Master Thesis

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Computer Science

Dental caries detection from bitewing X-ray images

Lukáš Kunt

Supervisor: prof. Dr. Ing. Jan Kybic
Field of study: Open informatics
Subfield: Artificial inteligence
May 2022

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Kunt** Jméno: **Lukáš** Osobní číslo: **478073**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra počítačů**

Studijní program: **Otevřená informatika**

Specializace: **Umělá inteligence**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Detekce zubních kazů z rentgenových snímků

Název diplomové práce anglicky:

Dental caries detection from bitewing X-ray images

Pokyny pro vypracování:

1) Familiarize yourself with state-of-the-art methods in object detection and computer-assisted dental caries detection.
2) Based on the studied literature, propose a solution to dental caries detection from X-ray images based on deep learning.
3) Implement a program that automatically detects caries from bitewing X-ray images.

Seznam doporučené literatury:

[1] Srivastava et al: Detection of Tooth caries in Bitewing Radiographs using Deep .Learning. NIPS 2017 workshop on Machine Learning for Health.
[2] Jae-Hong Lee et al: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. Journal of Dentistry 2018
[3] Moran et al: Classification of Approximal Caries in Bitewing Radiographs Using Convolutional Neural Networks. Sensors 2021
[4] Yusuf Bayraktar, Enes Ayan: Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. Clinical Oral Investigation, 2021.

Jméno a pracoviště vedoucí(ho) diplomové práce:

prof. Dr. Ing. Jan Kybic algoritmy pro biomedicínské zobrazování FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **23.02.2022** Termín odevzdání diplomové práce: **20.05.2022**

Platnost zadání diplomové práce: **19.02.2024**

prof. Dr. Ing. Jan Kybic
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

.
Datum převzetí zadání

Podpis studenta

© ČVUT v Praze, Design: ČVUT v Praze, VIC

Acknowledgements

Firstly, I would like to express gratitude to prof. Jan Kybic for supervision of this thesis and his willingness to help at any time.

Secondly, I would like to thank MDDr. Tichý was always eager to help, and without his dedication to creating the dataset, we would not have achieved the results we did.

Last but not least, I would like to thank my family for their support during my studies.

I would like to further emphasize my gratitude to my girlfriend Anna for her undying moral and encouragement throughout the creation of this thesis.

Declaration

I hereby declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the Methodi- cal instructions for observing the ethical principles in the preparation of university thesis.

Prague, May 20, 2022

Abstract

Dental caries is the most prevalent disease globally, with more than 3.5 billion people affected. The treatment of dental caries imposes a burden on health care in every country financially and timewise. Detection of the disease in its early stages can mitigate the impact on the cost of treatment and improve the patient's prognosis.

Bitewing X-ray imaging is the second most used method for dental caries detection after the visual-tactile method. Aproximal and an-early stage carious lesion can be easily overlooked by the visual-tactile exam, making the bitewing X-ray imaging very beneficial for early detection and a chance for recovery without the need for further dental treatment.

This Master's thesis addresses the problem of dental caries detection from bitewing images using convolutional neural networks. First, a dataset of 3889 bitewing images with 7257 annotated dental caries was created for the purpose of this thesis. We trained multiple architectures for object detection and compared their performance using it. In the end, we used an ensemble of models to obtain the best-performing model.

We have created a solution that can detect dental caries with a precision of 0.751 and a recall of 0.7. Furthermore, a second model for segmentation of dental restoration was created, achieving an IOU score of 0.676.

Keywords: dental caries detection, convolutional neural networks, dental restorations segmentation, ensemble, bitewing, X-ray image

Supervisor: prof. Dr. Ing. Jan Kybic
Biomedical imaging algorithms, FEE

Abstrakt

Zubní kaz je jedním z nejrozšířenějších onemocnění na světě postihující více než 3.5 miliard lidí. Léčba je náročná jak finančně, tak časově a zatěžuje zdravotnický systém ve všech zemích světa. Včasná detekce zubního kazu umožňuje tuto zátěž snížit a zlepšit pacientovu prognózu.

