bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pediatric Bone Age Assessment Using Deep Convolutional Neural Networks

Vladimir Iglovikov[1], Alexander Rakhlin[2], Alexandr Kalinin[3], and Alexey Shvets[4]

[1] Lyft Inc., San Francisco, CA 94107, USA
iglovikov@gmail.com
[2] National Research University of Electronic Technology, Zelenograd, Moscow, Russia
rakhlin@gmx.net
[3] University of Michigan, Ann Arbor, MI 48109, USA
akalinin@umich.edu
[4] Massachusetts Institute of Technology, Cambridge, MA 02142, USA
shvets@mit.edu

Abstract. Skeletal bone age assessment is a common clinical practice to diagnose endocrine and metabolic disorders in child development. In this paper, we describe a fully automated deep learning approach to the problem of bone age assessment using data from the 2017 Pediatric Bone Age Challenge organized by the Radiological Society of North America. The dataset for this competition is consisted of 12.6k radiological images. Each radiograph in this dataset is an image of a left hand labeled by the bone age and the sex of a patient. Our approach utilizes several deep neural network architectures trained end-to-end. We use images of whole hands as well as specific parts of a hand for both training and inference. This approach allows us to measure the importance of specific hand bones for the automated bone age analysis. We further evaluate performance of the method in the context of skeletal development stages. Our approach outperforms other common methods for bone age assessment.

Keywords: Medical Imaging, Computer-aided diagnosis (CAD), Computer Vision, Image Recognition, Deep Learning

1 Introduction

During organism development the bones of the skeleton change in size and shape, and thus a difference between a child's assigned bone and chronological ages might indicate a growth problem. Clinicians use bone age assessment in order to estimate the maturity of a child's skeletal system. Bone age assessment methods usually start with taking a single X-ray image of the left hand from the wrist to fingertips, see Fig. 1. It is a safe and painless procedure that uses a small amount of radiation. The bones on the X-ray image are compared with radiographs in a standardized atlas of bone development. Such bone age atlases are based on large numbers of radiographs collected from children of the same sex and age.

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Over the past decades, the bone age assessment procedure has been performed manually using either the Greulich and Pyle (GP) [6] or Tanner-Whitehouse (TW2) [23] methods. The GP procedure determines the bone age by comparing the patient's radiograph with an atlas of representative ages. The TW2 technique is based on a scoring system that examines 20 specific bones. In both cases, bone assessment procedure requires a considerable time. Only recently software solutions, such as BoneXpert [24], have been developed and approved for the clinical use in Europe. BoneXpert uses the Active Appearance Model (AAM) [4], a computer vision algorithm, which reconstructs the contours of 13 bones of a hand. Then the system determines the overall bone age according to their shape, texture, and intensity based on the GP or TW techniques. The accuracy of this system is around 0.65 years (7.8 months). However, it is sensitive to the image quality and does not utilize the carpal bones, despite their importance for skeletal maturity assessment in infants and toddlers.

Recent advances in deep learning and its applications to computer vision allowed many researchers to drastically improve results obtained with image processing systems particularly related to medical image analyses [2]. Deep learning based approaches are gaining more attention because in a number of cases they were shown to achieve and even overcome human level performance, making end-to-end image processing automated and sufficiently fast. In the domain of medical imaging, convolutional neural networks (CNN) have been successfully used for diabetic retinopathy screening [16], heart disease diagnostics [12], lung cancer detection [1], and other applications [2]. In the case of bone age assessment, a manually performed procedure requires around 30 minutes of doctor's time per a patient. When the same procedure is done using software based on classical

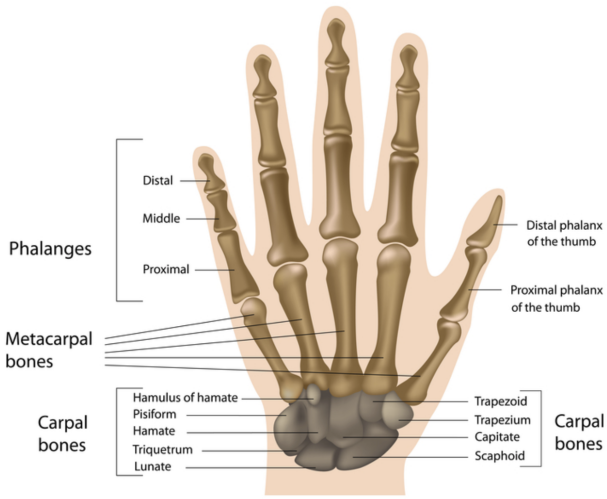


Fig. 1. Bones of a human hand and wrist (adapted from the Human Anatomy library [8]).

