Computerized Medical Imaging and Graphics 68 (2018) 61–70

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

An effective teeth recognition method using label tree with cascade network structure

Kailai Zhang[a,*], Ji Wu[a], Hu Chen[b], Peijun Lyu[b]

[a] Department of Electronic Engineering, Tsinghua University, Beijing, China
[b] Center of Digital Dentistry, Peking University School and Hospital of Stomatology & National Engineering Laboratory for Digital and Material Technology of Stomatology, Beijing, China

ARTICLE INFO

Keywords:
Teeth recognition
Convolutional neural network
Label tree
Cascade structure

ABSTRACT

In this article, we apply the deep learning technique to medical field for the teeth detection and classification of dental periapical radiographs, which is important for the medical curing and postmortem identification. We detect teeth in an input X-ray image and distinguish them from different position. An adult usually has 32 teeth, and some of them are similar while others have very different shape. So there are 32 teeth position for us to recognize, which is a challenging task. Convolutional neural network is a popular method to do multi-class detection and classification, but it needs a lot of training data to get a good result if used directly. The lack of data is a common case in medical field due to patients' privacy. In this work, limited to the available data, we propose a new method using label tree to give each tooth several labels and decompose the task, which can deal with the lack of data. Then use cascade network structure to do automatic identification on 32 teeth position, which uses several convolutional neural network as its basic module. Meanwhile, several key strategies are utilized to improve the detection and classification performance. Our method can deal with many complex cases such as X-ray images with tooth loss, decayed tooth and filled tooth, which frequently appear on patients. The experiments on our dataset show: for small training dataset, compared to the precision and recall by training a 33-classes (32 teeth and background) state-of-the-art convolutional neural network directly, the proposed approach reaches a high precision and recall of 95.8% and 96.1% in total, which is a big improvement in such a complex task.

1. Introduction

Although physiological characteristics such as face and fingerprint are the most common information for person recognition, it is difficult for them to provide accurate results when one is dead, especially died from disaster or traffic accident because these features lost easily. However, as the hardest part of human, teeth can remain unbroken after one's death, thus it is applicable for person recognition (Zhou and Abdel-Mottaleb, 2005; Nomir and Abdel-Mottaleb, 2005; Tohnak et al., 2007), even for the recognition of dead person (Valenzuela et al., 2000). Therefore, the teeth detection and classification is important for legal medical expert to do postmortem identification (Nomir and Abdel-Mottaleb, 2007; Jain and Chen, 2004).

Most adults have 32 teeth in total (there are eight teeth on each side of the jaw) as expected, which means there are 32 teeth position for us to detect. Meanwhile, the medical X-ray image is deficient because it is related to patients' privacy, and usually we cannot get as much X-ray image data as we want. So it is difficult to do teeth detection and classification based on a large number of class relative to a small number of image data. In order to achieve automatic teeth recognition, many scholars achieve the X-ray image preprocessing and teeth segmentation based on image processing and segmentation algorithms such as graph-based method (Felzenszwalb and Huttenlocher, 2004; Carreira and Sminchisescu, 2011; Boykov and Jolly, 2001), contour detection method (Arbeláez et al., 2011) and level set method (Li et al., 2006), which only need a small quantity of image data.

Shah et al. (2006) used active contour without edges technique to extract the contour of the teeth based on the intensity of the overall region of the tooth image. Therefore, while does not necessitate the presence of a sharp boundary between teeth, the region contour can be extracted in the presence of additive noise and in the absence of well-defined image gradients. Said et al. (2006) offered a mathematical morphology approach to the problem of teeth segmentation. Rad et al. (2013) used clustering (k-means) method for teeth segmentation step

* Corresponding author at: Faculty of Electronic Engineering, Tsinghua University, Beijing, China.
E-mail address: zhangkl17@mails.tsinghua.edu.cn (K. Zhang).

https://doi.org/10.1016/j.compmedimag.2018.07.001
Received 1 March 2018; Received in revised form 27 June 2018; Accepted 5 July 2018

and extracted multiple features of dental X-ray images using texture statistics techniques by gray-level co-occurrence matrix. Several articles present algorithms to classify and assign numbers to teeth in bitewing dental images. Mahoor and Abdel-Mottaleb. (2005) used Bayesian classification to classify the teeth into molars and premolars in a bitewing image, then assign an absolute number to each tooth based on the common numbering system used in dentistry. Aeini and Mahmoudi. (2010) deducted the exact location and size of the mesiodistal neck for each type of tooth and finished the teeth classification work using a linear model. Lin et al. (2010) proposed a binary linear support vector machine using the skew-adjusted relative length/width ratios of both teeth, pulps, and crown size as features to classify each tooth to molar or premolar, then a numbering scheme, which combined a missing teeth detection algorithm and a simplified version of sequence alignment commonly used in bioinformatics, was presented to assign each tooth a proper number. Yuniarti (2012) performed dental classification process which aims to classify the extracted tooth into molar or premolar using a binary support vector machine method before the teeth numbering process, obtaining good performance of classification in both bitewing radiographs and panoramic radiographs. Since the periapical radiographic images are common used in clinical more than bitewing and panoramic radiographs, and consists only of either upper or lower series of teeth, this type of radiograph is more representative to show disaster conditions. For example, the upper and lower jaws of a victim were usually disintegrated in the flight accident. Dental classification for periapical radiograph based on multiple fuzzy attribution is proposed by Tangel et al. (2013), where each tooth is analyzed based on multiple criteria such as area/perimeter ratio and width/height ratio.

