Bone Age Prediction Method Based on Convolutional Neural Network

Bin Hu, Lidan Shang, Feng Cheng, Hai Sheng, Jiayan Pan and Dong Yin

1 School of Information Science and Technology, University of Science and Technology of China, Jin Zhai Road 96, Hefei, Anhui 230026, China
2 Key Laboratory of Electromagnetic Space Information of CAS, Hefei, Anhui 230027, China
3 Hefei Angao Information Technology Co., Ltd., Block A 803, Hai Guan Road 9, High-tech Zone, Hefei, Anhui, China
4 First People Hospital of Wuhu City, Chizhu East Road 1, Jiujiang District, Wuhu, Anhui, China
5 Author to whom any correspondence should be addressed. E-mail: yindong@ustc.edu.cn

Abstract. How to accurately observe whether children are developing normally has always been one of the focuses of medical research. Based on the convolutional neural network, this paper proposed a bone age prediction method. Firstly, the hand bone data set was constructed. Secondly, the prior frame required by the neural network was obtained by K-means clustering. Finally, fine-tuning, training, detecting and cropping were performed to achieve bone age classification. The experimental results show that our method achieves the smallest average error and has strong practicability.

1. Introduction

Bone age is the abbreviation for bone age measurement. It is the best indicator of the maturity of adolescent bones. It reflects human body development more accurately than age. With the increase of age, the shape and size of human bones will change accordingly. If there is a difference between bone age and actual age, it implies some growth and development diseases. Bone age can be used for the diagnosis of diseases in children's growth and development and the prediction of height in adulthood. At the same time, in the sports field, bone age prediction can further prevent the phenomenon of age falsification in youth competitions, and can also be used in the selection of athletes and subsequent training guidance.

In the past few decades, the methods recognized internationally for assessing bone age mainly included atlas method and scoring method. The atlas method was mainly the GP atlas method[1], which compared the photographer's X-ray film with typical X-ray films at different ages, and then estimated bone age. The scoring method was mainly the TW3 scoring method[2], which was based on a scoring system that scored 20 special bones and calculated the total score, and then obtained the bone age measurement value. These methods required considerable time and were greatly affected by the subjectivity of the physician. And, bone age experts were extremely scarce, doctors were often required to read a large number of films in a short time, which further affected the accuracy. So, the intelligent bone age assessment could allow artificial intelligence to undertake the repetitive work of reading images and could be completely objective. Thereby, assisting doctors or even replacing
doctors for medical diagnosis, liberating doctors from tedious reading tasks, and releasing high-quality medical treatment are imperative. It will improve the accuracy and efficiency of medical diagnosis.

2. Related Work
BoneXpert[3] was an outstanding representative of intelligent bone age assessment, which had been accepted and used by some medical institutions in Europe. BoneXpert was based on a computer vision algorithm -- the active appearance model[4], by reconstructing the contours of 20 special bones of the human hand, and based on the GP map method or TW scoring method, according to the bone texture, strength, shape and other characteristics, to assess the bone age. However, BoneXpert was extremely sensitive to image quality, and did not use the information of the carpal bone, which had certain limitations.

Research on automatic bone age assessment based on deep learning began to appear in 2016. Researchers at the University of Catania in Italy proposed a 6-layer bonet[5] structure and corrected the position of hand bone images to complete the end-to-end training. However, due to the shallow network layers, the learning ability was not good. Harvard University Hyunkwang Lee1, Shahein Tajmir and others fine-tuned GoogleNet8[6], and performed pre-processing, image segmentation, image classification and other operations on bone pictures, and finally obtained a bone age assessment accuracy rate of 57.31% for female and 61.40%[7] for male. In 2018, Lyft, Alexander Rakhlin of the University of Massachusetts Institute of Technology and the University of Michigan, Vladimir Iglovikov and Alexey Shvets and others used a variety of convolutional neural networks, first used U-net[8] to segment the hand bone picture, and considered different hands. The influence of the bone area on the evaluation results of bone age, considering the regression and classification model comprehensively, finally obtained the result of the measured value error of 6.10 months[9]. In 2019, the Te Radiological Society of North America (RSNA) conducted a summary of its RSNA Pediatric Bone Age Machine Learning Challenge. By providing 14,236 radiographs of hand bones, the competition attracted 260 individuals or with the participation of the team, the average absolute errors of the top 5 winners of the competition were 4.2, 4.4, 4.4, 4.5, 4.5 months[10]. Compared with the traditional method, the bone age assessment system based on deep learning showed better performance in efficiency and accuracy.