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

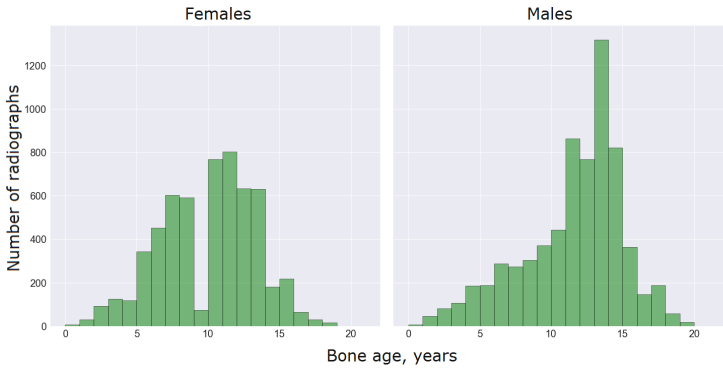


Fig. 2. Bone age distribution for females and males in the training dataset.

computer vision methods, it takes 1-5 minutes, but still requires substantial doctoral supervision and expertise. Deep learning based methods allow to avoid feature engineering by automatically learning the hierarchy of discriminative features directly from a set of labeled examples. Using a deep learning approach processing of one image typically takes less than 1 sec, while accuracy of these methods in many cases exceeds that of conventional methods. Deep neural network based solutions for bone age assessment from hand radiographs were suggested before [13, 14, 20]. However, these studies did not perform a numerical evaluation of the performance of their models using different hand bones. In addition, we find that the performance of deep learning models for bone age assessment can be further improved with better preprocessing and training networks on radiographs from scratch instead of fine-tuning from natural image domain.

In this paper, we present a novel deep learning based method for bone age assessment. We validate the performance of this method using the data from the 2017 Pediatric Bone Age Challenge organized by the Radiological Society of North America (RSNA) [18]. This data set is now freely available and can be accessed at [21]. In our approach, we first preprocess radiographs by segmenting the hand, normalizing contrast, detecting key points and using them to register segmented hand images. Then, we train several deep network architectures using different parts of images to evaluate how various bones contribute to the models' performance across four major skeletal development stages. Finally, we compare predictions across different models and evaluate the overall performance of our approach. We demonstrate that the suggested method is more robust and demonstrates superior performance compared to other commonly used solutions.

2 Preprocessing

The goal of the first step in the preprocessing pipeline is to extract a region of interest (a hand mask) from the image and remove all extraneous objects. Images were collected at different hospitals and simple background removal

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methods did not produce satisfactory results. Thus, there is a compelling need for a reliable hand segmentation technique. However, this type of algorithms typically requires large manually labeled training set. To alleviate labeling costs, we employ a technique called positive mining. In general, positive mining is an iterative procedure where manual labeling is combined with automatic processing. It allows us quickly obtain accurate masks for all images in the training set.

Overall our preprocessing method includes binary image segmentation as a first step and then the analysis of connected components for the post-processing of segmentation results. For the image segmentation, we use U-Net deep network architecture originally proposed in [17]. U-Net is capable of learning from a relatively small training set that makes it a good architecture to combine with positive mining. In general, the U-Net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. The contracting path follows the typical architecture of a convolutional network with convolution and pooling operations and progressively downsampled feature maps. Every step in the expansive path consists of an upsampling of the feature map followed by a convolution. Hence, the expansive branch increases the resolution of the output. In order to localize, upsampled features in the expansive path are combined with the high resolution features from the contracting path via skip-connections [17]. This architecture proved itself very useful for segmentation problems with limited amounts of data, e.g. see [9]. We also employ batch normalization technique to improve convergence during training [10].

