Zurich University
of Applied Sciences

School of
Engineering

IDP Institute of Data Analysis
and Process Design

Automated scoring of bone erosion for rheumatoid arthritis with deep neural networks

Zurich University of Applied Sciences

Author: Janick Rohrbach
Supervisor: Dr. Oliver Dürr
Prof. Dr. Beate Sick
Industrial Partner: Seantis GmbH
External Supervisor: Fabian Reinhard
Dr. Tobias Reinhard
Date: December 22, 2017

Declaration of originality
Project Thesis at the School of Engineering

By submitting this project thesis, the undersigned student confirms that this thesis is his own work and was written without the help of a third party.

The student declares that all sources in the text (including Internet pages) and appendices have been correctly disclosed. This means that there has been no plagiarism, i.e. no sections of the project thesis have been partially or wholly taken from other texts and represented as the student's own work or included without being correctly referenced.

Any misconduct will be dealt with according to paragraphs 39 and 40 of the General Academic Regulations for Bachelor's and Master's Degree courses at the Zurich University of Applied Sciences (Rahmenprüfungsordnung ZHAW (RPO)) and subject to the provisions for disciplinary action stipulated in the University regulations.

Zurich, December 22, 2017 Janick Rohrbach

Abstract

Abstract goes here

Acknowledgements

I would like to express my sincere thanks to my supervisors Beate Sick and Oliver Dürr who provided me with guidance and support during the writing of this thesis.

I would also like to thank Fabian Reinhard and Tobias Reinhard (Seantis GmbH) for their valuable inputs.

Further, I want to acknowledge the SCQM foundation which made this thesis possible by providing the comprehensive dataset. A list of rheumatology offices and hospitals that are contributing to the SCQM registries can be found on www.scqm.ch/institutions. The SCQM is financially supported by pharmaceutical industries and donors. A list of financial supporters can be found on www.scqm.ch/sponsors.

Contents

1 Introduction

This thesis shows a method for the automated scoring of x-ray images of patients with rheumatoid arthritis.

1.1 Background

Rheumatoid arthritis is caused by a malfunctioning immune system. It is therefore a type of autoimmune diseases. The immune system attacks healthy tissue instead of bacteria and viruses. This causes inflammation in the joints. Irreversible damage to the bone in the joint can occur, if the inflammation lasts for a long time. [1] Rheumatoid arthritis is incurable, merely the symptoms can be treated.

Today, the severity of the bone erosion is assessed by a trained rheumatologist by using x-ray images of hand and feet. This process takes several minutes per patient. This thesis shows, how recent advances in computer vision make it possible to automate this task. This leads to time savings which in return helps the rheumatologist to spend more time with the patient.

The Swiss Clinical Quality Management in Rheumatic Diseases (SCQM) Foundation runs a national registry of inflammatory rheumatic diseases. [2] They have collected anonymized patient data for over 10 years and provide us with x-ray images for this analysis.

Seantis GmbH is a Swiss company that develops data driven web applications for medical research, pubic administration and aviation. [3] For their customer SCQM they want to automate the bone erosion assessment. They already have a working algorithm, which detects the body part shown in the x-ray image. A second algorithm detects the joints in the image and extracts them as single images. These images are then used together with the bone erosion scores to train our model.

1.2 Related literature

There are several applications where convolutional neural networks are used in medical research.

A recent paper from Tajbakhsh et al. [4] investigated whether fine-tuning a pre-trained CNN is better than training a CNN from scratch when applied to medical images. They find that pre-trained networks with fine-tuning always outperformed or at least performed as well as CNNs trained from scratch. They further recommend a layer-wise fine tuning which seems to outperform shallow and deep tuning.

A study by Paul et al. [5] tried to classify osteoporosis by considering x-ray images of the bone. This task proved to be very difficult as the x-ray images from healthy patients look very similar to the ones of patients with the disease. By using a transfer learning approach they achieved a validation accuracy of 44.82 %.

