Multi-Factorial Age Estimation from Skeletal and Dental MRI Volumes

Darko Štern[1,2,⋆], Philipp Kainz[1], Christian Payer[2], and Martin Urschler[1,2]

[1]Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria
[2]Institute for Computer Graphics and Vision, Graz University of Technology, Austria

Abstract. Age estimation from radiologic data is an important topic in forensic medicine to assess chronological age or to discriminate minors from adults, e.g. asylum seekers lacking valid identification documents. In this work we propose automatic multi-factorial age estimation methods based on MRI data to extend the maximal age range from 19 years, as commonly used for age assessment based on hand bones, up to 25 years, when combined with wisdom teeth and clavicles. Mimicking how radiologists perform age estimation, our proposed method based on deep convolutional neural networks achieves a result of 1.14 ± 0.96 years of mean absolute error in predicting chronological age. Further, when fine-tuning the same network for majority age classification, we show an improvement in sensitivity of the multi-factorial system compared to solely relying on the hand.

Keywords: forensic age estimation, multi-factorial method, convolutional neural network, random forest, information fusion

1 Introduction

Age estimation of living individuals lacking valid identification documents currently is a highly relevant research field in forensic and legal medicine. Its main application comes from recent migration tendencies, where it is a legally important question to distinguish adult asylum seekers from adolescents who have not yet reached age of majority. Widely used radiological methods for forensic age estimation in children and adolescents take into account complementary biological development of skeletal [3, 12] and dental structures [2]. This allows an expert to examine progress in physical maturation related to closing of epiphyseal gaps and mineralization of wisdom teeth. Despite biological variation among subjects of the same chronological age (CA), hand bones are the most suitable anatomical site to follow physical maturation in minors, since epiphyseal gaps start closing at different times, with distal bones finishing earlier and e.g. the radius bone finishing at an age of about 18 years. However, the age range of interest for forensic age estimation is between 13 and 25 years. Therefore, additional anatomical sites are required in a multi-factorial approach to allow an extension of the age estimation range up to 25 years.

⋆ This work was supported by the Austrian Science Fund (FWF): P 28078-N33.

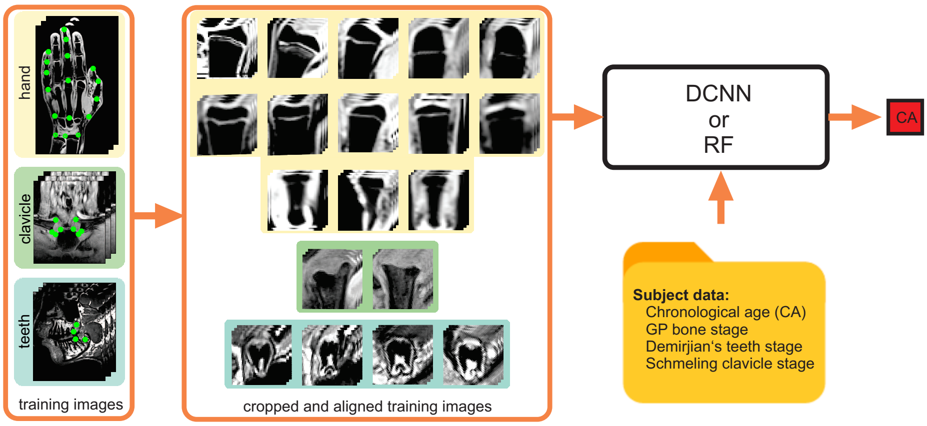

Fig. 1: Overview of our automatic multi-factorial age estimation framework. MRI volumes of hand, clavicle and wisdom teeth are cropped according to locations of age-relevant anatomical landmarks. A random forest (RF) or a deep convolutional neural network (DCNN) performs the nonlinear mapping between appearance information and chronological age.

The established X-ray imaging based multi-factorial approach [7] for estimation of biological age (BA) uses the Greulich-Pyle (GP) method [3] based on representative hand images of different age groups of a sample population, the Demirjian method [2] involving characteristic stages of wisdom teeth development, and the staging method of Schmeling [8] for assessing clavicle bone maturation. No standardized method exists for the combination of different sites, but, for majority age estimation, guidelines propose to use the minimum age of the most developed anatomical site as seen in a reference population [7].

