



Skeletal Bone Age Assessment Based on Deep Convolutional Neural Networks

Pengyi Hao, Yijing Chen, Sharon Chokuwa, Fuli Wu^(✉), and Cong Bai

Zhejiang University of Technology, Liuhe Road 288, Hangzhou, China
fuliwu@zjut.edu.cn

Abstract. Bone Age Assessment (BAA) is a pediatric examination performed to determine the difference between children's skeletal bone age and chronological age, the inconsistency between the two will often indicate either hormonal problems or abnormalities in the skeletal system maturity. Previous works to upgrade the tedious traditional techniques had failed to address the human expert inter-observer variability in order to significantly refine BAA evaluations. This paper proposes a deep learning method that detects and segments carpal bones as the region of interests within the left hand and wrist radiographs, and then feed the image data into a deep convolutional neural network. Tests are then made to determine whether it is more efficient to use full hand radiographs or segmented regions of interest, and also made comparisons with some CNN models. Evaluations show that the proposed method can dramatically increase the accuracy.

Keywords: Bone Age Assessment · Carpal bones region of interest Classification

1 Introduction

Bone Age Assessment (BAA) is a medical examination performed by pediatricians and pediatric endocrinologists to determine the difference between children's skeletal bone age and real age (in years). The inconsistency between the two will often indicate either hormonal problems or abnormalities in the skeletal system maturity [9, 10]. The results obtained from BAA can enable pediatricians to make the predictions such as an estimation of the required time for a child to grow, the estimated puberty age and a child's ultimate height. The assessment also gives an insight into how to monitor progress and eventually treat conditions that inhibit normal growth. Thus, BAA is a vital technique in pediatric endocrinology, orthodontics and pediatric orthopaedics for assessing childrens skeletal system maturity.

The most common and widely accepted traditional methods of performing BAA is by obtaining a radiograph of the left wrist, hand and fingers. The acquired X-ray image is then compared with samples of labelled images within a bone development standard atlas, comprising of children with the same age

and gender [2, 4, 10]. The oldest way for attaining a child's age is based on two versions of standardised main atlases; Greulich and Pyle (GP) atlas [4] and the Tanner-Whitehouse (TW) atlas [10]. While, the radiographs used in the GP method were obtained over eight decades ago, hence, it may be difficult to assess bone age accurately nowadays, While the TW is more objective than the GP [5], it takes relatively longer to perform an assessment using the TW.

BoneXpert was the first most celebrated attempt which utilizes an automated implementation for BAA, yielding a considerable high accuracy [15]. BoneXpert makes use of the active appearance model (AAM), to automatically segment 15 bones in the hand and wrist and then determine either the GP or TW bone age based on shape, intensity, and textural features. Although this approach has successfully managed to yield high accuracy over the traditional methods, it still bears some shortcomings; this system is depended on the tie between the bone age and chronological age and hence cannot give direct predictions. Previous works which used methods like canny edge detection with fuzzy classification made use of morphological features belonging to carpal bones but this method does not generalise to children above 7 years old [13, 16]. The other paper which used the method of SVD fully connected NN utilised fixed- sized feature vectors from SIFT descriptions with SVD [11], other papers used automatic carpal bone area extraction and dealing the information of carpal bone with support vector regression [8] or random forest regression [9] to do a prediction. These methods lifted the BAA to a new height of automation, however they faced data limitations for training and validation hence also lacked robustness.

Recently, with the coming of the big data age, deep learning is applied to more and more areas [6]. The medical field has also incorporated these techniques to replace some of the conventional methods, hence easing and accurately performing various applications [1, 8]. Deep CNNs have capabilities to discover multiple levels of distributed representations by learning high-level abstractions in medical image data [3]. In the field of bone age assessment, Spampinato et al. [14] tested several deep learning methods to assess skeletal bone age automatically. Lee et al. [7] mainly focused on the preprocess of DICOM images using a CNN model and then bone ages were predicted based on the common classification CNN model.