Zhou et al. [6] used a two-level ensemble of neural networks to identify lung cancer cells on x-ray images of the chest. The first-level ensemble classifies whether a cell is a cancer cell or not by using full voting. The second-level ensemble is used only on cells classified by the first-level as cancer cells. It differentiates between different cancer classes as well as a non-cancer class. This ensemble works with plurality voting. The authors state that this method achieves a high accuracy and a low rate of false negatives.

A report from Chen [7] showed the application of convolutional neural networks on x-ray images of hands to predict the developmental bone age. He achieves a top one and two accuracy of 46 % and 70 % respectively. This result is close to previously used methods which use manual segmentation and handcrafted features.

In a degree project Hensman and Masko [8] looked at the impact of imbalanced training data for CNNs. They find, that heavy imbalances have a strong impact on the performance and suggest oversampling of minority classes to improve the performance of the network.

1.3 Aim and scope of this thesis

The aim of this thesis is to predict bone erosion scores from x-ray images. We further examine how the bone erosion and the disease activity are correlated and use a time series of images to predict the course of the disease.

The work is based on images of the left hand only. There exist also images of right hands as well as left and right feet. But at this point in time, only the joints of left hands have been extracted from the images. We assume that the model will perform similar on the joints of the right hand. By fine-tuning the model on the images of joints from feet it should also perform well for those images.

1.4 Outline

section 2 provides...
section 3 describes..

2 Theory

2.1 Bone erosion scores

Figure 1 shows a sample x-ray image of two hands similar to the images received from the SCQM foundation. The five proximal interphalangeal (PIP) joints and the 5 carpometacarpal (MCP) joints per hand are shown with blue bounding boxes. A trained rheumatologist has scored each of those 10 joints per hand.

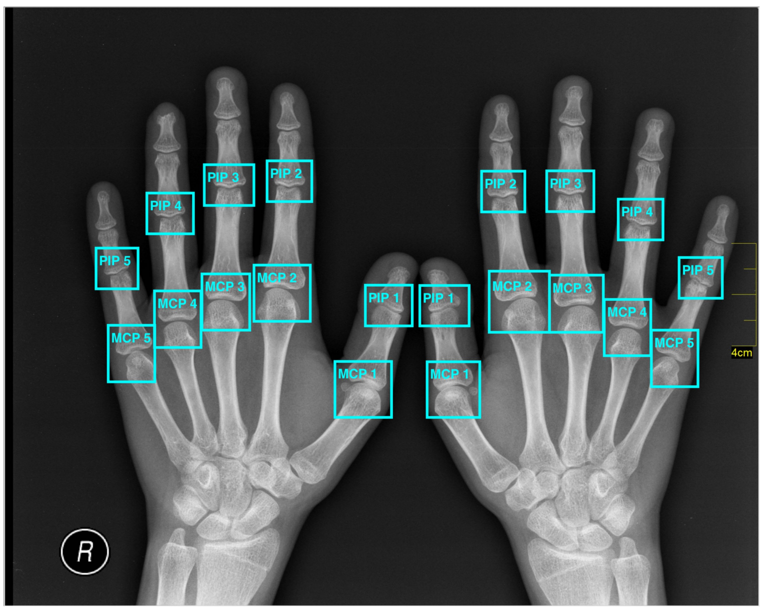

Figure 1: Proximal interphalangeal (PIP) joints and carpometacarpal (MCP) joints.

Original image by Nevit Dilmen (CC BY-SA) `https://commons.wikimedia.org/wiki/File:Medical_X\discretionary{-}{}{}Ray_imaging_OPC06_nevit.jpg`

2.1.1 Ratingen score

Stadium 0	=	normal joint
Stadium 1	=	one or more erosions, less than 20 % of the joint surface is eroded
Stadium 2	=	21 % - 40 % of the joint surface is eroded
Stadium 3	=	41 % - 60 % of the joint surface is eroded
Stadium 4	=	61 % - 80 % of the joint surface is eroded
Stadium 5	=	more than 80 % of the joint surface is eroded

2.2 Convolutional neural networks

Convolutional neural networks take an image as an input. The image then gets passed through several convolutional layers. These layers work as filters and detect different features in the image. The weights of these layers are combined to class scores. Andrey Karpathy provides a good overview over convolutional neural networks in his course notes for the Stanford class CS231n. [9]

3 Methods

Keras

4 Predicting Rau scores

4.1 Pre-processing

The pre-processing step formats the images into a suitable format that can be used as an input for the CNN.

