

Application of Deep learning in Bone age assessment

Yagang WANG, Qianni ZHANG, Jungang HAN and Yang JIA

School of Computer Science, Xi'an University of Posts and Telecommunications,
Shaanxi710121, China

Yagang Wang, Qianni Zhang, Jungang Han, Yang Jia; e-mail:
zhangqianni187@163.com

Abstract. Bone age assessment is a common method for evaluating the growth and development of children and adolescents. It can be used for diagnosing problems such as rickets and short stature during the growth of adolescents. The traditional bone age assessment method is the doctor's manual treatment of hand bone X-ray images, and comparing with medical standard pictures to achieve bone age assessment. In order to reduce the workload of doctors in identifying X-ray images and the subjective effect of doctors, the method is proposed by combining deep learning with medical images to implement automatic bone age assessment. Different methods are used to segment the hand bone X-ray images and transfer learning for classification and identification of bone age. The test results show that female bone age test precision were assigned 94.4% within 20 months, male test precision were assigned 90.5% within 20 months.

1.Introduction

Bone age assessment plays an important role in the growth and development of children and adolescents [1-3]. One of the earliest methods of measuring bone age was to count the number of ossification centers and compare them with charts. The other most commonly used methods were the Greulich and Pyle (GP) or Tanner-Whitehouse (TW2) methods. These two methods were based on the stage of bone development, and determined by searching in the evaluation criteria of bone age, followed by the Chinese skeleton maturity evaluation standard- CHN method [4-6]. Currently, these methods were used by doctors to observe the development degree of the hand bone X-ray images, and compare it with the standard graph, so as to determine the bone age, which takes a lot of time. In order to save labor as much as possible, there is a bone age measuring device on the market. Although this device saves a lot of labor, it uses printers, scanners and professional display screens, which requires high hardware [7]. Medical images have not only been analyzed by manual and medical devices, but also combined with deep learning, which has a good development prospect in bone age assessment [8-11].

This paper applies the method of deep learning to bone age assessment. The overall frame of bone age assessment is shown in Figure.1. First, hand bone X-ray images of children and adolescents are collected and preprocessed. Then, LeNet-5 and U-Net are used to segment the hand bone X-ray images. It avoids the time and resource waste of manual treatment, and finally uses transfer learning to classify bone age. Deep learning reduces the burden on doctors and improves the accuracy of bone age assessment.



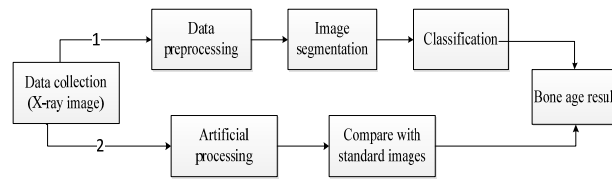


Figure 1. Bone age assessment frame

The traditional method is the part 2 and the automatic method adopted in this paper is the part 1 marked in Figure 1.

2. Data collection

The data come from the hand bone database and competition research, mainly including the X-ray images of 0-18 years old teenagers, which are stored in the form of “png”. All X-ray images are from the left hand. The main reason is that the commonly used hand is the right hand, and the daily life and work will directly affect the shape of bone, the possibility of causing greater damage. So the left hand is more appropriate. X-ray images of 5-18 years old adolescents were selected for operation, because the bone development of 0-4 years old children was not stable and had poor reference value.

3. Data preprocessing

Data preprocessing stage in the raw data is inconsistent data size, noisy and high dimension problem, main is to accomplish data standardization/normalization, containing ways such as geometric transformation (translation, scaling, rotation, etc.), pixel transform (the contrast and brightness transform, channel, transform, etc.), gray transformation, etc.

There are two main problems in the X-ray images. First of all, the images are larger than 1000*1000 pixels and have different sizes. As the input images of the segmentation and classification neural network are less than or equal to the input size of the images, it is necessary to standardize the images. In order to make the processed images clearer and not loss precision, adjust the images to 512*512 size under the condition of maintaining the original horizontal/vertical ratio, remove the excess places on the edge, and use zero-padding for the missing places (the value of filling is the value of the outer boundary of the images). Another problem is the lack of contrast between the background and the foreground of some pictures. In order to improve the visual effect of the pictures, the foreground is enhanced and the background is suppressed. In order to ensure the picture is true, the contrast of the pictures is enhanced.