Po detekci pohledem spojené s použitím zubařské sondy jsou bitewingové rentgenové snímky druhou nejvíce využívanou metodou pro diagnostiku zubního kazu. Časné a aproximální kariézní léze nejsou prvně zmíněnou metodou vždy spolehlivě diagnostikovány, což dává bitewingovým RTG snímkům značnou výhodu a šanci pro dřívější diagnostiku spojenou s možností zhojení léze bez nutnosti dalšího lékařského zásahu.

Tato diplomová práce se zabývá problémem detekce zubního kazu z bitewingových RTG snímků za použití konvolučních neuronových sítí. Pro účely této práce byl nejprve vytvořen dataset skládající se z 3889 bitewingových RTG snímků se 7257 anotovanými zubními kazy. Za jeho použití jsme natrénovali několik architektur pro detekci objektů a porovnali jejich výsledky. Nakonec jsme využili spojení modelů pro získání modelu s nejlepšími výsledky.

Vytvořili jsme řešení, které umožňuje detekci kazů s přesností 0.751 a citlivostí 0.7. Navíc byl vytvořen i druhý model pro segmentaci zubních výplní, který dosáhl IOU 0.676.

Klíčová slova: detekce zubních kazů, konvoluční neuronové sítě, segmentace zubních výplní, ensemble, bitewing, rentgenový snímek

Překlad názvu: Detekce zubních kazů z rentgenových snímků

Contents

1 Introduction 1

2 Medical background 3
2.1 Human teeth 3
2.1.1 Structure of teeth 4
Pulp-Dentin complex 4
Cementum 5
2.2 Dental caries 5
2.2.1 Cause 5
2.2.2 Epidemiology 6
2.2.3 Diagnosis 6
2.2.4 Treatment 9

3 Theoretical background 11
3.1 Computer vision tasks 11
3.1.1 Classification 11
3.1.2 Semantic segmentation . . . 11
3.1.3 Object detection 11
3.1.4 Instance segmentation 12
3.2 Data format in object detection 12
3.2.1 PASCAL VOC 12
3.2.2 COCO 12
3.2.3 YOLO 13
3.3 Metrics 13
3.3.1 Intersection over union (IOU) 13
3.3.2 Precision and recall 13
3.3.3 Mean average recall in MS-COCO (mAR) 16
3.4 Optimization 17
3.4.1 Optimizers 17
3.4.2 Weight decay 17
3.4.3 Learning rate schedulers . . 17
3.5 Artificial neural network (ANN) 18
3.5.1 Convolutional layer 19
3.5.2 Activation functions 19
3.5.3 Normalization layers 19
3.6 Transformer architecture 20
3.6.1 Transformers in computer vision 21
3.7 General architecture for object detection 21
3.8 Backbone models 23
3.8.1 ResNet 23
3.8.2 EfficientNet 24
3.8.3 Swin transformer 25
3.9 Detection models 25
3.9.1 Faster R-CNN 25
3.9.2 RetinaNet 26
3.9.3 EfficientDet 26
3.9.4 Models for image segmentation 27
3.10 Model ensembling in object detection 28

4 Related Work 31
4.1 Dental caries detection 31
4.1.1 Manual detection and classification 31
4.1.2 Dental caries segmentation 32
4.1.3 Dental caries detection . . . 33
4.2 Dental restorations segmentation 35

5 Dataset 37
5.1 Dental caries 37
5.2 Dental restorations 41

6 Project structure 43
6.1 Organization of the project . . . 43
6.1.1 Models 43
6.1.2 Modules 44
6.1.3 Transformations 44
6.1.4 Data-Modules 44
6.1.5 Trainer 45
6.1.6 Callbacks 46
6.1.7 Logging 46
6.2 Additional open-source software 46

7 Methods 47
7.1 Caries detection - baseline model comparison 47
7.1.1 Dataset 47
7.1.2 Image augmentations 47
7.1.3 Neural network models . . . 48
7.2 Improvements 50
7.2.1 Training protocol changes . 50
7.2.2 Group normalization 50
7.3 Model inspection 50
7.3.1 Size of backbone 50
7.3.2 Weight decay 51
7.4 Model ensembling 51

