Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Advances in Deep Learning-Based Bone Age Assessment Research

Zhenyu Du *

The First High School of Changsha, Changsha, China

* Corresponding Author Email: zebraprince2020@gmail.com

Abstract. In clinical and forensic medicine, bone age assessment (BAA) is crucial for tracking children's growth and development as well as diagnosing and identifying diseases. While statistical machine learning and mapping and expert discrimination are two examples of traditional BAA techniques, deep learning (DL)-based BAA techniques are progressively becoming more popular as deep learning technology advances. The use of convolutional neural networks in BAA is explained in length in this study, along with information on their development history, particular applications in the segmentation and regression stages of BAA, and related models. Currently, deep learning bone age assessment methods mainly include end-to-end assessment method and region of interest-based assessment method, and in regression, it has also experienced a stage from early exploration to gradual development to rapid development. However, the existing methods have problems such as insufficient attention to image detail information, partial reliance on manual annotation, inaccurate assessment results, and lack of in-depth study of local bone features. In the future, efforts should be made to make better use of image detail information, explore weakly supervised learning methods, conduct in-depth research on local skeletal features, and combine multimodal data to improve the performance of BAA and promote the wide application of this technology in clinical and other fields.

Keywords: Deep learning; bone age assessment; convolutional neural network; segmentation; regression.

1. Introduction

Bone age (BA), as opposed to chronological age (CA), is a measure of a person's biological and skeletal maturity [1]. BAA is crucial in clinical and forensic medicine specialties like pediatric endocrinology, orthopedics, and orthodontics. Given the many syndromes and endocrine issues associated with short height, which often manifest as delayed bone age development, evaluations of children's growth and maturation are particularly important. It can also be used to predict the final height of an adult [2]. The rising prevalence of precocious puberty, which has sparked interest in growth hormone therapy and childhood development, has also made identifying bone age more crucial. Legally speaking, when identifying records for children or teenagers are missing, assessments of physical maturity are utilized to approximate the unknown age of osteogenesis.

2. Methods of BAA

2.1. Development of BAA

Three broad categories can be used to classify BAA methods: assessment techniques based on mapping and expert judgment, like the GP and TW3 methods; statistical machine learning-based techniques, which typically involve using some conventional machine techniques to extract features from the hand and wrist X-ray film, then speeding up the conventional techniques or achieving the automatic assessment of bone age; and deep learning-based BAA techniques, which typically involve integrating deep learning technology into the field of BAA in order to achieve higher accuracy through deep mining data information. Deep learning-based BAA, which usually introduces deep learning technology into the field of BAA, and realizes a more accurate BAA method by mining data information at a deeper level.

Numerous research on deep learning-based BAA have been carried out in tandem with the quick development of deep learning-related technology. Instead of manually specifying the hand bone characteristics, the deep learning assessment automatically extracts the features using a neural network, in contrast to the traditional BAA. As related data quantities have increased and computer technology, especially the GPU, has improved in speed, deep learning has made major strides in several areas in recent years.

2.2. Convolutional Neural Networks (CNN)

Feed-forward neural networks serve as the foundation for Convolutional Neural Networks (CNN), a type of deep learning network constructed with a convolutional topology. The analysis and processing of the input data are ultimately accomplished by using convolutional operations to extract the complicated feature information from the input data through the creation of multi-layer network connections.

In 1987, Time-Delay Neural Networks (TDNN) were introduced as the first implementation of convolutional architectures, applying the backpropagation framework to speech recognition tasks. This pioneering work demonstrated the efficacy of convolutional neural networks through its operational performance.Two years later, CNN were enhanced through backpropagation integration, with subsequent application to handwritten digit recognition systems. This technological evolution culminated in the creation of the LeNet prototype, establishing foundational architecture for optical character recognition.

Due to the limitations of the computing power available at the time, the LeNet network's relatively simplistic structure—which only consists of two convolutional layers and two fully-connected layers—has significant limitations when it comes to processing complex data. In 1998, LeCun and his collaborators continued to apply convolutional neural networks to the handwriting recognition task with the previous learning strategy and added pooling layers to the previously proposed LeNet network structure, and proposed a new convolutional neural network LeNet-5 [3], which can filter the input features with the new pooling layer and reduce the overall computational effort. LeNet-5 defined the basic structure of modern convolutional neural networks, and it is the first convolutional neural network in the world. The alternating convolutional layer-pooling layer in LeNet-5's architecture can extract and choose the picture feature information for filtering, which plays a crucial role in advancing further research in this field. LeNet-5 also establishes the fundamental structure of contemporary convolutional neural networks. Fig. 1 illustrates the basic structure of a convolutional neural network by showing the three main parts of the LeNet-5 structure: the convolutional layer, pooling layer, and fully connected layer.

From AlexNet in 2012 to VGGNet, GoogLeNet, ResNet, Faster R-CNN, and others, convolutional neural networks have been created since then. Through the addition and removal of network layers, structure processing, and weight allocation, these network models have been continuously improved and developed, and their application has gradually moved from computers to computers. Applications of neural networks have gradually spread beyond computer vision to include natural language processing and many other fields.