In our algorithms, we use the generalized loss function

$$L = H - \log J \,, \tag{1}$$

where H is a binary cross entropy that defined as

$$H = -\frac{1}{n}\sum_{i=1}^{n}(y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)) \,, \tag{2}$$

where y_i is a binary value of the corresponding pixel i ($\hat{y}_i$ is a predicted probability for the pixel). In the second term of Eq. (1), J is a differentiable generalization of the Jaccard Index

$$J = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i\hat{y}_i}{y_i + \hat{y}_i - y_i\hat{y}_i}\right) \,. \tag{3}$$

For more details see [9].

In this work, we first manually label 100 masks using the online annotation service Supervisely [22] that takes approximately 2 min to process each image. These masks are used to train the U-Net model that then is used to perform hand segmentation on the rest of the training set. Since each image is supposed to contain only one hand, for each prediction we remove all connected components except for the largest one that is kept and further subjected to the standard hole filling protocol. This procedure allows us to predict hand masks in the unlabeled train set, however, visual inspection reveals that the quality of mask predictions is inconsistent and requires additional improvements. Thus, we visually inspect

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

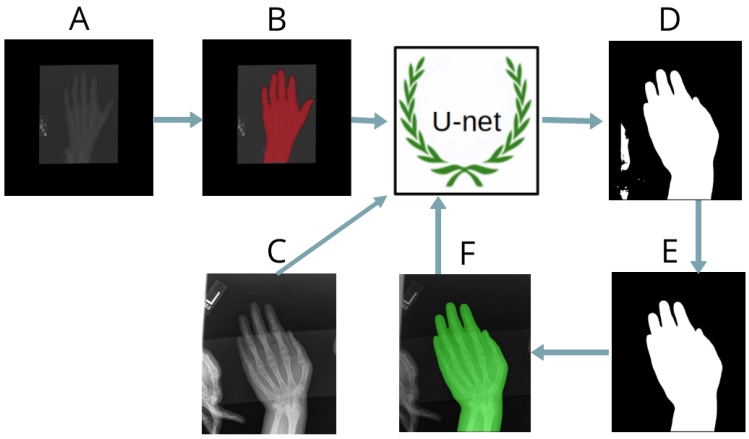


Fig. 3. Iterative procedure of positive mining utilizing U-Net architecture for image segmentation: (A) raw input data; (B) mask manually labeled with the online annotation tool Supervisely [22]; (C) new data; (D) raw prediction; (E) post processed prediction; (F) raw image with mask plotted together for visual inspection.

all predicted masks and keep only those of acceptable quality, while discarding the rest. This manual process allows us to curate approximately 3-5 images per second. Expanding the initial training set with the additional good quality masks increases the size of the labeled images for segmentation procedure and improves segmentation results. To achieve an acceptable quality on the whole training set, we repeat this procedure 6 times. Finally, we manually label approximately 100 of the corner cases that U-Net is not able to capture well. The whole iterative procedure is schematically shown in Fig. 3

Original GP and TW2 methods focus on specific hand bones, including phalanges, metacarpal and carpal bones, see Fig. 1. So, we choose to train separate models on several specific regions in high resolution and subsequently evaluate their performance. To correctly locate these regions it is necessary to transform all the images to the same size and position, i.e. to register them in one coordinate space. Hence, our model comprises two sub-models: image registration and bone age assessment of a specific region.

3 Key points detection

One of our goals is to evaluate the importance of specific regions of a hand for the automated bone age assessment. This opens remarkable opportunity of running a model on smaller image crops with higher resolution that might result in reduced processing time and higher accuracy. To crop a specific region, we have to *register* hand radiographs, or in other words, align them into a common coordinate space. To this end, we first detect coordinates of several specific key points of a hand.

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Then, we calculate affine transformation parameters (zoom, rotation, translation, and mirror) to fit the image into the desired position (Fig. 4).

Three characteristic points on the image are chosen: the tip of the distal phalanx of the third finger, tip of the distal phalanx of the thumb, and center of the capitate. All images are re-scaled to the same resolution: 2080×1600 pixels, and padded with zeros when necessary. To create training set for key points model, we manually label 800 radiographs. Pixel coordinates of key points serve as training targets for our regression model. Registration procedure is shown in Fig. 4.