Besides the lack of a standardized method for combining individual estimates, radiological methods also suffer from intra- and inter-observer variability when determining, from each anatomical site, the stages that define minimum age. While the use of more objective, automated image analysis for age estimation from X-ray data of the hand was already shown in [13, 9], no such approaches yet exist for orthopantomograms of the teeth or computed tomography images of the clavicle bones. A novel trend in forensic age estimation research is to replace X-ray based methods with magnetic resonance imaging (MRI), because legal systems in most countries disallow the application of ionizing radiation on healthy subjects. Recently, automatic methods for age estimation based on MRI data were developed [11, 10], nevertheless with the hand they also solely investigate a single anatomical site. To the best of our knowledge no automatic image analysis method for multi-factorial age estimation, irrespectively of the imaging modality, has been presented yet.

In this work, we investigate novel methods for multi-factorial age estimation from MRI data of hand bones, clavicles and wisdom teeth (see Fig. 1). Inspired by how radiologists perform staging of different anatomical sites, our methods automatically fuse the age-relevant appearance information from individual anatomical structures into a single chronological age. We compare a random forest (RF) based method [1] with two deep convolutional neural network (DCNN) architectures [4]. The first DCNN is trained on CA and the second one fine-tuned on CA after pre-training it using skeletal and dental age as BA estimates determined by an expert. The proposed methods are evaluated on an MRI database of 103 images by performing experiments assessing CA estimates in terms of regression, as well as distinction of minority/majority age, defined as having passed the 18th birthday, in terms of classification. Our results demonstrate the increase in accuracy and decrease in uncertainty when using the multi-factorial approach as compared to relying on a single anatomical structure.

2 Method

Following the established radiological staging approach involving different anatomical sites in a multi-factorial setup, after cropping of age-relevant structures we perform age estimation from cropped wisdom teeth, hand, and clavicle bones, either by the use of an RF or a DCNN architecture.

Cropping of age-relevant structures: Differently to [9], where a large data set of whole X-ray images is used for age estimation, our motivation for cropping age-relevant structures is to simplify the problem of regressing age from appearance information, such that it is also applicable for smaller data sets. Therefore, automated landmark localization methods as presented in [5] or [6] could be used to localize, align and volumetrically crop age-relevant anatomical structures from skeletal and dental 3D MRI data (see Fig. 1). From hand MRI we crop the same thirteen bones that are also used in the TW2 RUS method [12], similar to [10]. Four wisdom teeth are extracted from the dental MRI data using the locations of the centers of each tooth, and in clavicle MRI data the two clavicle bones are cropped based on four landmarks on the manubrium and two on each clavicle.

RF framework: Starting from the easily extensible framework for hand MRI age estimation proposed in [11], we additionally incorporate the selection of teeth and clavicle bones into each node of an RF. Thus, we allow the RF to select from which anatomical structure it extracts the features that are relevant for modeling the mapping between image appearance information and CA. After training it for regression of CA from all three anatomical sites, we denote this method RF-CA. Additionally, we train the same framework for majority age classification (RF-MAJ), and to compare to previous work [10], we also train an RF using BA as a regression target solely from the hand MRI data (RF-BA-HAND).

DCNN architecture: Identical feature extraction blocks consisting of convolution (conv) and pooling (pool) stages are used for individual cropped input volumes. Fusion is performed for anatomical sites separately giving a final representation of extracted features. Estimated CA is obtained by combining the extracted features from the three sites with a fully connected (fc) layer.

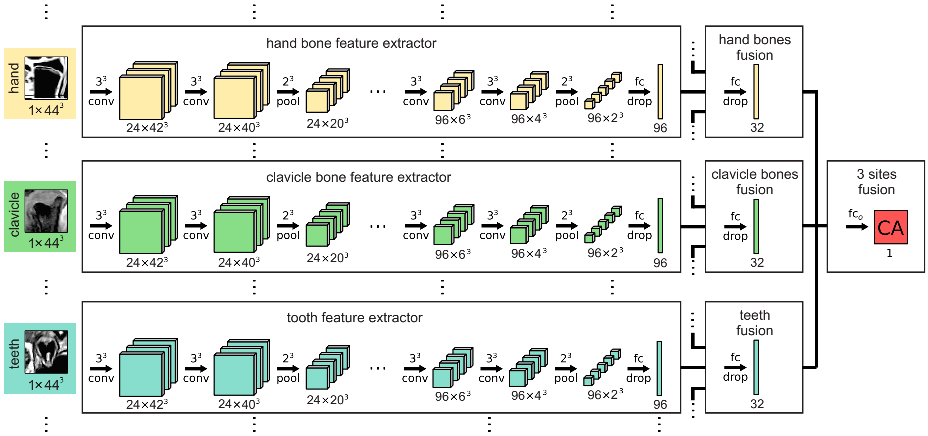

Fig. 2: DCNN architecture for multi-factorial age estimation.

