Automated Age Estimation from Hand MRI Volumes using Deep Learning

Darko Štern[1,⋆], Christian Payer[2], Vincent Lepetit[2], and Martin Urschler[1,2,3]

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

Abstract. Biological age (BA) estimation from radiologic data is an important topic in clinical medicine, e.g. in determining endocrinological diseases or planning paediatric orthopaedic surgeries, while in legal medicine it is employed to approximate chronological age. In this work, we propose the use of deep convolutional neural networks (DCNN) for automatic BA estimation from hand MRI volumes, inspired by the way radiologists visually perform age estimation using established staging schemes that follow physical maturation. In our results we outperform the state of the art automatic BA estimation method, achieving a mean error between estimated and ground truth BA of 0.36 ± 0.30 years, which is in line with radiologists doing visual BA estimation.

1 Introduction

Estimation of the progress of physical maturation of individuals, which in literature is referred to as biological age (BA) estimation, is an important topic in clinical medicine when determining endocrinological diseases in adolescents or for optimally planning the time-point of paediatric orthopaedic surgery interventions, e.g. for leg-length discrepancy correction [7]. Due to biological variation, BA differs from chronological age (CA). Nevertheless, in legal medicine, BA is used to approximate unknown chronological age (CA) when determining age in cases of criminal investigations or for asylum seeking procedures, where identification documents of children or adolescents are missing [9].

Widely used radiological methods for BA estimation are based on visual examination of ossification, i.e. epiphyseal plate fusion, of individual bones in X-ray images of the hand. In these examinations, radiologists exploit the fact that aging is not the same for all bones of the hand. Distal phalanges are the first to finish ossification while in radius and ulna maturation can be followed up to an age of around 19 years. Greulich-Pyle [3] (GP) is a preferred radiologic age estimation method as it is easy to use and fast to apply: All hand bones are simultaneously compared to the best matching reference image from an atlas

⋆ This work was supported by the province of Styria (HTI:Tech_for_Med ABT08-22-T-7/2013-13) and the Austrian Science Fund (FWF): P 28078-N33.

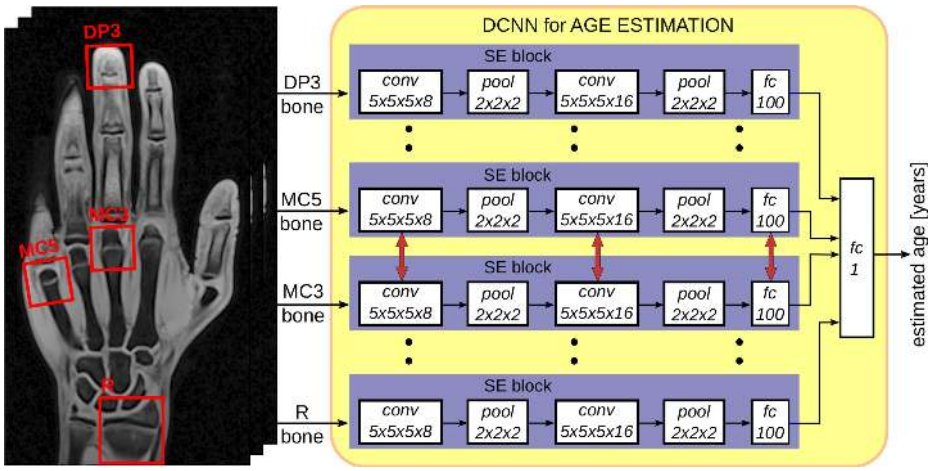

Fig. 1: Our DCNN architecture. Each stage estimation (SE) block performs dimensionality reduction and extracts age relevant features from cropped 3D bone volumes. By feeding feature outputs into a fully connected layer, age estimation is performed. Red arrows indicate weight sharing among SE blocks.

of radiographs. However, radiologists have to visually extract and mentally fuse information from different bones, which makes the GP method prone to intra- and inter-rater variability. In the more complex Tanner-Whitehouse [10] (TW2) method, which is considered to be more accurate than GP [9], visual examination of 13 selected hand bones is simplified by separating their ossification process into stages according to textual and visual descriptions. Individual bone age estimates are then transferred to an ossification score and fused according to a pre-defined nonlinear function. Both the ossification score and the fusion function were derived from characteristics of a sample population.