Key points model is implemented as a deep convolutional neural network, inspired by a popular VGG family of models [19], with a regression output. The network architecture is schematically shown in Fig. 5. The VGG module consists of two convolutional layers with the Exponential Linear Unit (ELU) activation function [3], batch normalization [10], and max pooling. The input image is passed through a stack of three VGG blocks followed by three Fully Connected layers. VGG blocks consist of 64, 128, 256 convolution layers respectively. For better generalization, dropout units are applied in-between. The model is trained with Mean Squared Error loss function (MSE) using Adam optimizer [11]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \,. \tag{4}$$

To reduce computational costs, we downscale input images to 130x100 pixels. At the same time, target coordinates for key points are re-scaled from $[0, 2079] \times [0, 1599]$ to the uniform square $[-1, 1] \times [-1, 1]$. At the inference stage after the model detects key points, we project their coordinates back to the original image size, i.e 2080×1600 pixels. To improve generalization of our model, we apply

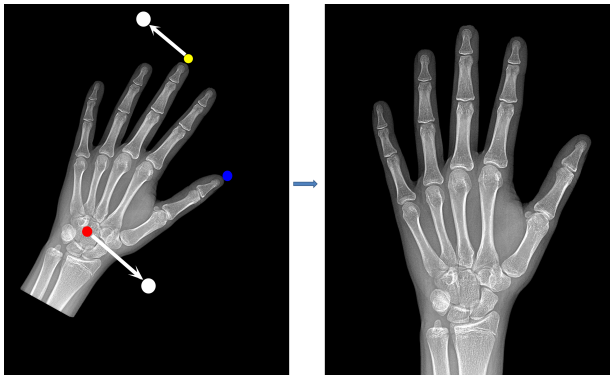


Fig. 4. Image registration. (Left) Key points: the tip of the middle finger (the yellow dot), the center of the capitate (the red dot), the tip of the thumb (the blue dot). Registration positions: for the tip of the middle finger and for the center of the capitate (white dots). (Right) Registered image after key points are found and the affine transformation and scaling are applied.

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

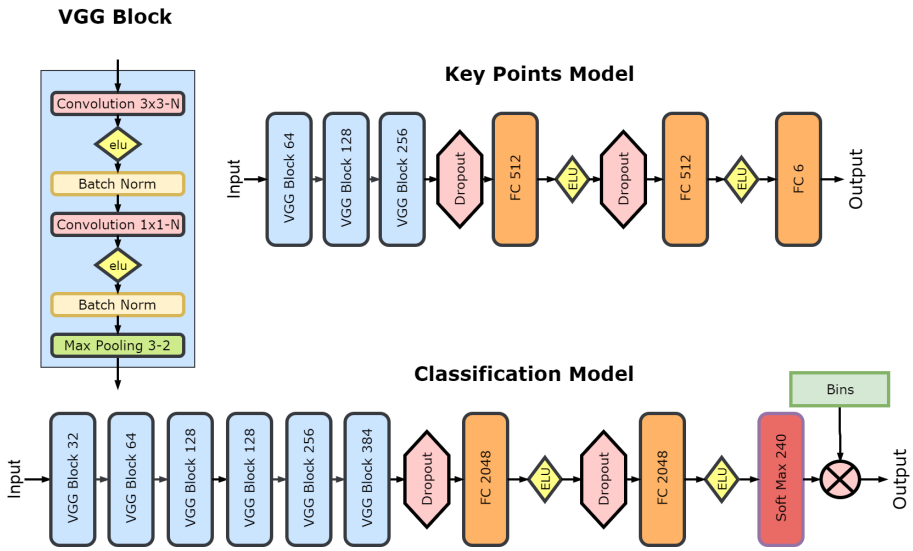


Fig. 5. VGG-style neural network architectures for regression (top) and classification (bottom) tasks.

input augmentations: rotation, translation and zoom. The model output consists of 6 coordinates, 2 for every key point.

At the next step, we calculate affine transformations (zoom, rotation, translation) for all radiographs. Our goal is to preserve proportions of an image and to fit it into uniform position such that for every image: 1) the tip of the middle finger is aligned horizontally and positioned approximately 100 pixels below the top edge of the image; 2) the capitate (see Fig. 1) is aligned horizontally and positioned approximately 480 pixels above the bottom edge of the image. By convention, bone age assessment uses radiographs of the left hand, but sometimes the images in the dataset get mirrored. To detect these images and adjust them appropriately the key point for the thumb is used. The results of the segmentation, normalization and registration are shown in the fourth row of Fig. 6.