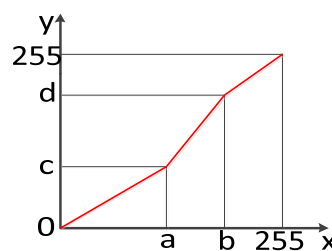


Figure 2. Linear transformation

$$y = \begin{cases} \left(\frac{c}{a}\right) * x & , x < a \\ \left(\frac{d-c}{b-a}\right) * (x - a) + c & , a < x < b \\ \left(\frac{255-d}{255-b}\right) * (x - b) + d & , x < b \end{cases} \quad (1)$$

In the Formula (1), where a and b are the transformation range, and d and c determine the slope of linear transformation ($a < b$, $c < d$ guarantee the function increase and ensure that the gray scale does not reverse).

Through the above methods, the data preprocessing process is completed, and the preprocessing results are applied to image segmentation and classification.

4.Segmentation

According to the data preprocessing procedure, the images still contains noisy (patients' information will be marked on the X-ray images in general medical pictures). The area of interest for bone age assessment is the part of the hand bone, so the hand bone in the images needs to be segmented. In order to be faster and more efficient segment 512*512 images, there are two methods: LeNet-5 network (based on CNN) and U-Net network (based on FCN).Through these two methods to segment effectively the images.

4.1 LeNet-5 network based on CNN

LeNet-5 network architecture as shown in Figure 3. Due to network input to 32 * 32 of the size of sample, firstly needs to deal with the size of 512 * 512 images. Matlab mouse operation to select the hand part of the images (including bone and tissue), get the rectangle width not less than 32 pixels to select all of the rectangular boxes which are cut into the size of 32 * 32 samples. As shown in Figure 4, the first row represents the selected 32*32 size bone sample, and the second row represents the selected 32*32 size tissue sample.

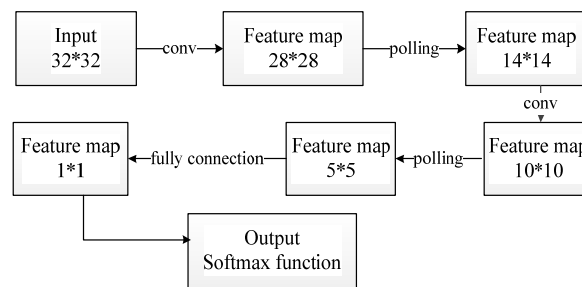


Figure 3. LeNet-5 network architecture

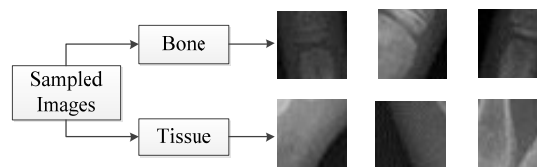


Figure 4. Sampled images

The samples were trained using LeNet-5 network to distinguish the hand from the non-hand parts. First, the sampled data that needs to be input into the network is converted into tfrecord format, which is faster than the original method. Then, it may be part of the hand to use the sliding window to perform scanning evaluation on the standardized picture. In the process of using the sliding window, multiple windows may be included or crossed. Since there are some non-hand areas in the selected areas of the X-ray images through the test, the maximum connected region algorithm is adopted to determine the part of the hand to achieve the images denoising effect.

4.2 U-Net network based on FCN

The U-Net network architecture is shown in Figure 5 [12]. In order to verify the particularity of U-Net network in the medical field, FCN network obtained through fine-tuning VGG-16 which was first used to train images [13]. FCN is characterized by deconvolution of the map at the last layer of full convolution, but the result after segmentation is not obvious enough. U-Net is adopted for segmentation. Its biggest characteristic is that the network structure is “U” (the process of down-sampling and up-sampling), and it has symmetry. The size of the images is halved through two convolution operations, the parameters are reduced by pooling, and the same mapping layer is reproduced on the basis of the FCN network structure (shown as copy and crop in Figure 5). In order to improve the effect of network training, geometric transformation (for example, rotation, contrast, etc.)

was adopted to enhance the images. During training, 512×512 images were transformed into npy format (including pictures and labels), and npy files corresponding to the test images were obtained and transformed into pictures.

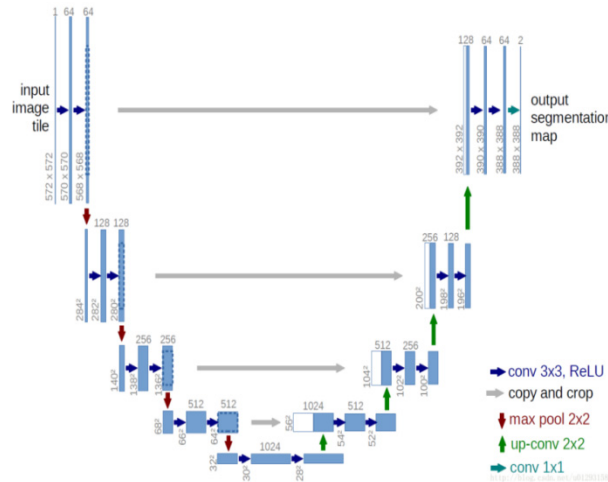


Figure 5. U-Net network architecture

Binary the image obtained, as shown in Figure 6. Add the processed image to the original image to ensure that only the hand parts appear on the original image. Segmentation results are shown in Figure 7.



