Tooth Restoration and Dental Work Detection on Panoramic Dental Images via CNN

Anil Gurses
Dept. of Electrical and Electronics Engineering
Istanbul Medeniyet University
Istanbul, Turkey
anilgurses@ieee.org

Ayse Betul Oktay
Dept. of Computer Engineering
Istanbul Medeniyet University
Istanbul, Turkey
abetul.oktay@medeniyet.edu.tr

***Abstract*—Automatic detection of dental work and type of restorations plays an important role for human identification and creation of reports for dental treatment at clinics. In this study, we employed three state-of-the-art convolutional neural networks (CNNs), which are GoogleNet, DenseNet and ResNet, for classification of dental restorations. Implants, canal root treatments, amalgam and composite fillings, dental braces and unrestored teeth are the classes that are detected by the networks. The CNNs are validated on a dataset including 3013 tooth images. DenseNet has 94% accuracy which is the highest accuracy among three CNN architectures. Dental braces and implants are detected with more accuracy than other dental work.**

***Index Terms*—CNN, tooth restoration, dental work detection, deep learning**

Fig. 1: A sample dental panoramic X-Ray image.

I. Introduction

Dental radiography is crucial for clinical treatment, diagnosis and surgery since the dental images include a lot of information about teeth and other structures. Different imaging modalities, such as computed tomography (CT), ultrasound, Magnetic Resonance Imaging (MRI) and X-ray are used at dental clinics [1]. Dental X-ray images are commonly used for detecting dental cavities, root canal treatment, pathologies, endodontic and periodontal procedures, and implant assessments. In addition, for human identification, forensic odontologists employ dental X-ray images [2], [3].

There are two types of dental X-ray images: intra-oral and extra-oral [4]. The intra-oral dental images include the bitewing and periapical images. Bitewing images show the details of the upper and lower teeth at a particular mouth area where the periapical images show the whole tooth from root to crown.

Panoramic images are the examples of extra-oral images. Extra-oral images are taken by positioning the patient between X-Ray source and X-Ray film. Panoramic images are commonly used on the examination teeth structure based diseases. They include all teeth on the mouth with the bones. Entire mouth is captured in an image. These type of images are taken for the examination of the dentures, braces, extractions and, implants. An example for a panoramic X-Ray image is shown at Figure 1.

Recent developments at the computer vision and deep learning fields have pioneered novel researches on different perspectives at medical imaging [5], [6]. There are many automated dental imaging based studies for dental treatment and diagnosis. For example, teeth are detected with their boundaries and numerated in [7] by faster R-CNN network which is one of the successful deep learning networks that has been successfully used on different areas. Jader et al. [8] have employed mask R-CNN algorithm for segmenting tooth from the X-ray images. They segmented the teeth with Mask R-CNN with high precision. Hwang et al. [9] presented a fully automatic approach with generative adversarial networks for designing dental crowns. In another study, You et al. [10] have shown a dental plaque segmentation with deep learning. Dental plaque is one of the main cause of dental diseases such as the sensitivity of tooth gum, tooth loss and, etc. The dental caries are detected with CNN [1], [11].

Yeshua et al. [12] classified the dental restorations into nine different classes and three negative classes. They first segmented teeth with adaptive thresholding and classified dental restoration with 20 features. The method is tested on 63 dental panoramic images containing 316 dental restorations and has 93% accuracy. Automated detection and classification of the dental work is important for detection of oral pathologies and creating reports for human identification. The study of [12] is the only study with machine learning for detection of dental work. However, it is based on low level image processing algorithms with many parameters and tested on a small dataset.

In this study, we propose to employ the state-of-the-art CNN architectures for detection of dental work. The dental work types including root canal treatment, dental implant,

978-1-7281-8073-1/20/ $31.00 ©2020 IEEE

filling and braces are classified with CNNs. In addition, the teeth with no restoration are detected. Examples of each class are shown in Figure 2. We used three CNN architectures which are GoogleNet, DenseNet and ResNet and provide a comprehensive analysis of these CNNs for the task of dental work classification. The methods are tested and validated on a dataset containing 3013 images.

The proposed method has many advantages. First, the method does not need any parameters. The automated detection of dental work can be used for human identification [13]. Since manual reports are formed by experts with time consuming and labor intensive procedure, this automatic reports will decrease costs. In addition, automatic detection of dental work is very important clinical treatment and assessing previous dental treatment.