4 Bone age assessment model

Although CNNs are more commonly used in classification tasks, bone age assessment is a regression task by nature. In order to access performance in both settings, we compare two types of CNNs: regression and classification. Both models share similar architectures and training protocols, and only differ in two final layers.

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

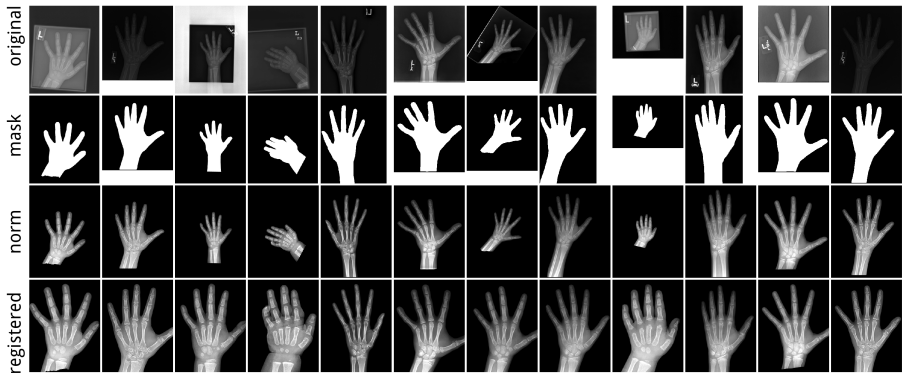


Fig. 6. Preprocessing pipeline: (first row) original images; (second row) binary hand masks that are applied to the original images to remove background; (third row) masked and normalized images; (bottom row) registered images.

4.1 Regression model

Our first model is a VGG-style CNN [19] with a regression output. This network represents a stack of six convolutional blocks with 32, 64, 128, 128, 256, 384 filters followed by two fully connected layers of 2048 neurons each and a single output (see Fig. 5). The input size varies depending on the considered region of an image, Fig. 7. For better generalization, we apply dropout layers before the fully connected layers. For regression targets, we scale bone age in the range $[-1, 1]$. The network is trained by minimizing Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{5}$$

with Adam optimizer. We begin training with the learning rate 10^{-3} and then progressively lower it to 10^{-5}. Due to a limited data set size, we use train time augmentation with zoom, rotation and shift to avoid overfitting.

4.2 Classification model

The classification model (Fig. 5) is similar to the regression one, except for the two final layers. First, we assign each bone age a class. Bone ages expressed in months, hence, we assume 240 classes overall. The second to the last layer is a softmax layer with 240 outputs. This layer outputs vector of probabilities of 240 classes. The probability of a class takes a real value in the range $[0, 1]$. In the final layer, the softmax layer is multiplied by a vector of distinct bone ages uniformly distributed over 240 integer values $[0, 1, ..., 238, 239]$. Thereby, the model outputs single value that corresponds to the expectation of the bone age. We train this model using the same protocol as the regression model.

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.3 Region-specific modelling

In accordance with the targeted features of skeletal development stages described in [5,6,23], we crop three specific regions from registered radiographs (2080×1600 pixel), as shown in Fig. 7:

1. whole hand (2000×1500 pixel)
2. carpal bones (750×750 pixel)
3. metacarpals and proximal phalanges (600×1400 pixel)

4.4 Experiment setup

We split labeled radiographs into two sets preserving sex ratio. The training set contains 11,600 images, and validation set contains 1,000 images. We create several models with a breakdown by:

1. type (regression, classification)
2. sex (males, females, mixed)
3. region (A, B, C)

Given these conditions, we produce 18 basic models ($2 \times 3 \times 3$). Furthermore, we construct several meta-models as a linear average of regional models and, finally, an average of different models.