In this paper, we propose a deep learning approach which initially employs segmentation of the region of interest (ROI) consisting of carpal bones from the left-hand wrist radiographs. Secondly, we propose a detection CNN for detecting the ROI in X-ray images and then we also give a classification CNN to predict the bone age of the patient. The advantages of our method include full automation over the use of manual techniques, the utilisation of the carpal bones over the whole hand, and robustness since this evaluation is extended over the age group from 0 to 18 years. Finally, we evaluate the method using whole hand and the method using ROIs, we also compare our CNN model with the famous VGGNet-16 [12]. Analysis of bone age is a complicated process even for experts; hence this paper's main objective is to utilise an automated bone age evaluation method which could reasonably reduce the cost of assessment of bone age, by decreasing

the time that radiologists spend in predicting the bone age and also lessening the bias caused by an inter-observer.

2 Proposed Method

In this section, the details of proposed BAA method will be given. It mainly includes three parts: data preprocessing, ROIs detection and segmentation, and classification. Its flowchart is shown in Fig. 1. The original radiographs are pre-processed at first, then a detection CNN is explored to obtain carpal bones that are called ROIs in this work. After training a classification CNN, we can get predictions of bone ages.

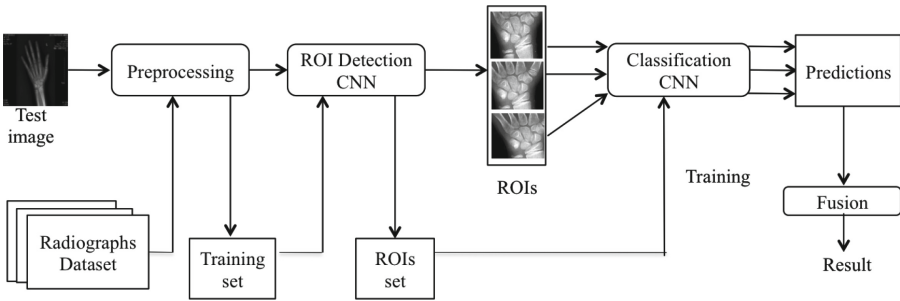


Fig. 1. The flowchart of proposed approach for bone age assessment.

2.1 Preprocessing

The original radiographs have different colour backgrounds with various sizes and the hospital annotations are situated at various positions around the radiograph. The first row in Fig. 2 shows some samples of the input radiographs used in this study. The original radiographs also have a variety of noises. These noises will give a wrong guidance in the way of learning features. So we need to preprocess them, which can not only eliminate the noises in the original images as much as possible but also reduce the image size for efficiently training. As we all know, bones are the most important part of the whole image since our target is bone age assessment. The preprocess should be able to distinguish bones from other noises. We sample pure background, background with line, background with words, bones with background and pure bones from the original radiographs, then train a network using VGG structure, with relu function and categorical crossentropy.

For every X-Ray image, we did sliding window operation with size 32×32 . Each sample patch by sliding window operation did a classification using the trained CNN model. Based on the classification result, we can get a hand bone

label-map by assigning pixels labeled as pure bones to white and other pixels to black. Since bones with background may have some unneeded information, then extracting the largest contiguous contour and filling the hand bone are needed to correct the output of model. Finally the hand bones are got from the original radiographs, some examples are shown in the second row of Fig. 2. These preprocessed images are then used as input of ROI detection CNN.

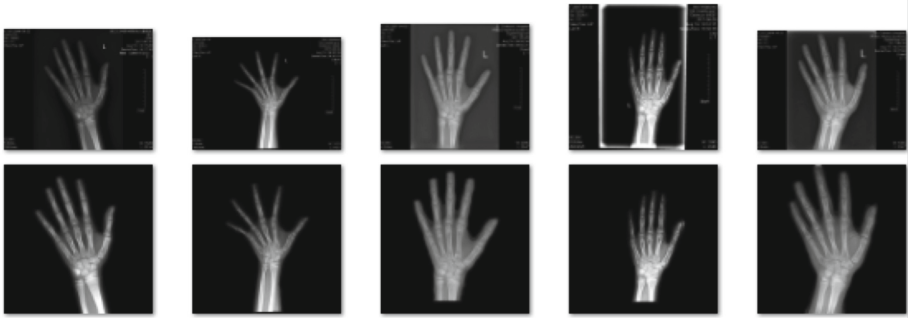


Fig. 2. Samples of radiographs. The first row shows the original radiographs. The second row shows their preprocessed results.

