

Age Estimation on Panoramic Dental X-ray Images Using Deep Learning

Sarah Wallraff¹, Sulaiman Vesal¹, Christopher Syben¹, Rainer Lutz²,
Andreas Maier¹

¹Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nürnberg,
Germany,

²Department of Oral and Maxillofacial Surgery, Friedrich-Alexander-University
Erlangen-Nürnberg, Germany
`sarah.wallraff@web.de`

Abstract. Dental panoramic X-ray images provide important information about an adolescent’s age because the sequential development process of teeth is one of the longest in the human body. Such dental panoramic projections can be used to assess the age of a person. However, the existing manual methods for age estimation suffer from a low accuracy rate. In this study, we propose a supervised regression-based deep learning method for automatic age estimation of adolescents aged 11 to 20 years to reduce this estimation error. To evaluate the model performance, we used a new dental panoramic X-ray data set with 14,000 images of patients in the considered age range. In an early investigation, our proposed method achieved a mean absolute error (MAE) of 1.08 years and error-rate (ER) of 17.52% on the test data set, which clearly outperformed the dental experts’ estimation.

1 Introduction

Milk and permanent teeth breakthrough with such regularity that there is a close correlation with the patient’s chronological age (CA). Furthermore, tooth development is only minimally affected by environmental and nutritional influences compared to skeletal growth. Since the dentition is usually very well preserved after death, its analysis is particularly suitable for determining the age and identity of deceased persons. A forensic examination of the teeth is also carried out to estimate whether a person has already reached the age of majority if no birth certificate is available [1,2]. Several manual methods have been developed to assess the CA of adolescents as a function of the degree of calcification of their milk and permanent teeth, such as the London Atlas [3] and Demirjian’s [4] method. These methods are based on the evaluation of dental X-rays. However, already after the breakthrough of the second molars at the age of about 13 years, the age determination becomes increasingly less accurate, as the changes in dentition are more irregular than in childhood [2]. The results of manual estimation methods

differ from the actual CA by about ± 1 year for adolescents on average. In up to 40% of the examined persons, the absolute deviation is even more than two years [5].

In 2017 Stern et al. [6] presented a deep learning based approach to determine the CA of adolescents. A deep convolutional neural network (DCNN) was trained with MRI images of the left hand, upper thorax, and jaw. Before training cropping of age-relevant structures was performed such that the regression is applicable for a relatively small data set. Therefore from the dental MRI data, only the four wisdom teeth were extracted. The approach was trained and evaluated on a data set of 3D MRI data of 103 male Caucasian volunteers in the age range of 13.01 to 24.89 years. The regression-based solution achieved a mean absolute error (MAE) of 1.3 ± 1.13 years.

Kahaki et al. [7] developed a DCNN for roughly classifying adolescent’s age only based on their dental X-ray images. First, the images have been segmented using fuzzy c-means, and afterward, convolutional filtering was used to extract the molar teeth. Then a DCNN was trained with those cropped teeth images captured of 356 Malaysian children aged between 1 and 17 years. The age categories were defined as 1-4, 5-7, 8-10, 11-13, and 14-17 years. On average the accuracy achieved was about 81.83% for this five-class problem.

In this paper, we investigate the feasibility of an extensive data set of 14,000 dental panoramic X-ray images for regression-based automatic age estimation on non-cropped panoramic X-rays to reduce the estimation error. To the best of our knowledge, no such extensive data set has yet been used for an automated age estimation on medical data. Compared to the mentioned related works, no pre-processing and cropping of images is necessary.

2 Material and Methods

2.1 Data

We use a clinical image data set with 110,595 dental panoramic X-rays. As teeth develop in a predictable sequence for the first 20 years of life [2], only images of patients aged 11 to 20 years were used for our approach, which leads to 14,000 images in total. Projections of the first ten years of life were omitted because there were relatively few images available from this age range, which would have resulted in a highly unbalanced data set. The age distribution is shown in Table 1. All images are labeled with the age of the patients in days, their sex and a patient identification number (PID). With regard to the patients’ sex, the data set is almost equally distributed. Our data set includes various types of clinical images such as low contrast X-rays and images of denture deviating from regular dentition caused by accidents, diseases and jaw malpositions. The data set was provided by our clinical partners². 244 images were split off as unique test data set. The age of the patients in this test data set was manually estimated by an senior physician for oral and maxillofacial surgery and three dental students specially trained in age estimation.

Table 1. Age distribution of the data set for patients in the age range of 11 to 20 years.

Age in years	11	12	13	14	15	16	17	18	19	20	Total
Number of images	975	1103	1362	1432	1621	1621	1653	1537	1501	1599	14019

2.2 Network Architecture

For our regression-based approach, we used the original ResNet18 [8] architecture. The fully-connected block at the end of the network architecture was adapted to have one node as output and consists of three fully-connected layers. Since a patient’s sex influences the development of teeth [2], two architectures were compared: A uni-sex architecture and a sex-specific architecture, which receives the sex of the patients in the last fully-connected layer as additional information. In initial tests, similarly promising results were achieved with other architectures from the ResNet, VGG and DenseNet families.