7.4.1 Data-format 51
7.4.2 Area-aware ensembling . . . 52
7.4.3 Assessing the importance of different models in ensembling 53
7.5 Dental restorations segmentation 54
7.5.1 Non-deep learning approach 54
7.5.2 Deep-learning approach . . . 55
7.5.3 Model training improvements 56

8 Results 57
8.1 Model comparison on different datasets 57
8.1.1 Stage one dataset 57
8.1.2 Stage two dataset 58
8.1.3 Stage three dataset 58
8.1.4 Stage four dataset 59
8.1.5 Stage five 59
8.2 Improvements 60
8.2.1 Training protocol improvements 60
8.2.2 Group normalization 60
8.3 Model inspection 61
8.3.1 Size of backbone 61
8.3.2 Weight decay 61
8.4 Ensembling 62
8.4.1 Manually-picked parameters 62
8.4.2 Grid search results 62
8.4.3 Assessing the importance of different models 63
8.5 Dental restorations segmentation 64
8.5.1 Non-deep learning approach 64
8.5.2 U-Net 66
8.6 Comparison of results with related publications 68
8.7 Visualization of models 68

9 Discussion and further suggestions 71
9.1 Comments on model comparison and their improvements 71
9.2 Ensembling 71
9.3 Dental restorations segmentation 72
9.3.1 Non-deep learning approach 72
9.3.2 U-Net 72
9.4 Comparison of results with related publications 72

10 Conclusion 73
Further work 73

A Bibliography 75

B Aditional results 83
B.1 Detection of detnal caries and segmentation of dental restorations 85

C Images 89
C.1 Predictions of the model 89
C.2 Model importance during ensembling 92
C.3 Augmented images 93

Figures

2.1 Structure of teeth 4
2.2 Bitewing X-ray image 7
2.3 Periapical X-ray image, source [1] 7
2.4 Panoramic X-ray image, source [2] 8

3.1 From the left: semantic segmentation, object localization, object detection, instance segmentation 12
3.2 Examples of IOUs for different overlaps between GT and predicted box, source. [3] 14
3.3 Standard and interpolated precision-recall curve, source [4] 16
3.4 Learning rate schedulers: Cosine annealing is red, Reduce learning rate on plateau has purple color . 18
3.5 Graphs of ReLU based activation functions, source [5] 20
3.6 Batch and group normalization layers with denoted axes, across, which the normalization statistics is computed 20
3.7 Architecture of transformer with two encoders and two decoders, source [6] 22
3.8 Architecture of ViT, source [7] 23
3.9 Architecture of different necks for feature fusion, source [8] 23
3.10 The schema of two-stage detection process of Faster R-CNN 24
3.11 General architecture for object detection, source [9] 24
3.12 Comparison of loss landscapes, source [10] 25
3.13 Hierachical structure of Swin Transformer compared with ViT, source [11] 26
3.14 Shift of local window for computation of self-attention, source [11] 26
3.15 Focal Loss, source [12] 26
3.16 Architecture of U-Net model, source [13] 27
3.17 Pseud code of NMS and soft-NMS, source [14] 28

4.1 Annotated bitewing radiograph and the same image post-processed, source [15] 33
4.2 Sample of data and predictions of the model by 34
4.3 Predictions of the model proposed by Srivastava et al., source [16] 35
4.4 Results of segmentation algorithm proposed by Yeshua et al.[17] 36
4.5 Cropped region from panoramatic image with multiple restorations, source [18] 36

5.1 The environment of CVAT with annotated carious lesion and dental restorations 38
5.2 Histogram of bonding boxes dimensions in the dataset 39
5.3 Histogram of the number of dental caries per image 40
5.4 Bitewing X-ray image on the left, pixel mask of the X-ray on the right (dental restorations have yellow color) 41
5.5 Histogram of restoration area in image, images without restorations omitted 42
5.6 Histogram of areas of restorations, 10 largest omitted . 42