The details of our individual identical DCNN blocks [4] are shown in Fig. 2, where we connect three stages consisting of two convolution and one max-pooling layer together with Rectified Linear Units ($ReLUs$) as nonlinear activation functions. Each block finishes with a fully connected layer, leading to a dimensionality reduced feature representation consisting of 96 outputs for each cropped input volume. Thus, we require another fully connected layer to fuse feature representations into a single feature vector for the different structures at each anatomical site. Finally, all three sites are fused with a fully connected layer to form a single continuous CA regression target. To reduce overfitting, we include drop-out regularization with a ratio of 0.5 into all fully connected layers except the last layer which solely has a single output. We denote this network DCNN-CA using CA as regression target. Since our network architecture is mimicking how radiologists perform staging of different anatomical sites, it readily supports the use of the assigned stages representing BA to pre-train the network weights of each individual site. This can be achieved by decoupling the last fully connected layer fc_o from the network and adding individual fully connected layers with a single output for each anatomical site. By training these individual networks with their respective radiological stage (e.g. DCNN-BA-HAND), we expect to achieve a better initialization of network weights compared to training DCNN-CA from scratch solely on CA. Fine-tuning of the pre-trained network on CA leads to our network DCNN-CA-RFND. Further, we use the same pre-trained network to directly predict whether a subject is an adult or a minor by fine-tuning network parameters for a classification target instead (DCNN-MAJ).

For training, we associate each training sample $s_n, n \in \{1, .., N\}$, consisting of thirteen cropped hand bone volumes $s^j_{n,h}, j \in \{1, .., 13\}$, 4 wisdom teeth $s^k_{n,w}, k \in \{1, .., 4\}$ and 2 clavicle regions $s^l_{n,c}, l \in \{1, 2\}$, with a regression target y^A_n. Depending on whether it is used for pre-training or for direct training/fine-tuning, here A is either chronological age CA or biological age BA defined as the average assigned radiological stage of the components of each anatomical site. Optimizing a DCNN architecture ϕ with parameters $\mathbf{w}$ is performed by applying stochastic gradient descent to minimize an L_2 loss on the regression target y^A:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \frac{1}{2} \sum_{n=1}^{N_s} ||\phi(s_n; \mathbf{w}) - y_n^A||^2. \tag{1}$$

For estimating whether a subject is a minor (m) or an adult (a) with DCNN-MAJ, we use the legally relevant chronological majority age threshold of 18 years to separate our subjects into two groups $\{m, a\}$. As an optimization function for this classification task we apply softmax computed as multinomial logistic loss:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \sum_{n=1}^{N_s} \sum_{j \in \{m,a\}} -y_n^j \log \frac{e^{\phi_j(s_n;\mathbf{w})}}{\sum_{k \in \{m,a\}} e^{\phi_k(s_n;\mathbf{w})}}. \tag{2}$$

To distinguish whether introducing multiple sites is beneficial for discriminating minors from adults, we apply the same classification loss for classification based solely on hand bones. Additionally, this network DCNN-MAJ-HAND facilitates a comparison with previous work on hand bone age estimation [10].

3 Experimental Setup and Results

Material: We apply our proposed method on a dataset of $N = 103$ 3D MRIs of the left hand, the upper thorax and the jaw, respectively. The three volumes for each subject were prospectively acquired from male Caucasian volunteers with known CA ranging between 13.01 and 24.89 years (mean±std: 19.1±3.5 years, 44 subjects were minors below 18 years) in a single MRI scan session. CA of subjects was calculated as difference between birthday and date of the MRI scan. T1-weighted 3D gradient echo sequences were used for acquiring the hand and clavicle data (resolutions of $0.45 \times 0.45 \times 0.9$ and $0.9 \times 0.9 \times 0.9$ mm^3), while teeth were scanned using a proton density weighted turbo spin echo sequence ($0.59 \times 0.59 \times 1.0$ mm^3).

Regarding biological ages y_n^{BA} as determined by a board-certified radiologist as well as a dentist, for the hand volumes the GP standard [3] y_n^{GP} was used, the wisdom teeth were assessed using the Demirjian system [2] y_n^{DM}, and finally the clavicles were rated with the Schmeling system [8] y_n^{SM}.