Automated image analysis methods for BA estimation have recently started to appear. Most prominent, the BoneXpert [11] method uses Active Appearance Models to automatically segment hand bones and employs principal component analysis to reduce age relevant shape and appearance feature information. For BA estimation, the fusion of individual estimations per bone is calibrated using the same pre-defined nonlinear function as in TW2. A new trend in BA estimation is the use of volumetric MRI, which may provide fundamentals for more accurate and reliable estimation, without harmful ionizing radiation. Applied to adolescents, the automatic MRI-based BA estimation method of [12] uses a random forest (RF) to separately regress CA from image intensity based features of 11 selected hand bones. There, a decision tree excluding metacarpal and phalanx information from age estimation of older subjects serves as a heuristic fusion strategy, making this method *ad hoc* and depending on parameter tuning. In our recently proposed method [13], we explore the capability of RFs for information fusion by allowing it to internally decide from which bones to learn a subject's CA. Thus, aging is treated as a global developmental process without the need

for pre-defined nonlinear functions [10, 11] or heuristic fusion schemes [12]. However, by introducing more bones into RF, the space from which hand-crafted features are generated is increased making the discrimination task harder. Thus, an image preprocessing step emphasizing epiphyseal plates by filtering artifacts and intensity inhomogeneities had to be introduced to simplify discrimination.

Very recently, deep convolutional neural networks (DCNN) have shown to be immensely successful in solving diverse machine learning and computer vision problems [6, 5], mainly due to their ability to automatically learn task-relevant features from large training datasets. In this work we follow this novel direction and explore the capability of DCNNs to automatically estimate a subject's age given 3D hand MRI volumes depicting ossification. We propose a novel DCNN architecture inspired by the best performing, visual based TW2 method to combine age information from individual bones in an automatic fashion by letting the DCNN learn the features relevant for age estimation. Thus, our DCNN mimics a radiologist performing age estimation, but with the goal to eliminate intra- and inter-rater variability. By working directly on 3D input volumes, our proposed DCNN outperforms the state of the art in 3D MRI BA estimation.

2 Method

Our deep neural network architecture for 3D MRI bone age estimation is presented in Fig. 1. Following the idea of the TW2 method, which estimates ossification stages of hand bones separately, our proposed DCNN consists of identical per-bone stage estimation (SE) blocks. They are designed to reduce dimensionality of appearance features in 3D bone volumes, thus capturing age relevant features defined by the ossification process. To achieve a continuous BA prediction, fusion of independent age scores from each bone is implemented in our DCNN by connecting the outputs of all SE blocks in a fully connected layer.

2.1 Image preprocessing

Based on a landmark localization algorithm such as [1] or [8], we automatically localize, align and crop the 13 bones that are also used in the TW2 method [10] (see Fig. 2a). To reduce image intensity variations, thus potentially simplifying the learning task, we experiment with an image pre-processing step that enhances the appearance of epiphyseal plates from its surrounding anatomical structures. This pre-processing step utilizes planarity of epiphyseal plates by generating a filtered image representation I_i^b for each cropped bone volume $b = \{1, .., N_B\}$ of hand $i = \{1, .., N\}$. It is based on an eigenanalysis of the Hessian matrix computed from the second image derivative, as inspired by [2]. Thus, to enhance plate structures we compute

$$I_i^b = \frac{1}{1 + exp\left(-\frac{|\lambda_1| - \zeta_1}{\zeta_2}\right)} \cdot exp\left(-\frac{|\langle \boldsymbol{v}_1, \boldsymbol{n}_z \rangle - 1|}{\zeta_3}\right), \tag{1}$$

where the left term exploits that in planar structures, response of Hessian eigenvalues is $|\lambda_1| \gg 0$, $|\lambda_{2,3}| \approx 0$. With $\zeta_1 = 40$ and $\zeta_2 = 5$, it is possible to enhance $|\lambda_1|$ inside the epiphyseal plate. The right term penalizes the deviation of Hessian eigenvector $\boldsymbol{v}_1$, i.e. the plane normal, from the longitudinal axis of the aligned bone $\boldsymbol{n}_z$, via their dot product, scaled by $\zeta_3 = 0.25$. In our experiments, we compare original image intensities (II) to preprocessed 3D bone volumes (FI) as an input to the evaluated algorithms, e.g. DCNN-II vs. DCNN-FI.