The images of the joints have varying exposure. Some images are very dark while others are very bright. It was therefore considered to apply a histogram equalization, which is a linear transformation that maps the lightest pixel to 255 and the darkest pixel to 1. However, this transformation did not improve the accuracy of the model and was not used for the final model.

The Rau scores of the labeled joints are highly imbalanced. As seen in Figure 2 most of the joints are healthy and received a score of 0. There are quite a few observations with scores between 0 and 25 but there are very little observations with scores higher than 25. Because the CNN minimizes

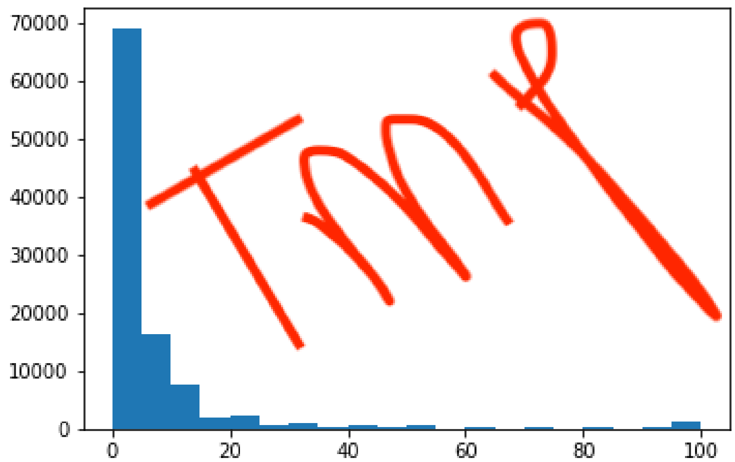

Figure 2: Distribution of the Rau scores

the overall loss-function, it would perform bad for the underrepresented part of the data. In this case, the model would be bad in predicting high scores. In order to make the model a good predictor for all cases, we introduced ...????????????

4.1.1 CNN architecture

Convolution
BatchNorm
Relu
Convolution
BatchNorm
Relu
MaxPooling

6x
2 Dense layers
Output layer Softmax

4.2 Classification

4.3 Regression

5 Results

6 Discussion

7 Conclusion

References

[1] American College of Rheumatology. *Rheumatoid Arthritis*. [Online; accessed 26-September-2017]. URL: https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Rheumatoid-Arthritis.

[2] SCQM. *About Us*. [Online; accessed 11-October-2017]. URL: https://www.scqm.ch/en/ueber-uns/.

[3] Seantis GmbH. *Data Driven Web Applications*. [Online; accessed 12-October-2017]. URL: https://www.seantis.ch.

[4] Nima Tajbakhsh et al. "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1299–1312. DOI: https://doi.org/10.1109/TMI.2016.2535302.

[5] Rahul Paul et al. *Make Your Bone Great Again : A study on Osteoporosis Classification*. University of South Florida, 2017. URL: https://arxiv.org/abs/1707.05385.

[6] Zhi-Hua Zhou et al. "Lung cancer cell identification based on artificial neural network ensembles". In: *Artificial Intelligence in Medicine* 24.1 (2002), pp. 25–36. DOI: https://doi.org/10.1016/S0933-3657(01)00094-X.

[7] Matthew Chen. *Automated Bone Age Classification with Deep Neural Networks*. Stanford University, 2016. URL: http://cs231n.stanford.edu/reports/2016/pdfs/310_Report.pdf.

[8] Paulina Hensman and David Masko. *The Impact of Imbalanced Training Data for Convolutional Neural Networks*. KTH Royal Institute of Technology, 2015. URL: https://www.kth.se/social/files/588617ebf2765401cfcc478c/PHensmanDMasko_dkand15.pdf.

[9] Stanford University Andrej Karpathy. *CS231n Convolutional Neural Networks for Visual Recognition*. [Online; accessed 26-September-2017]. URL: http://cs231n.github.io/convolutional-networks/.

List of Figures