Figure 6. Binary image



Figure 7. Segmentation results

4.3 Network comparison

Both of the above two networks are implementing hand bone segmentation. By comparing the two networks, the results are shown in Table 1:

Table 1. Network model comparison

	Input size	Structural Characteristics	Major time consuming
LeNet-5	32*32	Simple(5-layers)	Sliding window (size:32*32)
FCN	Any size	VGG(16-layers)	Training net

U-Net	512*512	Symmetry	Training net
-------	---------	----------	--------------

The classification criteria of bone age are mainly reflected in the differences in the areas concerned by the opponent bone. In order to classify the hand bone parts in the X-ray images better (highlight the bone part and weaken the other tissue parts), use partial region enhancement.

5. Classification

Transfer learning was used to classify bone age. It applies a well-trained model to a new problem through simple adjustment. In this paper, the model of google trained general images are transferred to the medical images for bone age assessment. Because the complex network requires more data tagging, but selects a GoogLeNet deep network with a depth of 22 layers. Fine-tuning GoogLeNet network on the ImageNet which has already been trained, and applying it to the classification of hand bones. The GoogLeNet network architecture is shown in Figure 8.

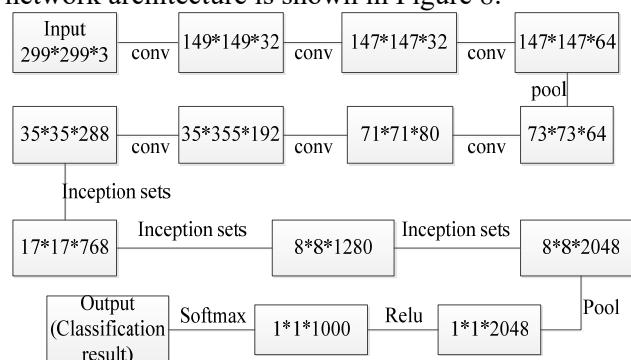


Figure 8. GoogLeNet network architecture

The classification model was trained by deep convolution neural network [14]. The network structure is mainly expanded by increasing the number of network layers (depth) and expanding the number of neurons (width), which has many shortcomings. First of all, as the network increases, the computing capacity will also increase, and the calculation process will be more complex, so the hardware requirements are higher. Secondly, as the network grows, more parameters need to be controlled. If the data set is too small, over-fitting may occur. Finally, the increase of network depth means that the number of network layers increases, causing the gradient instability. At the same time, the gradient will disappear, the model optimization is poor, and the classification accuracy is reduced.

In order to solve the above problems, the Inception module is introduced to reduce the required parameters while increasing the number of network layers and neurons. The inception architecture is shown in Figure 9.

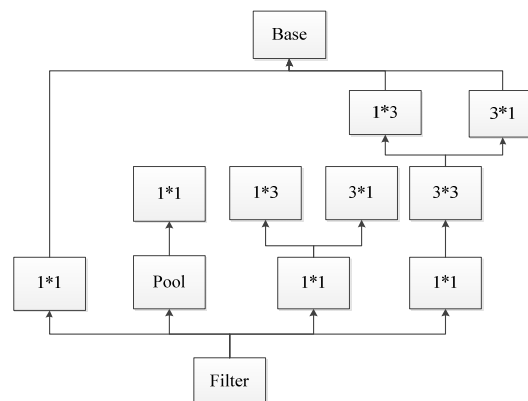


Figure 9. Inception architecture

Inception has remarkable features. The first part is the same as the previous convolution and pooling operation, and three inception module groups are added to the second part of the network. The

two-dimensional convolution of $n \times n$ is replaced by the one-dimensional convolution of $1 \times n$ and $n \times 1$, which solves the over-fitting problem and uses a few parameters to speed up the operation. There are three ways to fine-tune the network model:

(1) Feature extraction using the existing CNN: due to the high similarity between the trained data set and the constructed data set, and the small data set, fine-tuning may result in fitting. Therefore, by removing the last layer, the full connection layer can be used as the feature extraction device.

(2) Using the pre-trained model architecture: because the trained data set is highly similar to the constructed data set and the data set is large, the entire network can be fine-tuned without considering the fitting;

(3) Stratification of training department: the training data set has a low similarity with the constructed data set, and the data set is small and different. Therefore, training should not start from the top of the network, but should be characterized by the values of the first few layers of training, and then the linear classifier should be trained.