Similarly, we worked with bone age prediction experts to carry out more accurate bone age prediction.

3. Our Method

3.1. TW3 Method
The TW3 method had the nature of not being affected by race, age, and region. The specific process of TW3 bone age scoring method was: (1) took X-ray film on the wrist of the child's left hand; (2) judged the degree of bone development of some key bones in the X-ray film. In the TW3 bone age scoring method, 13 hand bone skeletons that needed to be analyzed were shown in figure 1.

![Figure 1. Left hand X-ray film.](image)

TW3 staged bone development to determine the degree of bone development. It was usually divided into 8-9 stages, which were represented by capital letters A-I.
In order to predict the bone age more accurately, we divided TW3 into more detailed. In our work, according to the small differences in bone development at the C level, it was divided into four levels C-, C, C1, and C2, therefore, were also called quartiles. The score of the bone age development degree was to grade the 13 bones marked with numbers in figure 1 for each grade, and each grade corresponded to a corresponding score. The corresponding scores obtained were summed and compared with the score table. The corresponding bone age value could be obtained.

3.2. CNN
Convolutional Neural Network (CNN) was a special form of feed-forward neural network. It was influenced by the animal’s visual receptive field mechanism and had a local connection and weight sharing mechanism. Convolution was a common operation method, and its physical meaning was the weighted sum of one function on another function. Its mathematical expression is shown in Formula (1).

\[ y(n) = \sum_{i=-d}^{d} x(i)h(n-i) = x(n) * h(n) \]

Where, \( x(n) \) is the input signal and \( h(n) \) is the convolutional variable sequence.

The layers in a convolutional neural network could be roughly divided into an input layer, an output layer, a convolutional layer, a pooling layer, and a fully connected layer. The activation function was unified by using the rectified linear unit (Relu) function, as shown in Formula (2).

\[ F(X) = \max(0, W * X + B) \]

Where \( F(X) \) is the output, \( W \) is the convolution kernel weights, \( X \) is the input, and \( B \) is the bias.

In image processing, the image convolutions were all discrete convolutions, and the essence of discrete convolution was linear transformation, that was, the convolution kernel with the same parameters was applied to each small block of the image. The operation of the convolution layer was similar to a sliding window sliding through the picture in sequence. The convolution kernel was multiplied by the value at the corresponding position of the picture and accumulated, and the accumulated value was used as the output of this position.

The pooling layer used a certain value in the region to represent the entire region. It could perform feature selection on the feature map and gradually reduced the size of the feature map, thereby effectively reducing the parameters in the convolutional neural network and reducing the operation during the network learning process. The amount could also be mitigated by overfitting. Common pooling layers included the largest pooling layer and the average pooling layer. Maximum pooling was to select the maximum value in the subregion, and average pooling was to calculate the average value of the subregion.

3.3. Presented Method
The bone age assessment model based on the convolutional neural network proposed in this paper was mainly constructed using YOLOV3[11] and ResNet[12], and the two were connected through the intermediate cropping module to form the YOLOV3-SC-ResNet model. The strategy adopted was that first step was the target detection of the key area of the X-ray film of the hand bone image (YOLOV3), then was the bone of the key area of the detected hand bone cut out from the X-ray film (SC), and the samples were obtained for rating. In the grade evaluation stage, this paper used ResNet as the basis to determine the grade of each bone. In order to improve the accuracy of the evaluation, the loss function of the classification stage was optimized and adjusted. Finally, the grade of each bone and the evaluation value of bone age could be obtained. The proposed overall framework is shown in figure 2.
1) Generating a priori box

In the key region detection model based on neural network, it needed to provide the size of the prior frame to the network model. In this paper, the size of the prior frame was obtained by K-means clustering in machine learning. Before performing K-means clustering, the xml file generated by labelImg needed to be converted into a txt file.