II. Method

There are many neural network architectures that have been developed for computer vision tasks from different fields. The CNN architectures have different performances for medical image analysis tasks. In this study, we evaluate three different CNN architectures which are DenseNet [14], GoogleNet [15] and ResNet [16] for dental work classification with fine-tuning.

A. ResNet

He et. al [16] introduced a residual block for neural networks in general. These blocks are making shortcuts between layers and carrying the previous features to the next layers by skipping a set of layers. ResNet has performed great achievements based on accuracy and training performance. In this study, ResNet architecture with 18 layers is used.

B. DenseNet

DenseNet is built on the ResNet architecture and it includes additional inputs from all preceding layers and transfer it to subsequent layers with dense blocks. In Resnet, features from the previous layers are passed into the next layers. Each layer on the ResNet network are feed by the sum of the 2 inputs. One from the current layer and the other one is the previous layer. However, every layer on DenseNet is feed by concatenation of all the previous layers which is also called as cascaded layers. These dense blocks has acquired better results at many applications.

C. GoogleNet

GoogleNet [15] is built on inception blocks. Inception blocks consist of parallel four convolution processes. After those processes, they are concatenated in depth with each other. GoogleNet has achieved great results as well as other networks. It's mainly used for object detection on standard images. The standard images pixel values mostly distributed equally between 0 and 255. The X-Ray images diverse from the standard images because of the distribution of the pixel values. Therefore, neural network performances could be more different than expected on the X-Ray images.

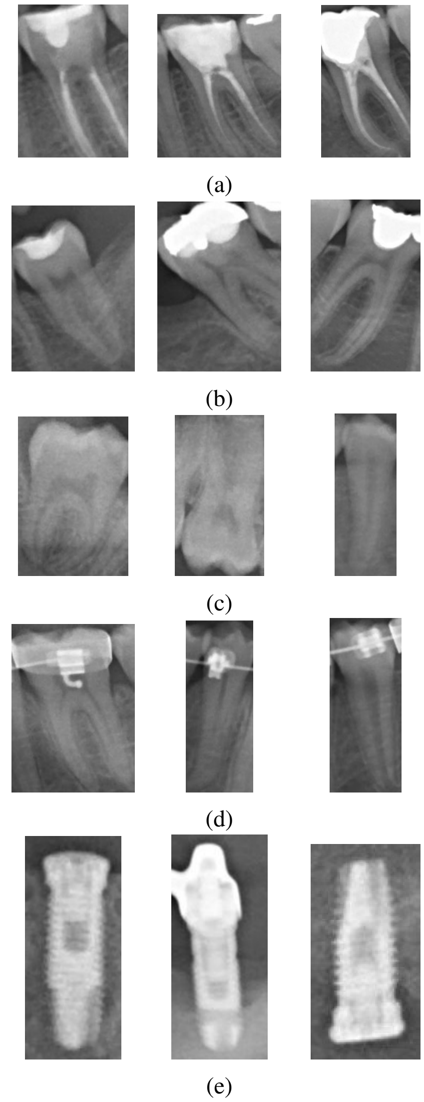
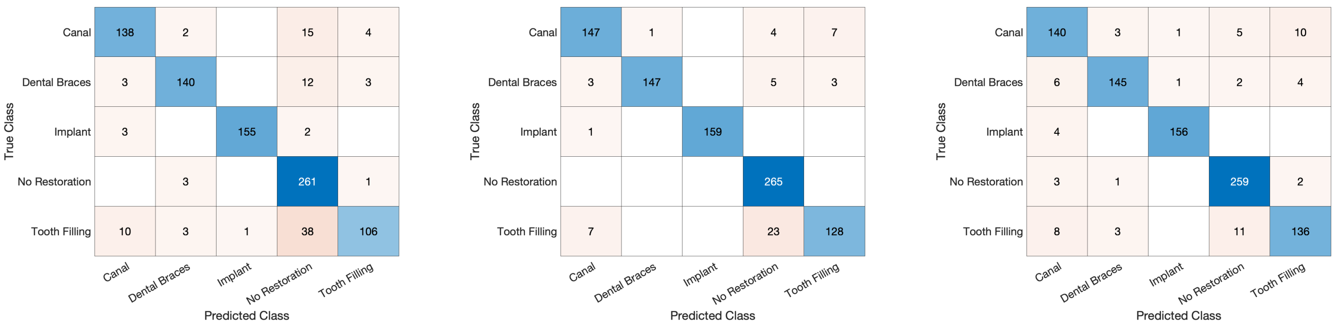


Fig. 2: Each row shows a restoration class: (a) canal root treatment, (b) filling, (c) no restoration, (d) dental braces and (e) implant.