5 Results

The performance of all models is evaluated on validation data set, as presented in Fig. 8. The leftmost column represents the performance of a regression model for both sexes. For carpal bones (region B) this model has the lowest accuracy with MAE 9.43 months. The region of metacarpals and proximal phalanges (region C) has higher accuracy with MAE equal to 8.42 months. MAE of the whole image (region A) is 8.08 months. The linear ensemble of the three regional models outperforms all of the above models with MAE 7.52 months (bottom row). This regional pattern MAE(B) > MAE(C) > MAE(A) > MAE (ensemble) is further observed for other model types and patient cohorts with few exceptions. Separate regression models for male and female cohorts (second and third columns) demonstrated higher accuracy when compared to those trained on a mixed population. The ensemble of regional models has MAE equal to 6.30 months for males and 6.49 months for females (bottom row). For males, the top performing region is the whole hand (A) that has MAE equal to 6.67 months. In contrast, for the female cohort region of metacarpals and proximal phalanges (C) has MAE equal to 6.79 months and this result is the most accurate across the three

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

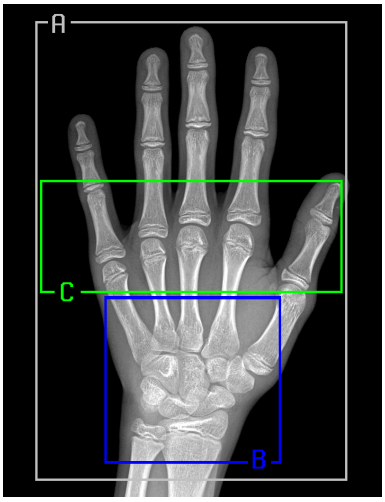


Fig. 7. A registered radiograph with three specific regions: (A) a whole hand; (B) carpal bones; (C) metacarpals and proximal phalanges.

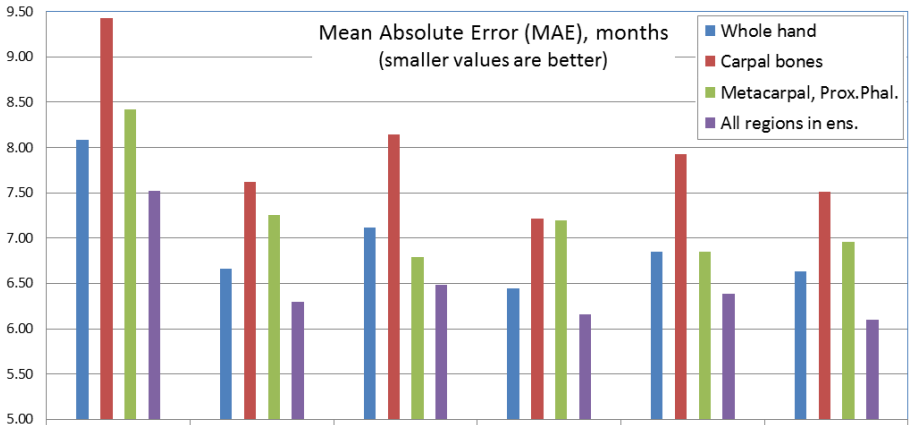
regions. Classification models (fourth and fifth columns) perform slightly better than regression networks. The ensemble of the regional models has MAE equal to 6.16 months for males and 6.39 months for females (bottom row). For the male cohort the whole hand region (A) has the highest accuracy with MAE equal to 6.44 months. For the female cohort the result produced using the metacarpals and proximal phalanges region (C) is on par with that obtained using the whole hand and has MAE equal to 6.85 months for both of them.

In the last column, we analyze the ensemble of classification and regression models. As shown, the MAEs of regional models follow overall pattern MAE(B) > MAE(C) > MAE(A). The ensemble of the regional models (bottom row) has the best accuracy with MAE equal to 6.10 month. This result clearly outperforms state-of-the-art results with MAE equal to 7.8 months for the BoneXpert software [24] and 9.84-11.16 months for the recent application of deep neural networks [14].

Next we evaluate our method in the context of the age distribution. Following [5, 14], we consider four major skeletal development stages: pre-puberty, early-and-mid puberty, late puberty, and post-puberty. Infant and toddler categories were excluded due to scarcity of the data: the development data set contained only 90 radiographs for bone age less than 24 months and only 9 of them were presented in our validation subset. The model accuracies for the four skeletal development stages, for different regions and sexes are depicted in the Fig. 9. Note that the skeletal development stages are different for males and females. Based on this data, we report two important findings.