2.2 DCNN architecture

We construct our SE blocks inspired by the LeNet architecture [6], due to its excellent capabilities for dimensionality reduction and feature extraction, while requiring only a small number of model parameters. Thus, as the first layer of each 3D SE block (see Fig. 1), we use a convolutional layer (*conv*) with a filter size of $5 \times 5 \times 5$ pixels and 8 filter outputs followed by a Rectified Linear Unit ($ReLU$) as nonlinear activation function. After activation, outputs are subsampled with a MAX pooling layer (*pool*) of size $2 \times 2 \times 2$. The same convolution, activation and pooling step is then repeated with 16 convolution filter outputs. The last layer of an SE block is fully connected (*fc*) with 100 outputs, again followed by a $ReLU$ activation unit. To prevent overfitting, we include dropout regularization with a ratio of 0.5 into the *fc* layer. Thus, each SE block reduces the dimensionality of the input bone volume by extracting a feature vector of size 100. This feature vector captures the same single ossification score that a rater performing BA estimation with the TW2 method would generate for an individual bone. According to the TW2 staging scheme, the third and fifth bones of the same finger group (i.e. metacarpals, proximal-, middle- and distal phalanges), show the same physical maturation process, i.e. ossification scores are identical for their ossification stages. For the bones of these groups, we therefore emulate this concept by weight sharing (TW2ws) among different layers of the DCNN's SE blocks (indicated by red arrows in Fig. 1), leading to the methods DCNN-TW2ws-II, DCNN-TW2ws-FI. The final age estimation output is obtained after fully connecting the extracted per-bone feature vectors, with a single continuous age prediction output.

Besides training our DCNN on CA, we experiment with regressing BA as determined by a radiologist, since it has a smaller deviation from the "true" biological age that we are aiming to estimate. Thus, each training sample $s_n, n \in \{1, .., N_s\}$ is associated with an age y_n^A, which is either BA or CA with $A \in \{BA, CA\}$ depending on the experiment. Using stochastic gradient descent optimization, the DCNN ϕ with parameters $\mathbf{w}$ is trained to minimize the L_2 loss:

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

Unlike clinical medicine applications where only BA estimation is required, in legal medicine CA is approximated with BA, but additionally it is often important to answer the question whether a person meets the legal criteria of

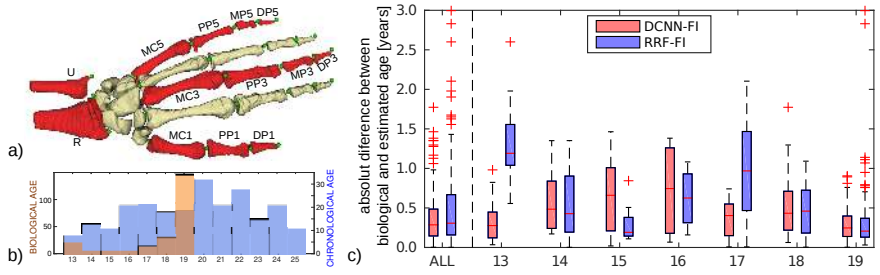

Fig. 2: a) DCNN performs age regression on the same 13 bones as TW2. b) Biological and chronological age distributions of our N=240 dataset. c) BA estimation results separately for age groups, comparing best performing RRF-FI and DCNN-FI methods.

e.g. having reached the age of criminal responsibility or majority age. Thus, we use the same DCNN architecture to discriminate between minors (m) from adults (a) by separating all testing subjects into two classes defined by the legally relevant chronological majority age threshold of 18 years. For this classification task we use the softmax loss computed as multinomial logistic loss:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \sum_{n=1}^{N_s} \sum_{j \in \{m,a\}} -y_n^{A,j} \log \frac{e^{\phi_j(s_n;\mathbf{w})}}{\sum_{k \in \{m,a\}} e^{\phi_k(s_n;\mathbf{w})}} \quad . \tag{3}$$

3 Experimental Setup and Results

Material: We use a dataset of $N = 240$ T1-weighted 3D gradient echo hand MRI volumes ($294 \times 512 \times 72$ voxels at a resolution of $0.45 \times 0.45 \times 0.9mm^3$) acquired from male Caucasian volunteers with known CA between 13 and 23 years. For evaluation, we investigate the same $N_B = 13$ bones as in TW2 (compare Fig. 2a). CA of subjects is calculated as difference between birthday and date of the MRI scan. BA is estimated based on the GP method [3] by consent of two radiologists, since TW2 method is currently not used for radiological age estimation from MRI data, because the scoring system and the nonlinear fusion function have not been adapted. We assign an age of 19 years to all subjects with CA of 19 or above, since due to finished physical development, all epiphyseal plates have vanished, thus no age relevant features can be extracted.