D. Implementation Details

In this study, transfer learning is employed where the CNN architectures are pre-tranined networks on ImageNet dataset. Before the training, the input images are normalized and resized to 224×224 pixels in size. After these preliminary steps, inputs are fed to the CNNs and the training process took approximately five hours on a computer with an Nvidia Titan X GPU.

The implementation is done with the PyTorch [17] library. The pre-trained networks are trained on Imagenet dataset [18]. In this study, we used these trained models with fine tuning as mentioned before. Fine tuning is removing the fully connected layers, so fully connected layers are completely removed from the existing network and replaced with the new fully connected block. That fully connected block consists of a linear layer and softmax layer. The linear layer input and output is arranged



(a) GoogleNet-Confusion Matrix (b) DenseNet-Confusion Matrix (c) ResNet-Confusion Matrix

Fig. 3: Confusion matrices of the CNNs.

based on the last layer of the network. Also, the layers are frozen in the training stage for the purpose of maintaining the weights from ImageNet training.

In this study, four types of dental restorations are examined. These restorations are canal root treatment, implant treatment, dental braces and, tooth fillings. Also, tooth without any restoration is added as another class as no restoration for prediction. Optimization process is one of the key points in the training stage. We used stochastic gradient descent for optimizing the network. Weight decay is applied to optimization for avoiding overfitting. It penalizes the large weights that come from the network.

III. Dataset

The dataset provided by Jader et al. [19] is used for evaluation of the deep architectures. It is an open dental panoramic X-Ray dataset that contains many images with different types of dental work. An example of a panoramic dental X-Ray image is shown at Figure 1. For this study, the teeth images with different types of dental work images are marked by an expert. The single tooth images are cropped and its category is labeled. There are totally 3013 marked tooth images where 507 of them are implants, 544 of them have canal root treatment, 521 have amalgam or composite fillings, 515 of them have tooth braces, and 928 of them do not have any dental restoration.

2113 of the tooth images are used for the training the CNN architectures. The remaining 900 images are used for testing process of the CNNs. In addition, the training data is split into classes where each class have nearly same number of images. Each category has almost the same amount of images in the training set.

IV. Experiments

For training the CNNs, 2113 cropped tooth images are used. The rest of the images are used for testing of the methods. In order to evaluate the performance of the CNNs, precision, recall and accuracy metrics are used. Let T_p be the true positives, F_p be the false positives, T_n be true negatives and F_n be the false negatives of the classification results. The accuracy, precision and recall are calculated as follows:

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (1)$$

$$Precision = \frac{T_p}{T_p + F_p} \quad (2)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (3)$$

The performance precision and recall values of each dental restoration class are shown in Table I. DenseNet has higher precision and recall values -between 89% and 100%- than GoogleNet and ResNet. Dental implant class has the highest detection values with respect to other restorations. The recall and precision values of GoogleNet are lower than DenseNet and ResNet for most of the tooth restoration types. Implants and braces are classified with higher rates than other dental work with all CNNs. This may be caused by the low intra-class variance between implants.

The amalgam and composite fillings have different appearances and the number of samples in the training set do not seem enough for training of restorations. The highest precision and recall values are detected with the DenseNet. GoogleNet has lower precision and recall values than other two CNNs.

The accuracies of the CNN architectures are shown at Table II. DenseNet has 94% accuracy while ResNet has 92% accuracy. Since DenseNet is an extension of Resnet, DenseNet performs better than ResNet. The accuracy of GoogleNet is 89% and GoogleNet does not perform well as other CNNs.

The confusion matrices of each CNN are shown in Figure 3. GoogleNet detects 155 of the implant category with high accuracy. GoogleNet has failed on tooth filling classification and 38 teeth with filling are classified as no restoration class. This is also common thing at other CNNs.

DenseNet detects most of the restoration types with high accuracy except fillings. 23 teeth with fillings are classified as non-restoration. DenseNet classifies all of the braces accurately and 147 of 148 braces.

ResNet has lower accuracy than DenseNet at all classes. It performs better than Densenet at all classes except implants.