6.1 Structure of modules and their configuration files 44
6.2 Folder structure of the project 45

7.1 Structure of the data in .json file used to store model predictions . 52

7.2 Schema of the search of hyperparameters and weights for ensembling 52

8.1 Difference in AP@.5 amog EfficientDet-D4 model with batch normalization and group normalization 61
8.2 AP@.5 of Faster-RCNN model with varying weight decay values. The metric is computed on the validation part of the dataset during training. 62
8.3 AP@.5 of YOLOv5 model with varying weight decay values. The metric is computed on the validation part of the dataset during training. 62
8.4 Importance of different models during ensembling with different architectures 65
8.5 From the left: X-ray image, ground-truth pixel mask, thresholded image, removal of border pixels, morphological opening 65
8.6 From the left: X-ray image, ground-truth pixel mask, the output of the model, output processed by morphological opening, output post-processed by morphological opening and closing. 66
8.7 IOU throughout the training of U-Net model for different loss functions 67
8.8 Value of IOU metric based on the size of kernels K_o, K_c in morphological operations that were used for post-processing 67
8.9 Number of false positives per image for a given value of recall 69
8.10 Percentage of nondetected dental caries based on the precision of the model 69
8.11 Segmented dental restorations in yellow, predicted dental caries in pink and ground truth of dental caries in green. We see a single false positive detection on the top right of the image. The author of the dataset acknowledges it to be a missing ground truth label 70

B.1 Four correct detection of dental caries and one falsely-positive . . 86
B.2 Segmented dental restorations in yellow, predicted dental caries are pink. We see a successful segmentation of dental restorations and a single falsely positive prediction of dental caries 87

C.1 X-ray image with ground truth boxes and model's predictions. Tho model correctly predicts 11 out of 12 dental caries in the image. . . . 89
C.2 The model predicts correctly five dental caries and has one falsely negative prediction 90
C.3 Prediction of the model in image with extensive amount of dental restorations 91
C.4 The model predictions on image with dental bridge 92
C.5 Importance of different models during ensembling with the same architectures and backbones . . . 92
C.7 No transformation applied . . . 93
C.6 Importance of different models during ensembling with the architecture and different backbones 93
C.8 Gaussian blur applied 94
C.9 Rotation applied 94
C.10 Gamma correction applied . . 95
C.11 Translation applied 95
C.12 The whole augmentation pipeline applied 96

Tables

4.1 Results of models proposed in [16] and [19], compared with three dentists, modified 34

5.1 Statistics of bounding boxes that denote position of carious lesions 40

7.1 Maximal batch sizes that fit into 12GB GPU for a given model 49
7.2 Hyperparameter search-space for area-aware model ensembling . . 53
7.3 Hyperparameter search-space for model ensembling 53
7.4 Comparison of the models involved in ensembling by statics of their $AP@.5$ metrics 54
7.5 Hyperparameter search space for restorations segmentation pipeline 55

8.1 Comparison of the trained models on the stage one dataset 58
8.2 Comparison of the trained model on the stage two dataset 58
8.3 Comparison of trained models on the test part of stage three dataset 59
8.4 Performance comparison of multiple models trained on the stage four dataset 59
8.5 Performance comparison of multiple models based on mean average precision metrics 60
8.6 Comparison of AP values between different models trained by an improved training protocol 60
8.7 Comparison of $AP:@5$metric for different backbones of YOLOv5 architecture 61
8.8 WBF ensembling of multiple models, where we handpicked the parameters of the ensembling process. The models were trained on the stage four dataset. 62
8.9 Hyperparameter values of ensembling methods found by a grid-search 63
8.10 Average precision of models ensembled with parameters from Table 8.9 63
8.11 Average precision of ensemble models 64
8.12 Average recall of models ensembled by parameters from Table 8.9 64
8.13 Precision, recall, and F-score based on the confidence threshold for different ensembling methods 64
8.14 Best hyper-parameters for non-deep learning pipeline 65
8.15 Results of non-deep learning approach to dental restorations segmentation given the hyperparameters in Table 8.14 . . 66
8.16 Optimal parameters for model post-processing found by a grid-search 67
8.17 Results of U-Net models 67
8.18 Comparison of results of this thesis with results in related publications 68