Experimental Setup: The results of all experiments were computed in an eight-fold cross validation. In each cross-validation round, 30 subjects are tested, while the remaining 210 subjects are used to generate a training dataset of $N_s = 1050$ samples filling up the distribution of training samples (see Fig. 2b) to a uniform distribution over age. For increasing the training dataset, cropped volumes of each bone are slightly rotated and translated around the estimated anatomical landmarks defining the bone. To reduce the number of parameters optimized by the DCNN, the part of the bone volume that contains the epiphyseal plate is used and all bone images are resized to $40 \times 40 \times 40$ pixels. Implemented in the *Caffe* framework [4], our DCNN was optimized with stochastic

Table 1: Errors in BA estimation when training on intensity (II) or filtered (FI) images using BA or CA as regression target are given as mean (± standard deviation) absolute differences between estimated and ground truth age in years.

method	BA → BA	CA → BA
RRF-FI [13]	0.52 ± 0.60	0.62 ± 0.58
RRF-II [13]	0.61 ± 0.62	0.74 ± 0.62
DCNN-FI	**0.36 ± 0.30**	**0.56 ± 0.44**
DCNN-II	0.42 ± 0.36	0.60 ± 0.47
DCNN-TW2ws-FI	0.39 ± 0.30	0.60 ± 0.45
DCNN-TW2ws-II	0.43 ± 0.35	0.66 ± 0.52

gradient descent with a maximal number of iterations 10^4, momentum 0.9 and learning rate 10^{-4}. For estimation of BA and classification to discriminate between minors and adults, we experimented with training our DCNN on both BA and CA. The results from training the DCNN on original intensity images (II) and on the filtered images (FI) as explained in Section 2.1 are compared. As a baseline method we use the age estimation based on random regression forests (RRF) as proposed in [13], with the only difference that in this work we used an increased number of training samples, the same as for the DCNN.

Results: Using either intensity (II) or filtered (FI) volumes for training, the results of BA estimation trained BA and CA are given in Table 1. Detailed results separately for each biological age group of the best performing DCNN-FI compared with RRF-FI are presented in the box-whiskers plot in Fig. 2c. The contingency table of classifying subjects as being minor or adult is given in Table 2. All results are compared with the RRF age estimation method of [13].

4 Discussion and Conclusion

Inspired by TW2 [10], which is considered the most accurate radiological hand bone age estimation method due to its fusion of independent per-bone estimates, we have designed our novel automatic age estimation DCNN using an architecture mimicking the TW2 method. Limited by dataset size, our design choice to use the LeNet architecture [6] as a building block for our method was motivated by keeping the number of DCNN weights as low as possible. To further reduce model complexity, we also experimented with sharing network weights between SE blocks of bones that undergo the same physical maturation process, an idea borrowed from the TW2 staging scheme. However, as shown in Table 1, we experienced no performance gains, which might be due to our limited number of training images but could also indicate subtle different physical maturation processes in bones where the TW2 staging system assigns the same score. Compared to the selection of hand-crafted features in RFs, DCNNs internally learn to generate the features relevant for age estimation. This comes at the cost of requiring a larger number of training data, therefore we obtain additional training data by augmentation with synthetic transformations. We found that when using the

Table 2: Classification error when determining majority age is given as true positive (TPR), false positive (FPR), true negative (TNR) and false negative (FNR) rate.

adulthood	method	trained age	TPR	FPR	TNR	FNR
$BA > 18$	DCNN-FI	BA	100.0	0.0	100.0	0.0
	radiologist	BA	98.7	28.6	71.4	1.3
$CA > 18$	DCNN-FI	BA	98.7	28.6	71.4	1.3
	DCNN-FI	CA	98.7	3.6	96.4	1.3

pre-processed filtered images (FI) as input to our DCNN, a higher estimation accuracy compared to raw intensity images could be achieved. Thus, in accordance with [13], by suppressing image intensity variations and enhancing the appearance of the ossifying epiphyseal plate from the surrounding anatomical structures it was possible to simplify the learning task for the DCNN.