2.2 ROI Detection and Segmentation

Figure 3 presents the process of ROI detection and segmentation. A sliding window with a size of 100×100 is used to move across the whole preprocessed radiograph while returning us a set of 100×100 images for each radiograph. This set gives an intuition of the likely location of the ROI.

Since the upper top of the original radiographs did not contain the ROI, these were ignored, hence one 512×512 image produced 25 sub-images. Most of these slices did not include the ROI; they were either dark patches or included the phalanges, metacarpal and a large portion of the ulna and radius bones. In order to remove the dark image slices, the mean pixel intensity for each image was calculated. The average mean pixel intensity for all images was 130 but this could not be used as the intensity threshold due to the presence of some outliers, because hospitals use different radiography machines producing various intensities for the radiographs. After considering a couple of intensity thresholds, it performed better in eliminating the dark images which did not include ROI when chose threshold to be 85. After this elimination process, the resulting set still had images that did not contain carpal bones. Part of the set was then labelled as positive (containing ROI) and negative (without ROI) in order to pave way for a detection CNN for the ROI. This set was then fed into a detection CNN with the structure as shown on Fig. 4. The model achieved an accuracy of 93% in detecting the ROI. The images predicted correctly from the

detection CNN consisted of 3 ROI images of size 100×100 per one original full hand image; they were then fed into the BAA classification CNN.

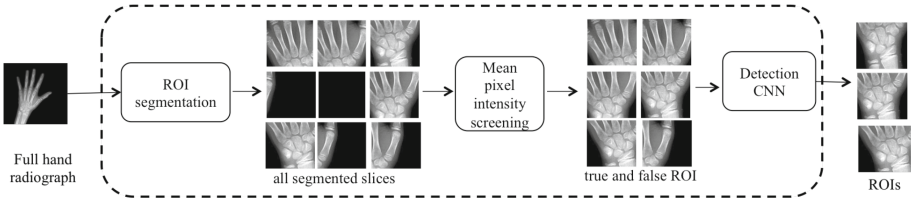


Fig. 3. The flowchart of the ROI segmentation and ROI detection process.

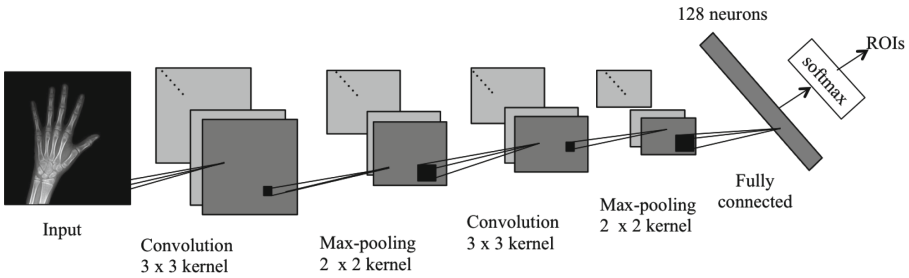


Fig. 4. ROI detection CNN architecture.

2.3 Classification CNN

We give a classification CNN here, whose structure is illustrated in Fig. 5. All the Conv Blocks are similar, with some convolutional 2D layers and ReLU activation layers, end with MaxPooling 2D layers. Fully-connected block includes a Flatten layer, two Dense layers with ReLU activation. Then through a Dropout layer and a softmax Dense layer, we can get the class label for the input. The convolution kernels are 24, 48, 96, 200 respectively, and both of the two Dense layers have 596 filters. We also explore VGGNet-16 [12] in this work. We implemented these two networks under the open source deep learning library Keras. Upon training, an SGD optimizer was used with a learning rate of 0.001. We carried out of 100 epochs with different epochs. As for VGGNet-16, the batch size is 15. For the proposed one, the batch size is 30.