As loss function, the Huber-Loss with $\delta = 1$ was chosen. Stochastic Gradient Decent (SGD) with a learning rate of 0.001 was used as optimizer. To make use of the capacities of our large data set, the architecture was pre-trained for regression with the images that do not fall into the age group considered and whose PID does not appear in the training, validation and test data set.

Training 5-fold cross-validation was used to obtain more stable results. Per fold, approximately 10,000 images were used for training, 2,500 for validation and 244 for testing. Thereby we ensured a unique data split, which ensures that images from one and the same patient are only used for (pre-)training, validation or testing. To compensate for the high variability of patient positioning for the data acquisition, we augment the training images by random flipping, rotating by ± 5 degrees, slight cropping along the y-axis, and variations in brightness and contrast. Since the X-ray images vary in height and width, all images were resized with interpolation to 1024×1024 pixels.

Metrics Different metrics were used to assess the accuracy of the estimation. First, MAE, standard deviation (std) and mean error (ME) were analyzed. Since an exact daily estimation is not necessary for our approach, the metrics accuracy and error rate are introduced. The accuracy rate describes the percentage of estimates that differ from the actual age by less than ± 1 year, while the error rate describes the percentage of estimates that differ from the actual age by more than ± 2 years.

3 Results

The performance of the trained uni-sex network on the test data set is shown in Figure 1 depending on the patient’s sex. The linear regression line for male patients runs almost parallel to the line for optimal estimation. Male patients are generally estimated slightly too young. Female patients’ age is overestimated

until the age of 16 years and underestimated after that. This is inline with the earlier tooth development of females reported in [2]. With a ME of -0.174 and a MAE of 1.122 years are predictions worse for female than for male patients with a ME of -0.191 and a MAE of 0.951 years.

Using the sex-specific architecture, the accuracy for female patients could be improved from 50.67% to 54.24% and the error rate was reduced from 19.12% to 15.75% but leads on the other hand to an even worse ME of -0.309 and a MAE of 1.180 years. For male patients, the sex-specific training did not lead to an improvement in performance with an accuracy of 55.08% instead of 55.75% and an error rate of 14.97% instead of 15.58%.

For a better understanding of the considered image features, the trained uni-sex architecture was analyzed with Grad-CAM++ [9]. Figure 2 presents two examples of accurate estimates and Figure 3 inaccurate estimates. The focused regions are colored as a heat-map. It is noticeable that those regions do not differ between accurate and inaccurate estimates. In both cases, the area of teeth is mainly focused. The image description also shows the mean of the clinical expert’s estimates (CE), in comparison to the automatic estimation (AE). After careful evaluation of the results, it was observed that images of patients whose age was strongly misjudged do not share common characteristics such as poor image quality or conspicuous disease- and patient-specific features. It is particularly noticeable for the inaccurate estimates that the experts similarly misjudged

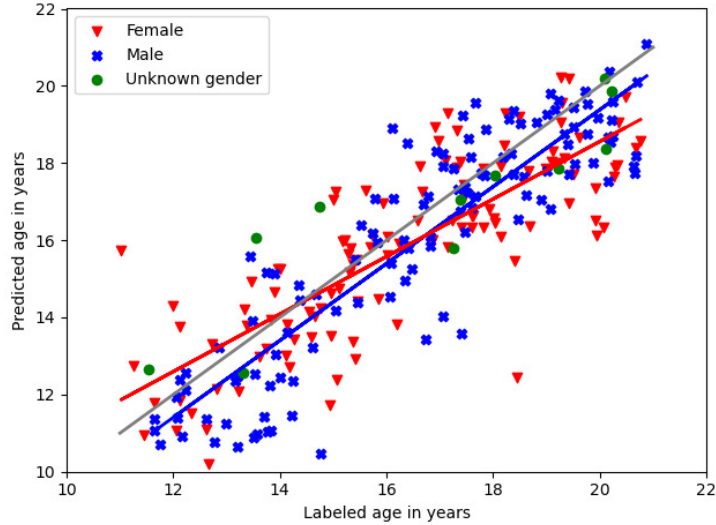


Fig. 1. Accuracy of age estimation on the test data set of the uni-sex regression-based approach depending on sex of the patients. Grey line: linear locations of perfect estimation, blue line: linear regression of estimation for male patients, red line: linear regression of estimation for female patients.

Table 2. Performance comparison between our uni-sex regression-based approach and the clinical evaluation in years on the test data set.