2) Fine-tuning

Due to the heavy workload of labeling pictures, this paper only marked 190 X-rays of hand bones. In order to solve the problem of small data sets, based on the idea of transfer learning, this paper adopted a fine-tuning method. Fine-tuning was mainly achieved by loading pre-training weights.

3) Training

Considering the calculation overhead and the problem that the size of the input image should meet an odd multiple of 32, the size of the image was uniformly 1056*1056, the batch size was set to 2 during training, the epoch was 500, and the initial learning rate was set to 0.001. During the training phase, the change of the loss value was constantly detected, and the fluctuation of the loss value was recorded. During the training process, it could be found that when training reached 90 times, the loss began to gradually stabilize.

4) Testing output

In the testing stage, the remaining hand bone images to be tested were sent to the network, and the trained weights were loaded at the same time. The network model was used to return the position of the frame and drew it on the original image to achieve the final detection output.

5) Cropping

In view of the YOLOV3-SC-ResNet model proposed in this paper, in the second stage, the image of the key bones in the hand bones needed to be input to the ResNet network, so on the basis of the output image detected in the previous step, the key area must be cropped. It mainly used the position and score of the frame returned by the network to perform cropping. When a type of bone was repeatedly detected, a detection frame with high confidence was selected for cropping.

6) Classifying bone level

According to the actual development of boys and girls, their bone levels needed to be classified separately. For the girl's hand bone picture, according to the principle of quartile, its 13 bones needed to be divided into 358 categories in total, the input images were all the bone images, and then a classifier with a total of 358 categories was trained. The same was true for boys. In the picture of hand bones, its 13 bones needed to be divided into 347 categories. In the classification process, the size of all kinds of skeleton pictures obtained by cropping was standardized to 224*224, the pre-processed pictures were used as the input of ResNet, the epoch was 200, and the batch-size of each batch was 64. In the training process, it could be found that when training reached 30 times, the loss began to stabilize gradually.

4. Experimental and Analysis

4.1. Experimental Environment

- Operating system: Ubuntu 16.04
- Language: Python 3
- Frame: Keras+Tensorflow
- Computing resources: TITAN Xp
4.2. Hand Bone Data Set
The data used in this paper were left-handed X-ray pictures of Chinese girls and boys aged 3-14. Among them, there were 4,710 pictures of girls' hand bones and 2,332 pictures of boys' hand bones. In the target detection stage, the training data was 190 hand bone pictures marked manually, and the key bone position detection was performed on the remaining pictures, then the key bone position was cropped and the subsequent level classification was performed. The original image of the hand bone data and the marked result graph are shown in figure 3(a) and figure 3(b) respectively.

![Figure 3(a)](source_of_hand_bone_image)
![Figure 3(b)](image_of_hand_bone_labeled_with_label_img)

In the bone classification stage, for the hand bone data of girls, 60252 key regions were segmented through target detection, and then 8614 pieces were randomly selected as the verification set and 43074 pieces were used as the training set. In the test phase, 667 hand bone images were used to pass the target 8564 key regions were detected and segmented, which were used as the test set. For boy hand bone data, 29831 key regions were segmented through target detection, and then 4369 sheets were randomly selected as the verification set and 21086 sheets as the training set. 348 hand bone images were used to segment 4376 key regions through target detection, which were used as the test set.

4.3. Analysis
During the target detection process, the loss during the training phase is shown in figure 4(a), and the detection schematic diagram in the test phase is shown in figure 4(b).