Fig. 5. Structure of the simplified version of VGGNet-13.

2.4 Data Augmentation

Overfitting is a major problem in deep learning algorithms if the dataset is too small therefore data augmentation techniques were used to enlarge the dataset and increase robustness. Rotations ranging from 0° to 350° with 10° increments produced 36 synthetic images. The brightness was altered by adding an integer ranging from 10 to 70, resulting in 7 synthetic images. If there exists one pixel larger than 200 in the original X-ray image, we change for decreasing brightness by adding an integer ranging from -10 to -70 .

3 Experiments

3.1 Dataset and Evaluations

The Children’s Hospital, Zhejiang University School of Medicine of China, provided us 945 radiographs from patients with chronological age of 0–18 years, including the patients chronological age and the bone age. Figure 6 shows the bone age radiographs distribution for both male and female left hand and wrist. Due to the limitation of the dataset size, gender was not considered as this would further reduce the dataset.

For evaluations, we will compare two methods and two classification models. In the first method, noted as M1, the whole hand that encompassed of all the left hand wrist bones (i.e. phalanges, metacarpal, carpal, ulna and radius bones) was preprocessed as shown in Sect. 2.1 and then used as input image to the classification CNN shown in Sect. 2.3. Images of two different sizes 512×512 and 224×224 were then used as the model input to the classification CNN. The second method is denoted as M2 that is described in Sect. 2, which mainly used the automatically cropped ROIs, then ROIs were used as the model input to the classification CNN. Fusion rules (i.e. the average rule and the maximum rule) were employed for calculating the accuracy of the second method (M2) because 3 ROI images were present for each original full hand radiograph. M2’s CNN network gave 3 predictions for the 3 singular ROI, thus when using the average rule the mean of these 3 predictions is the final result; and when using the maximum rule the 3 predictions were evaluated to find the most frequent prediction which would then be the final prediction. At the same time, we will also compare the performance of our proposed classification model given in Sect. 2.3 and the commonly used VGGNet-16 using different sizes of inputs.

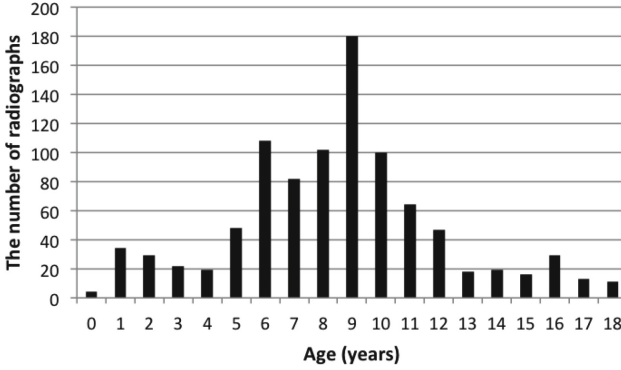


Fig. 6. Bone age radiographs distribution in the used dataset.