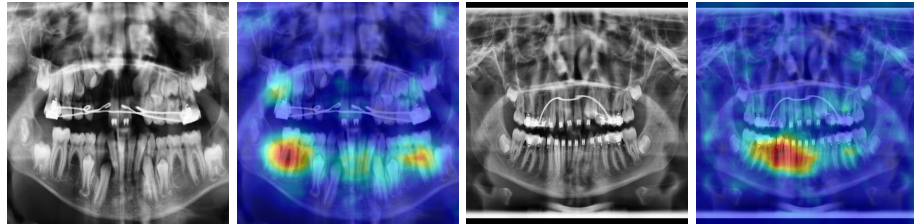
Methods	ME	std	MAE	Accuracy	Error rate
Automatic estimation (uni-sex)	-0.30	± 1.41	1.08	54.61%	17.52%
Automatic estimation (sex-specific)	-0.24	± 1.39	1.12	54.59%	15.13%
Clinical estimation	0.04	± 1.85	1.50	36.17%	30.14%

the age. Furthermore, they confirmed that forensic experts would also evaluate the areas highlighted by Grad-CAM++ primarily.

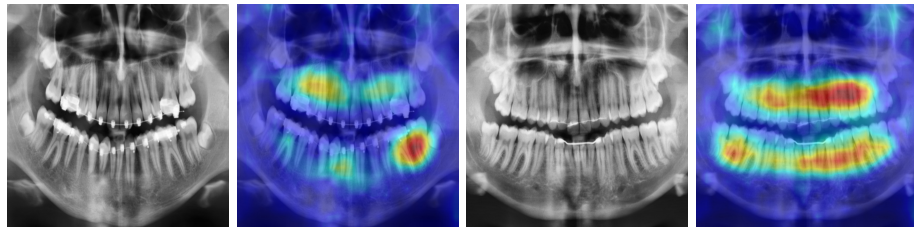
Compared to the age estimation by the clinical experts, our uni-sex and sex-specific approaches achieve better results on the test data set as shown in Table 2. A lower MAE, std and error rate were achieved as well as a higher accuracy rate. Our approaches tend to underestimate or overestimate mean age, respectively.

4 Discussion

In this paper, we have shown that regression-based deep learning for CA estimation of adolescents on dental panoramic X-rays achieves better results than manual estimation methods conducted by clinical experts. The proportion of



(a) 11 years (4253 d) (b) 11.5 years (AE), (c) 14 years (5164 d) (d) 14 years (AE), 9 years (CE) 16 years (CE)

Fig. 2. Input image and Grad-CAM++ visualization for two images with *accurate* estimates.

(a) 13 years (4910 d) (b) 15.5 years (AE), (c) 16 years (6175 d) (d) 19 years (AE), 15.5 years (CE) 18 years (CE)

Fig. 3. Input image and Grad-CAM++ visualization for two images with *inaccurate* estimates.

severe misjudgments can be reduced by using an architecture that takes the patient's sex as additional input.

The available extensive data set allows training an architecture capable of learning age-specific features without first cropping to relevant areas. This is advantageous compared to related deep learning approaches [6,7].

The achieved MAE of 1.1 years in the age range of 11 to 20 years is similar to the a MAE of 1.3 years of the multi-factorial approach based on MRI data by Stern et al. [6] considering patients in the age from 13 to 24 years although only the images of one body region were used.

Kahaki et al. [7] introduced a classification algorithm based on X-rays cropped to the molar regions for the age classes 1-4, 5-7, 8-10, 11-13, and 14-17 years with an accuracy of 81.83%. This is outperformed by the error rate of 15.13% of the proposed sex-specific approach which describes the percentage of estimates that differ from the actual age by more than ± 2 years. This shows that consideration of the entire dentition and a higher number of images allows age estimation even for older patients, who are therefore more difficult to estimate [2].

The available data set provides extensive data for the age range from 11 to 20 years while the range from 0 to 11 is underrepresented. In future, a representative data set from 0 to 20 years would be desirable to train the network.

References

1. AlQahtani SJ, Hector MP, Liversidge HM. Accuracy of Dental Age Estimation Charts: Schour and Massler, Ubelaker, and the London Atlas. *Am J Phys Anthropol.* 2014;154(1):70–78.
2. Müller N. Zur Altersbestimmung beim Menschen unter besonderer Berücksichtigung der Weisheitszähne. Universität Erlangen-Nürnberg: Dissertation; 1990.
3. AlQahtani SJ, Hector MP, Liversidge HM. The London Atlas of Human Tooth Development and Eruption. *Am J Phys Anthropol.* 2010;142(3):481–490.
4. Demirjian A, Goldstein H, Tanner JM. A New System of Dental Age Assessment. *Hum Biol.* 1973;45(2):211–227.
5. Khorate MM, Dinkar AD, Ahmed J. Accuracy of age estimation methods from orthopantomograph in forensic odontology: A comparative study. *Forensic Sci Int.* 2013;184.
6. Stern D, Kainz P, et al. Multi-factorial Age Estimation from Skeletal and Dental MRI Volumes. In: *Machine Learning in Medical Imaging*. Cham: Springer International; 2017. p. 61–69.
7. Kahaki SMM, Nordin J, et al. Deep convolutional neural network designed for age assessment based on orthopantomography data. In: *Neural Computing and Applications*. vol. 32. Springer; 2019. p. 9357–9368.
8. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 2015; p. 770–778.
9. Chattopadhyay A, Sarkar A, et al. Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks. *Proc IEEE Workshop Appl Comput Vis.* 2018; p. 839–847.