![Figure 4(a)](target_detection_stage_loss)
![Figure 4(b)](schematic_diagram_of_detection_results)

In the detection stage, the recall rate was 98.4%, and the accuracy rate was close to 100%, that was the object with almost no detection errors, and 98.4% of the areas that needed to be detected were detected by the model. In the bone classification stage, 667 images of girls' hand bones and 348 images of boys' hand bones were tested to obtain the grade of each bone in the picture and calculated the bone age obtained by network evaluation.

Finally, on the test set, the accuracy rate of the top-1 classification of girls' bone levels was more than 40%, the accuracy of top 5 was more than 90%, and the accuracy rate of top 1 classification of
boys’ bone levels was more than 42%, top 5 was above 92%. At the same time, for each hand bone picture, the 13 bone level evaluation values obtained by the network were added, and the bone age was obtained by comparing the level sum-bone age correspondence table, considering that only 13 bones were detected (in actual use, if there is a missed detection phenomenon, you can manually cut the missed detection area, this assumption is reasonable in actual circumstances), and finally realized the average absolute error of the girl’s bone age evaluation value and the actual bone age was 0.245 years (2.94 months), the average absolute error between the boy’s bone age assessment and the actual bone age was 0.365 years (4.38 months).

5. Conclusion
In this paper, TW3 bone age assessment method was used to further subdivide bone grades. Although the accuracy of top 1 is not high in the classification stage, the average absolute error of the final bone age assessment was extremely small because the grade division was sufficiently detailed and state-of-the-art.

However, it should be recognized that there were still some problems in this experiment, which needed to be improved in the following aspects: (1) Increasing the amount of data, compared with other related research, the amount of data used in this experiment was small, so there were still larger room for improvement; (2) The performance of the detection stage would greatly affect the final bone age evaluation value. There were still missing detection phenomena in the experiment, which would have a greater impact on the final bone age evaluation.

6. Acknowledgments
This work was supported by Hefei Angao Information Technology Co., Ltd. Data was provided by the First People's Hospital of Wuhu City.

7. References
[1] Greulich W W and Pyle S I 1959 J. Radiographic Atlas of Skeletal Development of The Hand and Wrist. The American Journal of The Medical Sciences, 238(3): p 393
[2] Malina R M and Beunen G P 2002 J. Assessment of skeletal maturity and prediction of adult height (TW3 method). American Journal of Human Biology, 14(6): p 788-789
[3] Thodberg H H, Kreiborg S, Juul A, et al. 2009 J. The BoneXpert method for automated determination of skeletal maturity. IEEE transactions on medical imaging, 28(1): p 52-66
[4] Cootes T F, Taylor C J, Cooper D H, et al. 1992 J. Training models of shape from sets of examples. Proceedings of the British Machine Vision Conference, 9: p 9-18
[5] Spampinato C, Palazzo S, Giordano D, et al. 2017 J. Deep learning for automated skeletal bone age assessment in X-ray images. Medical Image Analysis, 36: p 41-51
[6] Szegedy C, Liu W, Jia Y, et al. 2015 Proc. Int. Conf. On Computer Vision and Pattern Recognition p 1-9
[7] Lee H, Tajmir S, Lee J, et al. 2017 J. Fully Automated Deep Learning System for Bone Age Assessment. Journal of Digital Imaging, 30(4): p 427 – 441
[8] Ronneberger O, Fischer P and Brox T. 2015 J. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv: 1505.04597v1
[9] Iglovikov V, Rakhlin A, Kalinin A, et al. 2018 M. Pediatric Bone Age Assessment Using Deep Convolutional Neural Networks. Springer, Charm, p 300-308
[10] Halabi S S, Prevedello L M, Kalpathy-Cramer J, et al. 2019 J. The RSNA pediatric bone age machine learning challenge. Radiology, 290(2): p 498-503
[11] Redmon J and Farhadi A. 2018 J. YOLOv3: An Incremental Improvement[J]. arXiv preprint arXiv: 1804.02767.
[12] He K, Zhang X, Ren S, et al. 2016 Proc. Int. Conf. On Computer Vision and Pattern Recognition p 770-778