Due to the consideration of the size and similarity of the data set, the fine-tuning method in this paper adopted the replacement model of the last layer of fully connected layer to realize the classification of bone age.

6. Experimental results

Children and adolescents have different emphases on the side of the hand bone at different ages. It can be divided into the following three stages: middle, distal phalanges and carpal bone formation from 4 to 9 years old, radius ossification center increases, ulna ossification center appears and grows rapidly; from 9 to 15 years old, most of the articular surface of the ossification center is completely determined, and the wrist bone is basically mature; after the 15 years old, the metacarpal bone is connected with each backbone, and the bones on the wrist are basically closed.

On the basis of the above classification, this paper continues to divide into 6 categories. Transfer learning is used to train the classification samples and determine the classification model. In order to test the model, both single images and multiple images were tested (when single images were tested, the category of pictures was mainly judged). The test accuracy of multiple images reached over 90%, and the experimental results of multiple images is shown in Table 2.

Table 2. Evaluation results

	Man(20 months)	Woman(20 months)
Accuracy	90.5%	94.4%

7. Conclusions

As the health problems of adolescents are getting more and more attention, it is of great practical value to evaluate the bone age of adolescents. In order to reduce the burden on doctors, people are more and more interested in the study of automatic bone age measurement [15-17]. In this paper, deep learning is applied to medicine, and Matlab is used to preprocess X-ray images. Then the images are segmented, and on the basis of transfer learning, GoogLeNet is combined for classification operation. A series of automated bone age assessments make the results more accurate and assist doctors in diagnosis to a certain extent.

References

- [1] Martin, D. D., Deusch, D., Schweizer, R., Binder, G., Thodberg, H. H., & Ranke, M. B. (2009). Clinical application of automated greulich-pyle bone age determination in children with short stature. *Pediatric Radiology*, 39(6), 598.
- [2] Liu, J., & Liu, J. (2017). A review of the development of bone age assessment system. *Life Science Instruments*.
- [3] Zhang, A., Gertych, A., & Liu, A. B. J. (2007). Automatic bone age assessment for young children

- from newborn to 7-year-old using carpal bones. *Computerized Medical Imaging & Graphics*, 31(5), 299-310.
- [4] Milner, G. R., Levick, R. K., & Kay, R. (1986). Assessment of bone age: a comparison of the greulich and pyle, and the tanner and whitehouse methods. *Clinical Radiology*, 37(2), 119-21.
- [5] Frisch, H., Riedl, S., & Waldhor, T. (1996). Computer-aided estimation of skeletal age and comparison with bone age evaluations by the method of greulich-pyle and tanner-whitehouse. *Pediatric Radiology*, 26(3), 226-231.
- [6] Kreitner, K. F., Schweden, F. J., Riepert, T., Nafe, B., & Thelen, M. (1998). Bone age determination based on the study of the medial extremity of the clavicle. *European Radiology*, 8(7), 1116.
- [7] Luo, W., Li, M., Yun, P., & Xiao, H. (2014). [design and appliation of a kind of bone age automatic analysis and assessment device]. *Chinese Journal of Medical Instrumentation*, 38(4), 264-266.
- [8] Liu, J., Qi, J., Liu, Z., Ning, Q., & Luo, X. (2008). Automatic bone age assessment based on intelligent algorithms and comparison with tw3 method. *Computerized Medical Imaging & Graphics*, 32(8), 678.
- [9] Litjens, G., Kooi, T., Bejnordi, B. E., Aaa, S., Ciompi, F., & Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(9), 60-88.
- [10] Hu, T. H., Wan, L., Liu, T. A., Wang, M. W., Chen, T., & Wang, Y. H. (2017). [advantages and application prospects of deep learning in image recognition and bone age assessment]. *Journal of Forensic Medicine*, 629-634.
- [11] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu, P. J., & Lo, B., et al. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical & Health Informatics*, 21(1), 4-21.
- [12] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. , 9351, 234-241.
- [13] Wu, X. (2015). Fully convolutional networks for semantic segmentation. *Computer Science*.
- [14] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems* (Vol.60, pp.1097-1105). Curran Associates Inc.
- [15] Spampinato, C., Palazzo, S., Giordano, D., Aldinucci, M., & Leonardi, R. (2016). Deep learning for automated skeletal bone age assessment in x-ray images ☆. *Medical Image Analysis*, 36, 41-51.
- [16] Lee, H., Tajmir, S., Lee, J., Zissen, M., Yeshiwas, B. A., & Alkasab, T. K., et al. (2017). Fully automated deep learning system for bone age assessment. *Journal of Digital Imaging*, 30(4), 427-441. Simu, S., Lal, S., Fadte, K., & Harlapur, A. (2017). Fully automatic segmentation of phalanges from hand radiographs for bone age assessment. (1), 1-26.