The automatic identification algorithm established in the above literatures mostly extract the image features manually from dental bitewing radiographs. Although some of them have high precision on their test set, the workload of feature extraction is large, and the performance of the image detection and classification is greatly dependent on the quality of the extracted features. Unfortunately, for the dental periapical radiographs, there are a large area of background and it makes a big distraction. So it is difficult for us to do detection using the same method as the dental bitewing radiographs. What is more, their image processing method is based on the teeth with no defects, which seems to be impossible for patients who come to the hospital or people dying from traffic accident with broken teeth. Moreover, there are common cases including tooth loss, decayed tooth and filled tooth, which happen more frequently in real situation, so that single traditional method may not work in such complex cases. Fig. 1 shows some examples of complex dental periapical radiographs.

In last few years, deep learning technology especially the convolutional neural network (CNN) has been developed rapidly. The state-of-the-art methods such as Deep Convolutional Networks (Simonyan and Zisserman, 2014), fully convolutional networks (Shelhamer et al., 2017), Spatial Pyramid convolutional networks (He et al., 2015) and Resnet (He et al., 2015) have reached very high precision for image recognition. Deep learning approaches not only greatly reduce the workload of people, but can also extract better features that are difficult to recognize by human being. At present, in the field of image recognition, the deep learning algorithm reduced Top5 error rate from 26% to 15% through the construction of deep CNN, and further reduced it to 11% by increasing the complexity of the network structure (Krizhevsky et al., 2012). So far, the deep learning algorithm has almost reached the intelligent level of human brains. What is more, there have been some applications of deep learning methods in medical image field such as detection of myocardial infarction (Xu et al., 2017), brain tumor segmentation (Havaei et al., 2015) and lung nodule segmentation (Ronneberger et al., 2015; Wang et al., 2017). The authors get good results on their own task, which show the strong representation ability of neural networks.

To improve the robustness in teeth detection and classification, in this paper, we uses the label tree with deep learning techniques, especially the cascade network structure based on CNN to detect 32 classes of teeth position in total. The label tree is used to give each tooth several labels, which can decompose the task to some sub-tasks. We use a CNN to handle with each sub-task so that these single CNNs combine to a cascade network structure. What is more, we use several strategies to improve the performance, which achieves the automatic detection and classification of teeth in dental periapical radiographs with improved efficiency and reduced mental burden.

The main contributions of our method are: (1) the use of robust deep learning method for teeth recognition, which can deal with many complex cases. To the best of our knowledge, no public research has used the deep neural network for this problem so far; (2) the novel design of the label tree, which divide the problems to several sub-problems, and it can deal with the lack of data in this problem; (3) the application of the cascade network structure, which has much better performance than the direct single network; (4) the assisted strategies aimed to the specific sub-problem, which help to improve the performance obviously.

The remainder of this paper is organized as follows. Section 2 will introduce our label tree with the cascade network structure, including several cascade networks and a logical refine module. Section 3 will provide the key strategies we use to improve the performance. The results and comparisons are provided in Section 4, which demonstrate that our method is more effective when the image data is lacked.

2. System framework

We use CNN in this task. For this teeth detection and classification task, there are 32 teeth position in total. A direct method is to train a CNN that is proved to have great performance in large image database, such as faster-RCNN (Ren et al., 2015) and R-FCN (Dai et al., 2016) with whole training data, which direct give each tooth candidate a label of 33 classes (32 teeth and the background) for an input image. However, it is difficult to collect enough medical training data and their labels to train such a large multi-class network, which leading to a poor performance in this task (Section 4 will show the experiment results).

A single detection network cannot make full use of the information on a small training set. We provide more information to the network and limit the network to a small task that it can deal with. For the teeth classification task, a tooth can be divided to a set of labels according to whether a tooth is an up tooth or a lower tooth, a left tooth or a right tooth, and the teeth position by the arrangement of teeth. We combine these labels to get a whole mark, such as

RUT8, RUT7, …, RUT2, RUT1, LUT1, LUT2, …, LUT7, LUT8

RDT8, RDT7, …, RDT2, RDT1, LDT1, LDT2, …, LDT7, LDT8

we can use a tree to express these labels, which contains the information we try to utilize in the detection and classification process, see in Fig. 2. Difference from Deng et al. (2011), Rehreuer et al. (1998), we build the label tree manually by analyzing the task rather than learning a label tree.

The label tree has several layers which the root node is a tooth candidate. Each layer gives the tooth candidate a label, and the bottom layer gives the detailed position of the tooth. Each tooth has a label set generated from the label tree, and each layer represents a component of the mark. In our label tree, the top layer divides into two branches according to whether a candidate is a tooth (T), the second layer divides the teeth into upper teeth (UT) and the lower teeth (DT), the third layer is used to judge the position of teeth (such as UT1 to UT8), and the bottom layer divides the teeth into the left teeth (LT) and the right teeth (RT), a group of labels can lead to a specific tooth. We design a cascade network structure according the label tree, see in Fig. 3. In fact, the label tree is a decomposition of the original task, it gives each tooth several labels and each label is related to a sub-task, which is easier for us to handle with and provide more information to the whole system.