First, unlike Lee *et al.* [14], we do not observe better results when training on carpal bones compared to other areas. With the two exceptions, the metacarpals and proximal phalanges provide better accuracy than the carpal bones do. These

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



	Regression mixed	Regression males	Regression females	Classific. males	Classific. females	Regression Classific. ensemble
■	8.08	6.67	7.12	6.44	6.85	6.63
■	9.43	7.62	8.15	7.22	7.92	7.52
■	8.42	7.25	6.79	7.19	6.85	6.96
■	7.52	6.30	6.49	6.16	6.39	6.10

Fig. 8. Mean absolute errors on the validation data set for regression and classification models for different bones and sexes. Colors correspond to different regions. Table: regions are shown in rows, models in columns. There is a total of 15 individual models and 9 ensembles.

exceptions are pre-puberty for the male cohort and post-puberty for the female cohort, where the accuracy of the carpal bones is higher. However, the validation dataset sizes for these two skeletal development stages (117 for males in pre-puberty and 16 for females in post-puberty) are too small to draw statistically significant conclusion. At the same time, we notice that Gilsanz and Ratib [5] proposed carpal bones as the best predictor of skeletal maturity only in infants and toddlers. Thereafter, we find no sound evidence to support the suggestion that the carpal bones can be considered as the best predictor in the pre-puberty, see [14].

The second interesting finding is the influence of the dataset on the accuracy of a model. For the both sexes the accuracy peaks at the late-puberty, the most frequent age in the data set. This dependency is particularly evident in the male cohort, where the accuracy essentially mirrors data distribution (see Fig. 2). This finding is very important as it suggests a straightforward way for the future improvement of the method.

6 Conclusion

In this study, we investigate the application of deep convolutional neural networks to the problem of the automatic bone age assessment. The automatic bone age

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

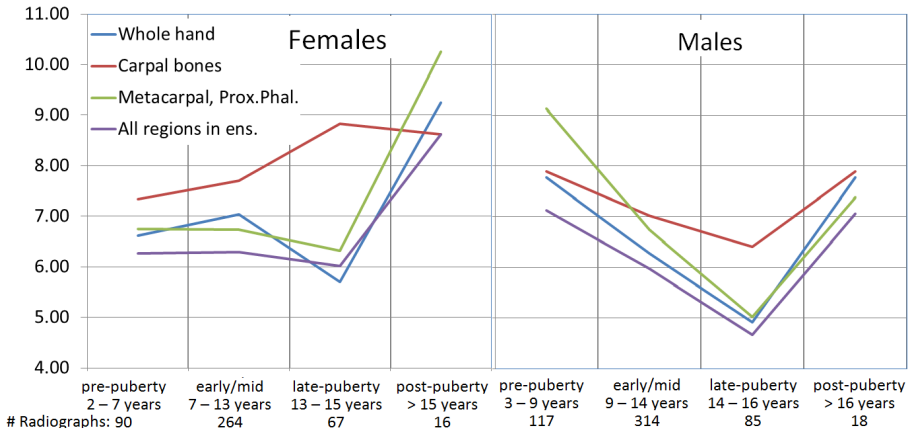


Fig. 9. Mean absolute error in months as a function of skeletal development stages for different sexes. Different colors on the plot correspond to different regions of a radiograph. For males and females the development stages are labelled at the bottom of each plot.

assessment system based on our approach can estimate skeletal maturity with accuracy similar to that of an expert radiologist and surpasses existing automated models, e.g. see [14]. In addition, we numerically evaluate different zones of a hand in bone age assessment. We find that bone age assessment could be done just for carpal bones or for metacarpals and proximal phalanges with around 10-15% increase in error compared to the whole hand assessment. Therefore, we can establish bone age using just part of the radiogram with high enough quality, lowering computational overheads.

Despite the challenging quality of the radiographs, our approach succeeds in image preprocessing, cleaning and standardization. These transformations, in turn, greatly help in improving the robustness and performance of deep learning models. Moreover, the accuracy of our approach can be improved even further. First, our solution could be easily combined with other, more complex network architectures, such as Resnet [7]. Another way to improve it is to substantially extend the training data set by additional examples. Furthermore, bone ages that serve as labels for training may also be refined based on work of independent experts. Thereby, implementing these simple steps could potentially lead to a development of a state-of-the-art bone assessment software system. It would have a potential for the deployment in the clinical environment in order to help doctors in making a final age bone assessment decision accurately and in real time, with just one click. Moreover, a cloud-based system could be deployed for this problem and process radoigraphs independently of their origin. This potentially could help thousands of doctors to get a qualified evaluation even in hard-to-reach areas.