Experimental Setup: A cross-validation with four folds was used to compute results for all experiments. In each cross-validation round, one fourth of the available datasets were used for testing, while the remaining subjects were used to generate training datasets for the RF and DCNN methods. To allow a meaningful evaluation over the given age range, the test sets were chosen by sampling according to the approximately uniform age distribution of our data, where random sampling required that at least two datasets from each bin were chosen for each fold. During training of both RF and DCNN, images were augmented

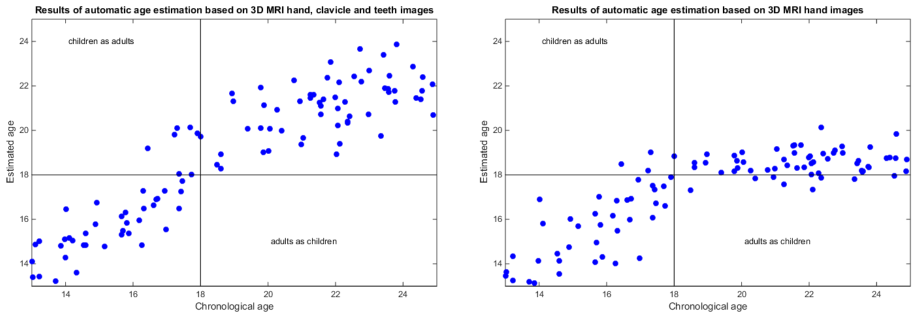

Fig. 3: Regression results for chronological age estimation based on DCNN-CA-RFND (left) and DCNN-CA-HAND restricted to hand MRI data (right).

on-the-fly by translating, rotating, scaling and shifting intensities of the cropped volumes. The size of cropped volumes after resampling with trilinear interpolation was $44 \times 44 \times 44$ pixels for all anatomical sites, respectively. We implemented our RF as a regression forest consisting of 100 trees, maximal tree depth of 20, and the number of candidate features/thresholds per node was set to 100/10, respectively. For the DCNN the *Caffe* framework[1] was used. Optimization was done with stochastic gradient descent with a maximum of $2 \cdot 10^4$ iterations, mini-batch size 1, momentum 0.99, and learning rate 10^{-5}. To evaluate our CA regression results, we compute mean and standard deviation of absolute errors between predicted and ground truth age. Classification experiments are evaluated by inspecting confusion matrices (TP,TN,FP,FN) and assessing accuracy (ACC), specificity (SPEC) and sensitivity (SENS), with the latter indicating the percentage of subjects classified as minors who are actually minors.

Results: For estimating biological age from the hand, absolute deviation results are 0.62 ± 0.58 for RF-BA-HAND and 0.33 ± 0.31 years for DCNN-BA-HAND. The mean absolute error of predictions compared to CA are 1.93 ± 1.26 for RF-CA, 1.3 ± 1.13 for DCNN-CA, and 1.14 ± 0.96 years for DCNN-CA-RFND. Regarding majority age classification, Table 1 shows the confusion matrix as well as accuracy, sensitivity and specificity results for the compared methods.

4 Discussion and Conclusion

Motivated by the lack of a standardized way of fusing age estimates from multiple complementary anatomical sites, and with the aim to reduce observer variability, we proposed two automated methods for age estimation of the living from MRI data in the age range of 13 to 25 years. It is for the first time that such a large age range was studied by an automatic approach. To verify that our proposed methods achieve state-of-the-art results on a reduced age range between 13 and

[1] Y. Jia, GitHub repository, https://github.com/BVLC/caffe/

Table 1: Classification results for the majority age experiments (N=103).

	TP	TN	FP	FN	ACC	SENS	SPEC
DCNN-MAJ	39 (37.9%)	55 (53.4%)	4 (3.9%)	5 (4.8%)	91.3%	88.6%	93.2%
DCNN-MAJ-HAND	36 (35.0%)	57 (55.3%)	2 (1.9%)	8 (7.8%)	90.3%	81.8%	96.6%
RF-MAJ	21 (20.4%)	58 (56.3%)	1 (1.0%)	23 (22.3%)	76.7%	47.7%	98.3%
RF-MAJ-HAND	25 (24.3%)	59 (57.3%)	0 (0.0%)	19 (18.4%)	81.6%	56.8%	100%

19 years, we first compared them with previous work from [10], based solely on hand MRIs for estimation of BA provided as bone age stages by a radiologist. For RF-BA-HAND, we achieved a mean absolute error of 0.62 ± 0.58 years on our data set, while the best RF result in [10] was 0.52 ± 0.60 years, but required a pre-processing step. Without any pre-processing but training directly on image intensity, our result of DCNN-BA-HAND (0.33 ± 0.31 years) was better than the overall best results of 0.36 ± 0.30 years reported in [10]. However, results should be interpreted carefully, since different data sets were used.