TABLE I: Precision and Recall Scores

Restoration Type	DenseNet		GoogleNet		ResNet	
	Precision	Recall	Precision	Recall	Precision	Recall
No Restoration	89%	100%	80%	98%	94%	98%
Canal Root Treatment	93%	92%	90%	87%	87%	88%
Fillings	93%	92%	93%	67%	89%	86%
Implant	100%	99%	99%	97%	99%	97%
Dental Brace	99%	93%	95%	89%	95%	92%

TABLE II: ACCURACY SCORES

	GoogleNet	DenseNet	ResNet
Accuracy	89%	94%	93%

The quantitative results show that the dental restoration types and unrestored teeth can be classified with high accuracy with CNNs. The special structures like implant and dental braces have the highest accuracy. The fillings class has the lowest accuracy at all CNNs since most of the fillings are very small and the pixel intensity of the filling is low especially at composite fillings.

V. CONCLUSIONS

In this study, we used state-of-the-art CNN architectures for detection of dental restorations. The CNN architectures pretrained on ImageNet are trained on a large dataset for classification of four dental restoration types and unrestored teeth. The tooth braces, implants, canal root treatment and fillings are the dental work classes and no restoration is another class.

The CNNs are tested on 900 tooth images and DenseNet has performed with 94% accuracy the among the all three CNNS. ResNet has 93% accuracy while GoogleNet has 89% accuracy. DenseNet is determined to work well for dental work detection.

Implants are detected higher accuracy with all CNNs while the fillings are detected with lowest accuracy. The generated dental work information is very valuable for different type of tasks. In the future work, we are planning to perform human identification with the generated dental work reports of the panoramic dental images.

VI. ACKNOWLEDGMENT

The authors thank to Sinem İlda Birdal for manually marking the dental work for ground truth.

REFERENCES

[1] C.-W. Wang, C.-T. Huang, J.-H. Lee, C.-H. Li, S.-W. Chang, M.-J. Siao, T.-M. Lai, B. Ibragimov, T. Vrtovec, O. Ronneberger, P. Fischer, T. F. Cootes, and C. Lindner, "A benchmark for comparison of dental radiography analysis algorithms," *Medical Image Analysis*, vol. 31, pp. 63 – 76, 2016.

[2] A. B. Oktay, "Human identification with dental panoramic images," *IET Biometrics*, vol. 7, no. 4, pp. 349–355, 2017.

[3] X. Zhong, D. Yu, Y. S. Wong, T. Sim, W. F. Lu, K. W. C. Foong, and H.-L. Cheng, "3d dental biometrics: Alignment and matching of dental casts for human identification," *Computers in Industry*, vol. 64, no. 9, pp. 1355 – 1370, 2013, special Issue: 3D Imaging in Industry.

[4] R. Kumar, N. Khambete, and E. Priya, "Extraoral periapical radiography: An alternative approach to intraoral periapical radiography," *Imaging science in dentistry*, vol. 41, pp. 161–5, 12 2011.

[5] J.-J. Hwang, Y.-H. Jung, B.-H. Cho, and M.-S. Heo, "An overview of deep learning in the field of dentistry," *Imaging Science in Dentistry*, vol. 49, p. 1, 03 2019.

[6] A. Betul Oktay, "Tooth detection with convolutional neural networks," in *2017 Medical Technologies National Congress (TIPTEKNO)*, 2017, pp. 1–4.

[7] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific Reports*, vol. 9, 12 2019.

[8] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic x-ray images," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 400–407.

[9] J.-J. Hwang, S. Azernikov, A. A. Efros, and S. X. Yu, "Learning beyond human expertise with generative models for dental restorations," 2018.

[10] W. You, A. Hao, S. Li, Y. Wang, and B. Xia, "Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments," 03 2020.

[11] J. Lee, D. Kim, S. Jeong, and S. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of Dentistry*, vol. 77, pp. 106–111, Oct. 2018.

[12] T. Yeshua, Y. Mandelbaum, R. Abdalla-Aslan, C. Nadler, L. Cohen, L. Zemour, D. Kabla, O. Gleisner, and I. Leichter, "Automatic detection and classification of dental restorations in panoramic radiographs," vol. 16, pp. 221–234, 05 2019.

[13] L. Lorton, M. Rethman, and R. Friedman, "Computer-assisted post-mortem identification (capmi) system: A computer-based identification program," *Journal of Forensic Sciences*, vol. 33, pp. 977–984, 1988.

[14] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016. [Online]. Available: http://arxiv.org/abs/1608.06993

[15] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[17] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[19] G. Jader, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives," *Expert Systems with Applications*, vol. 107, 02 2018.