periodontal diseases, or, in the case of clear aligner appliances, did not wear them long enough [37].

For the neural network to be able to effectively solve orthodontic problems, it is necessary to load the appropriate amount of data obtained from patients. In the case of rare abnormalities, the amount of available data will be not sufficient for AI to recognize them - therefore it will take a long time for the neural network to be as efficient at solving clinical problems at the same level of effectiveness as an orthodontist. The last problem in training the neural network is the time the orthodontist has to spend annotating DICOM, JPG, and STL files.

Conclusion
We give an overview of the various works concerning each method and compare them based on achieved results. The results of the reviewed studies indicate that deep learning methods employed in orthodontics can be far superior in comparison to other high-performing algorithms. Therefore, we believe that deep learning approaches will play a significant role in orthodontics.

References


102