B.1 Comparision of trained models on the train part of stage three dataset 83
B.2 Precision, recall, and F-score based on the confidence threshold for the models trained on stage four datset 83
B.3 Precision, recall, and F-score based on the confidence threshold. Models were trained on stage-five dataset 83
B.4 Average recall of models trained by improved training protocol . . . 84
B.5 Precision, recall, and F-score based on the confidence threshold for different models, trained by the imporved training protocol 84

B.6 Average recall of models ensembled by parameters from Table 8.9 84
B.7 Precision, recall, and F-score based on the confidence threshold for different ensembling methods 84

Chapter 1

Introduction

As of 2017, dental caries was the most prevalent disease globally[20][21], with more than 3.5 billion affected people. Despite the advancement of technology in the medical field, the prevalence has not decreased, hence imposing a burden on health care in every country. In the US, more than six percent of total health care expenditures were targeted toward dental care in 2016[22].
Machine learning and especially neural networks have improved remarkably over the last decade, surpassing human-level performance in the ImageNet classification task in 2015[23]. Since then, deep learning models' error rates on the ImageNet dataset have become four times smaller[24]. This significant improvement in deep learning led to its wast adoption across many fields, including medical imaging. Deep-learning models exceeded human-level performance on specific tasks, such as breast cancer detection[25] or arrhythmia detection[26].
This thesis aims to develop a deep-learning-based model that will be able to detect dental caries in bitewing X-ray images. The position of every carious lesion is denoted by a minimal bounding box around the lesion. This model aims to give dentists an opportunity to cross-check their decision regarding the presence of carious lesions in the X-ray image. Furthermore, detecting caries' position directly from the image benefits dentists from saving information about dental caries in a digital form without their intervention. Having the data in a digital format could help dentists communicate the problem to a patient by overlaying the position of dental caries over an X-ray image or provide them an option to save that information for monitoring lesion progression over time. Last but not least, software like this would be helpful in education, where dentistry students would be able to train their ability to recognize dental caries without the need for a tutor.

The structure of this thesis is as follows: In Chapter 2, the medical background is introduced, describing human dentition and dental caries. Chapter 3 presents the fundamentals of the techniques used in this work. In Chapter 4, related work in automatic caries detection is discussed. Chapter 5 describes the dataset that was created for this thesis. In Chapter 6, we introduce the reader and the structure of the object detection framework that we created and used for caries detection. In Chapter 7, we propose a solution to the detection of dental caries. The results achieved by the proposed methods are presented in Chapter 8. Chapter 9 discusses the results and methods that we used to obtain them. In Chapter 10, we summarise the work and suggest future improvements.

Chapter 2

Medical background

2.1 Human teeth

Human dentition is composed of two sets of teeth - primary and permanent. The primary, also called deciduous, consists of 20 teeth and begins to erupt at six months of age. This dentition is completely replaced at the approximate age of 13 years by a permanent set of teeth, including 32 teeth. These can be divided into four classes based on function and form. Namely, those classes are:

Incisors

A total of 8 incisors teeth are found in primary and permanent dentition. They are located at the oral cavity entrance, and their primary purpose is to cut and shear food. They are essential for a smile's esthetics and play a vital role in phonetics.

Canines

A total number of four canines are located at the corners of dental arches, dividing them into a frontal and a lateral part. They have a triangular shape with a single cusp tip on the incisal edge. The structure is associated with their ability to seize, pierce, tear and cut food. Along with the incisors, they are essential for esthetics.

Premolars

Premolars are teeth found only in permanent dentition, being the successional teeth of all primary eight molars. Premolars share functional characteristics of canines and molars - they both seize and grind food thanks to their anatomy.

Molars

Human dentition contains 12 permanent molars with no deciduous predecessors. Their leading role is crushing and grinding food to dimensions appropriate for swallowing. Broad occlusal surfaces make them capable of this task. Molars are prone to dental caries due to deep grooves that run across the occlusal surface of