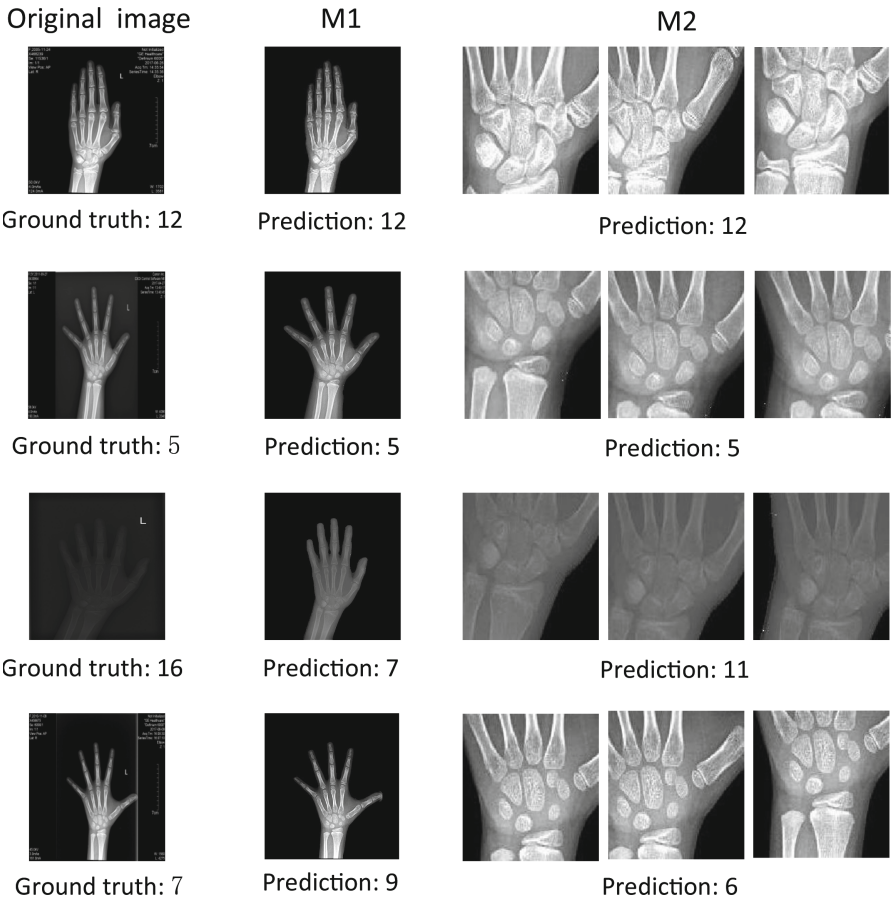
3.2 Results

Initially 140 radiographs from the dataset were randomly chosen encompassing of the age from 0 to 18 years. These images were used for testing models. The accuracy results obtained from the methods are shown in Table 1. Firstly, the results show that VGGNet-16 can get a little higher accuracy of totally correct than our model under the M1 method no matter with the input size of 512×512 or 224×224 . But our model performed a litter better than VGGNet-16 when radiographs were assigned an age within 1 year of ground truth. By using M2, our model got much higher accuracy than VGGNet-16. Our model can assign 43.3% radiographs to be the correct age, and 74.2% radiographs were assigned an age within 1 year of ground truth and 88.1% were assigned an age within 2 years of ground truth. What is more, no matter VGGNet-16 or our model, using the input with 512×512 performs better than the input size of 224×24 ; the results also show that utilising automatically segmented ROIs is a viable method for significantly improving the accuracy in comparison to using the full hand; the results attained portray that the average rule outperforms the maximum rule. The accuracy of totally correct for each model and each method is not very high. The reason may be that the age groups from 0 to 4 years have bad preprocessing result which restricts the system from being applicable to all age groups. The other reason is that the imbalance in the age groups distribution in the dataset, for example there are more radiographs for the age groups from 6 years to 11 years in contrast to the age groups from 12 years to 18 years.

Sample predictions achieved by the our classification CNN are shown in Fig. 7 illustrating the ground truth versus the model's prediction. Here, the predictions of M2 used average rule. In Fig. 7, the first two rows are correctly predicted, both methods can get the same bone age with ground truth. The last two rows are wrong predictions that may due to the presence of soft tissue reduced the hand-to-background ratio. Using a machine with a GPU of Nvidia GeForce GT 730M 2G, it averagely takes one and half minutes from reading an original radiograph to get the final bone age assessment, which is much more efficient than traditional methods.

Table 1. Comparison between different methods using different classification models in terms of accuracy.