For discussing results presented in Tables 1 and 2, it is important to understand that "true" BA, which we want to estimate, is the average stadium of physical development for individuals of the same CA. Therefore, the estimation of "true" BA would require a large dataset of subjects with given CA that statistically represents biological variation. Since our limited dataset can only partly cover biological variation in the target age range, we use BA as estimated by a radiologist as ground truth for training and testing, although the deviation from "true" BA that is introduced by the radiologist [9] can never be corrected by an algorithm. Moreover, the reported inter-observer variation for radiographic images varies in the range of 0.5 to 2 years, depending on the age, sex and origin of the examined population [9]. In clinical medicine applications, when biological age is required, training our DCNN-FI method on BA estimated by radiologists shows higher accuracy (0.36 ± 0.30 y) compared to training on CA (0.56 ± 0.44 y) on our dataset. The higher error can be explained by biological variation in the training dataset using CA. As shown in Table 1, our best performing DCNN-FI method outperforms previous work when estimating BA. When interpreting the detailed results in Fig. 2c, it has to be noted that the improvement of RRF-FI upon our previous work [13] is due to a larger, synthetically enhanced training dataset and the method being trained and evaluated on BA. Depending on the used population, results of the prominent automatic BoneXpert age estimation method [11] were reported between 0.65 and 0.72 years when compared to radiologists GP ratings for X-ray images of male boys, but further comparison to our method has to be taken with care due to the differences in datasets.

In legal medicine applications, recent migration tendencies lead to challenges, when asylum seekers without identification documents have to be discriminated according to having reached majority age. As can be seen in Table 2, our DCNN-FI classifier trained on BA is able to perfectly discriminate between subjects below and above 18 years of BA in our dataset. Nevertheless such a perfectly discriminating classifier trained on BA makes a larger error by classifying 28.6% of minors to be adults, the same error that radiologists make when approximating BA with CA using the GP method. Thus, better discrimination can not be

achieved by a classifier when using BA defined by radiologists for training. We further retrained our classifier using CA for training and achieve significantly better discrimination of legal majority age, misclassifying 3.6% minors to be adults. This observed behavior is in line with literature showing that BA estimated with the GP method has the tendency to underestimate CA [9] due to advanced physical maturation in nowadays population, while GP is based on radiographs that were acquired in the 30s of the last century.

In conclusion, our proposed DCNN method has proven to be the best automatic method for BA estimation from 3D MR images, although it has to be used carefully in legal medicine applications due to the unavoidable misclassification when discriminating minors from adults, which is caused by biological variation.

References

1. Ebner, T., Štern, D., Donner, R., Bischof, H., Urschler, M.: Towards Automatic Bone Age Estimation from MRI: Localization of 3D Anatomical Landmarks. In: MICCAI 2014, Part II. LNCS, vol. 8674, pp. 421–428. Springer (2014)
2. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer (1998)
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. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional Architecture for Fast Feature Embedding. In: Proc. ACM Int. Conf. on Multimedia (MM'14). pp. 675–678 (2014)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. In: NIPS. pp. 1097–1105 (2012)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient based learning applied to document recognition. Proceedings of the IEEE 86(11), 2278–2323 (1998)
7. Lee, S.C., Shim, J.S., Seo, S.W., Lim, K.S., Ko, K.R.: The accuracy of current methods in determining the timing of epiphysiodesis. Bone Jt. J. 95-B(7), 993–1000 (2013)
8. 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, 1862–1874 (2015)
9. Ritz-Timme, S., Cattaneo, C., Collins, M.J., Waite, E.R., Schuetz, H.W., Kaatsch, H.J., Borrman, H.I.: Age estimation: The state of the art in relation to the specific demands of forensic practise. Int. J. Legal Med. 113(3), 129–136 (2000)
10. 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)
11. 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)
12. Štern, D., Ebner, T., Bischof, H., Grassegger, S., Ehammer, T., Urschler, M.: Fully Automatic Bone Age Estimation from Left Hand MR Images. In: MICCAI 2014, Part II. LNCS, vol. 8674, pp. 220–227. Springer (2014)
13. Štern, D., Urschler, M.: From Individual Hand Bone Age Estimates to Fully Automated Age Estimation Via Learning-Based Information Fusion. In: ISBI (2016)