Due to biological variation of subjects at the same chronological age, estimation of CA as required in forensic medicine is a much harder task compared to regression of BA. Our main contribution in this work is the extension of the maximal age range for CA regression from 19 years, as possible with solely using hand images, to 25 years, when including data from three anatomical sites. This extension is clearly visible by comparing the scatter plots from DCNN-CA-RFND and DCNN-CA-HAND in Fig. 3. The plot corresponding to the extended age range regression further confirms that after the development of the hand has finished, uncertainty in estimating CA increases since less age relevant features are available. Compared to training DCNN-CA from scratch on chronological age, pre-training of the DCNN-CA-RFND network on the radiological staging results of the individual anatomical sites improved the mean absolute CA regression error from 1.3 ± 1.13 to 1.14 ± 0.96 years. Leading to a much higher error of 1.93 ± 1.26 years, we found that RF-CA was not able to achieve competitive age estimation results for the whole age range by selecting intensity features from all three anatomical sites, while the DCNNs superior feature extraction capabilities proved to be more powerful despite the low amount of training data.

A specific challenge in forensic age estimation is majority age classification of asylum seekers lacking valid identification documents, under the ethical constraint that legal authorities need to avoid misclassifications of minors as adults, i.e. requiring high sensitivity. While the RF based methods do not show competitive classification results, our DCNN-MAJ-HAND network that uses majority age as a binary classification target achieves a classification accuracy of 90.3%, while misclassifying 8 out of 44 minor subjects as adults (see Table 1). Refining the pre-trained multi-factorial network combining all three anatomical sites on majority classification (DCNN-MAJ) improves classification accuracy to 91.3%. More importantly, the number of misclassified minors is reduced to 5 out of 44, an improvement in sensitivity which greatly impacts the involved subjects to their advantage.

A limitation of our results is the low number of 103 studied subjects, so generalization of these results has to be done carefully until a larger set has been evaluated. Additionally, our method was evaluated on volumes cropped using manually annotated landmarks instead of predictions from a landmark localization algorithm. However, since our training stage involved random translational and rotational transformations in the data augmentation step, we expect the same performance using accurate and robust localization algorithms like [5, 6].

In conclusion, we have demonstrated with our proposed DCNN method that multi-factorial age estimation based on the three anatomical sites (hand, wisdom teeth, clavicle) can be used to automatically estimate chronological age in the living, extending the age range up to 25 years. However, caution has to be taken when it is used for deciding whether a subject is a minor or an adult.

References

1. Breiman, L.: Random Forests. Mach. Learn. 45, 5–32 (2001)
2. Demirjian, A., Goldstein, H., JM, T.: A new system of dental age assessment. Hum. Biol. 45(2), 211–227 (1973)
3. Greulich, W.W., Pyle, S.I.: Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press, Stanford, CA, 2nd edn. (1959)
4. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998)
5. Lindner, C., Bromiley, P.A., Ionita, M.C., Cootes, T.F.: Robust and Accurate Shape Model Matching Using Random Forest Regression-Voting. IEEE Trans. Pattern Anal. Mach. Intell. 37(9), 1862–1874 (2015)
6. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing Heatmaps for Multiple Landmark Localization using CNNs. In: Int. Conf. Med. Image Comput. Comput. Interv. pp. 230–238. Springer (2016)
7. Schmeling, A., Geserick, G., Reisinger, W., Olze, A.: Age estimation. Forensic Sci Int 165(2-3), 178–181 (2007)
8. Schmeling, A., Schulz, R., Reisinger, W., Muehler, M., Wernecke, K.D., Geserick, G.: Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. Int J Leg. Med 118(1), 5–8 (2004)
9. Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R.: Deep learning for automated skeletal bone age assessment in X-ray images. Med. Image Anal. 36, 41–51 (2017)
10. Štern, D., Payer, C., Lepetit, V., Urschler, M.: Automated Age Estimation from Hand MRI Volumes using Deep Learning. In: Int. Conf. Med. Image Comput. Comput. Assist. Interv. pp. 194–202. Springer (2016)
11. Štern, D., Urschler, M.: From Individual Hand Bone Age Estimation to Fully Automated Age Estimation via Learning-Based Information Fusion. In: 2016 IEEE 13th Int. Symp. Biomed. Imaging. pp. 1–5 (2016)
12. Tanner, J.M., Whitehouse, R.H., N, C., Marshall, W.A., Healy, M.J.R., Goldstein, H.: Assessment of skeletal maturity and predicion of adult height (TW2 method). Academic Press, 2nd edn. (1983)
13. Thodberg, H.H., Kreiborg, S., Juul, A., Pedersen, K.D.: The BoneXpert Method for Automated Determination of Skeletal Maturity. IEEE Trans. Med. Imaging 28(1), 52–66 (2009)