To conclude, in the final stage of the RSNA2017 Pediatric Bone Age Assessement challenge our solution has been evaluated by organizers using the test

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

set. This data set consisted of 200 radiographs equally divided between sexes. All labels have been hidden from participants. Based on organizers' report our method achieved MAE equal to 4.97 months, which is higher compared to the performance on the data withheld from the training set. The explanation of such improvement may be hidden in more accurate labelling of the test set or better quality of the radiographs. Each of radiographs in the test data set was evaluated and crosschecked by three experts independently, compared to a one expert that labelled training radiographs. This again demonstrate the importance of the input from domain experts.

Acknowledgments The authors would like to thank Open Data Science community [15] for many valuable discussions and educational help in the growing field of machine/deep learning.

References

1. 2017 Data Science Bowl, Predicting Lung Cancer: 2nd place solution write-up, Daniel Hammack and Julian de Wit. http://blog.kaggle.com/2017/06/29/2017-data-science-bowl-predicting-lung-cancer-2nd-place-solution-write-up-daniel-hammack-and-julian-de-wit/ (2017), online; accessed December 12, 2017
2. Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Xie, W., Rosen, G.L., et al.: Opportunities and obstacles for deep learning in biology and medicine. bioRxiv p. 142760 (2017)
3. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289 (2015)
4. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Training models of shape from sets of examples. In: BMVC92, pp. 9–18. Springer (1992)
5. Gilsanz, V., Ratib, O.: Hand bone age: a digital atlas of skeletal maturity. Springer Science & Business Media (2005)
6. Greulich, W.W., Pyle, S.I.: Radiographic atlas of skeletal development of the hand and wrist. The American Journal of the Medical Sciences 238(3), 393 (1959)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Human Anatomy Library: Anatomy of the left hand. http://humananatomylibrary.com/anatomy-of-the-left-hand/anatomy-of-the-left-hand-3d-human-anatomy-model-human-anatomy-library/ (2016), online; accessed December 12, 2017
9. Iglovikov, V., Mushinskiy, S., Osin, V.: Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. arXiv preprint arXiv:1706.06169 (2017)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. pp. 448–456 (2015)
11. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Korshunova, I.: Diagnosing heart diseases with deep neural networks. https://irakorshunova.github.io/2016/03/15/heart.html (2016), online; accessed December 12, 2017

bioRxiv preprint doi: https://doi.org/10.1101/234120; this version posted December 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Larson, D.B., Chen, M.C., Lungren, M.P., Halabi, S.S., Stence, N.V., Langlotz, C.P.: Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology p. 170236 (2017)
14. Lee, H., Tajmir, S., Lee, J., Zissen, M., Yeshiwas, B.A., Alkasab, T.K., Choy, G., Do, S.: Fully automated deep learning system for bone age assessment. Journal of Digital Imaging pp. 1–15 (2017)
15. Open Data Science (ODS). https://ods.ai, online; accessed December 12, 2017
16. Rakhlin, A.: Diabetic retinopathy detection through integration of deep learning classification framework. bioRxiv p. 225508 (2017)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
18. RSNA Pediatric Bone Age Challenge. http://rsnachallenges.cloudapp.net/competitions/4 (2017), online; accessed December 12, 2017
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., Leonardi, R.: Deep learning for automated skeletal bone age assessment in X-ray images. Medical image analysis 36, 41–51 (2017)
21. Stanford University Artificial Intelligence in Medicine & Imaging: Bone age images used in the 2017 RSNA bone age challenge competition. https://aimi.stanford.edu/available-labeled-medical-datasets (2017), online; accessed December 12, 2017
22. Supervisely. https://supervise.ly/, online; accessed December 12, 2017
23. Tanner, J., Whitehouse, R., Cameron, N., Marshall, W., Healy, M., Goldstein, H.: Assessment of skeletal maturity and prediction of adult height (TW2 method). Academic Press, London (1983)
24. Thodberg, H.H., Kreiborg, S., Juul, A., Pedersen, K.D.: The BoneXpert method for automated determination of skeletal maturity. IEEE transactions on medical imaging 28(1), 52–66 (2009)