	Type	Correct	± 1 year	± 2 years
VGGNet-16	M1 (224×224)	0.371	0.571	0.779
	M1 (512×512)	0.464	0.641	0.771
	M2 (Maximum)	0.411	0.645	0.834
	M2 (Average)	0.428	0.684	0.862
Proposed CNN	M1 (224×224)	0.364	0.593	0.721
	M1 (512×512)	0.414	0.673	0.712
	M2 (Maximum)	0.412	0.698	0.844
	M2 (Average)	0.433	0.742	0.881

**Fig. 7.** The examples of bone age assessments using two methods based on our classification CNN.

4 Conclusions

In this paper, we gave a deep-learning method to automatically detect and segment carpal bones as a region of interest. We also gave a classification CNN for BAA evaluation. This work made a comparison of the use of full hand and wrist and the carpal bones while utilising a data driven approach with a classification CNN. The technique proposed in this study has great capability to yield a much higher accuracy result however the presence of a number of limiting factors restrained its performance, like the amounts of data. Given a larger training dataset this accuracy can be improved.

Acknowledgments. The work is supported by Zhejiang Provincial Natural Science Foundation of China under grants No. LY18F020034, LY18F020032, and National Natural Science Foundation of China under grants No. 61502424 and No. 61801428 and partially supported by the Ministry of Education of China under grant of No. 2017PT18 and the Zhejiang University Education Foundation under grant of No. K18-511120-004 and No. K17-511120-017. This work is also supported by The Research of Real Doctor AI Research Center.

References

1. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans. Med. Imaging* **35**(5), 1207–1216 (2016)
2. Gaskin, C., Kahn, S., Bertozzi, J., Bunch, P.: *Skeletal Development of the Hand and Wrist: A Radiographic Atlas and Digital Bone Age Companion* (2011)
3. Greenspan, H., van Ginneken, B., Summers, R.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging* **35**(5), 1153–1159 (2016)
4. Greulich, W., Pyle, S.: *Radiographic Atlas of Skeletal Development of the Hand and Wrist*. Stanford University Press, Stanford (1959)
5. Khan, K., Elayappen, A.: Bone Growth Estimation Using Radiology (Greulich-Pyle and Tanner-Whitehouse Methods). In: Preedy, V. (ed.) *Handbook of Growth and Growth Monitoring in Health and Disease*. Springer, New York (2012). https://doi.org/10.1007/978-1-4419-1795-9_176
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436 (2015)
7. Lee, H., et al.: Fully automated deep learning system for bone age assessment. *J. Digit. Imaging* **30**(4), 427–441 (2017)
8. Liskowski, P., Krawiec, K.: Segmenting retinal blood vessels with deep neural networks. *IEEE Trans. Med. Imaging* **35**(11), 2369–2380 (2016)
9. Bonawitz, C.A.: *Practical pediatric imaging: diagnostic radiology of infants and children*. *Acad. Radiol.* **6**(10), 607 (1999)
10. Poznanski, A.: Assessment of skeletal maturity and prediction of adult height (TW2 method). *Am. J. Dis. Child.* **131**(9), 1041–1042 (1977)
11. Seok, J., Hyun, B., Kasa-Vubu, J., Girard, A.: Automated classification system for bone age x-ray images. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 208–213. IEEE (2012)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *ICML* (2015)

13. Somkantha, K., Theera-Umpon, N., Auephanwiriyaikul, S.: Bone age assessment in young children using automatic carpal bone feature extraction and support vector regression. *J. Digit. Imaging* **24**(6), 1044–1058 (2011)
14. Spampinato, C., Palazzo, S., Giordano, D.: Deep learning for automated skeletal bone age assessment in x-ray images. *Med. Image Anal.* **36**, 41–51 (2016)
15. Thodberg, H.H., Kreiborg, S., Juul, A., Pedersen, K.: The bonexpert method for automated determination of skeletal maturity. *IEEE Trans. Med. Imaging* **28**(1), 52–66 (2009)
16. Zhang, A., Gertych, A., Liu, B.J.: Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones. *Comput. Med. Imaging Graph.* **31**(4–5), 299–310 (2007)