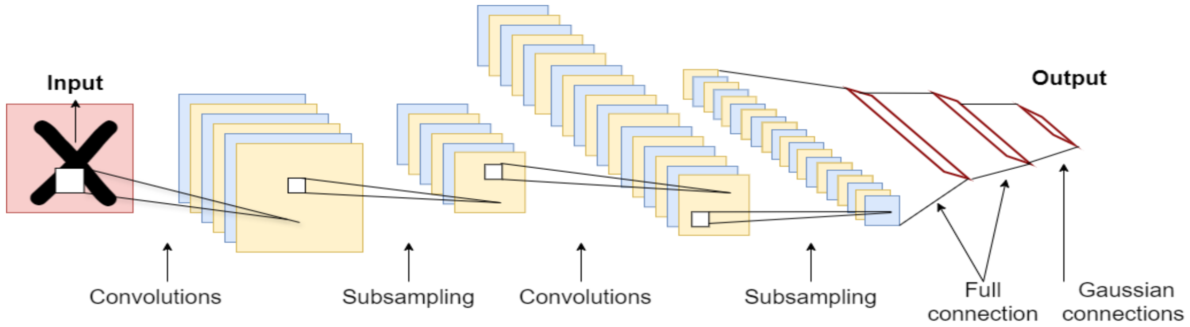

Fig. 1 LeNet-5 structure diagram.

Each convolutional layer has multiple convolutional kernels, each of which typically carries a bias parameter and corresponds to a weight. The convolutional layer's primary function is to extract the input features of the convolutional neural network. Following feature extraction via the convolutional layers, the pooling layer applies additional feature selection or information filtering to the resulting feature map. After several convolutional and pooling layers, the convolutional neural network is connected to one or more fully connected layers. Convolution and pooling are followed by the fully connected layer, which integrates the important local information and converts the feature matrix of the pooled layer into a one-dimensional feature vector, often represented by many neurons. The output layer, which typically follows the final fully connected layer, is the convolutional neural network's final output.

Fig. 2 BAA Procedure.

3. BAA steps

The number of studies looking into skeletal age determination has significantly increased in recent years. Segmentation and regression are the two primary processes used in the majority of these research. While segmentation is the process of eliminating regions of interest from an image, regression is the method used to forecast the age of bones from the segmented image. Fig. 2 shows the flow of the research.

3.1. Segmentation

Convolutional neural networks have the advantage of being able to extract and classify features, which can simplify the entire process of assessing bone age and make BAA simpler and more accurate. The task of assessing bone age by deep learning can be viewed as a classification problem, where X-ray images hand are categorized into appropriate bone age classes. Current deep learning techniques for determining bone age can be broadly divided into two categories: region-of-interest-based techniques and end-to-end techniques.

3.1.1. End-to-end assessment methods

The end-to-end assessment method uses full hand bone X-ray images with only image-level annotation information to feed a convolutional neural network, enabling autonomous BAA. Spampinato et al. [4] proposed a lightweight convolutional neural network model with five convolutional layers, a pooling layer, and a fully-connected layer to assess bone age across different races, age ranges, and genders. A mean absolute inaccuracy of 9.48 months was achieved. In the North American Radiological Society's (RSNA) Machine Learning Challenge for BAA in Children, Alexander et al. [5] won first place using a deep learning model based on the InceptionV3 architecture with an additional layer of dense connections. This enabled the network to comprehend the relationship between pixel and gender information, with a mean absolute error of 4.27-4.5 months on the RSNA dataset. He et al. [6] introduced an end-to-end BAA method based on a deep residual network with lossless photo compression. After the original image has been compressed using a compression module, features are extracted from the compressed image using a residual network. By using a regression model with an improved loss function, the method also forecasts bone age. Evaluation methods that extract bone age result values from complete images are often faster and simpler to utilize, but they are often not interpretable. Since the nuances of the bone age photos are not given enough attention, the experimental results are often less than ideal, less accurate, and have higher mistakes.

3.1.2. Region of Interest (ROI) Assessment Methods

A crucial first step in determining the ROI (region of interest) is segmentation. The various segmentation techniques are described here.

In 2015, researchers at the University of Freiburg introduced the U-Net deep learning architecture [7], a convolutional neural network often used for image segmentation tasks in the medical field, including bone age assessment. Its name comes from its unique form, which consists of an expansion route and a contraction path (Fig. 3). The contraction route, which consists of a convolutional layer and a maximum pooling layer, reduces the spatial resolution of the input picture and adds additional feature maps to extract important abstract elements, such as hand X-rays, which are used to learn skeletal structures in BAA. Following the convolutional layer on an extended route, an upsampling layer reduces the amount of feature maps, improves spatial resolution, and eventually produces a segmentation mask. A key component of U-Net, the jump connection joins the expansion and contraction routes and uses the high-level properties of the contraction path to improve the precision of segmentation. A regression model then determines the bone age of the patient based on the bone formations shown in the segmented image. In medical picture segmentation tasks, such as determining bone age, U-Net has demonstrated remarkable performance.

A variant of the DenseNet architecture, Fully Convolutional DenseNet (FC DenseNet) is a neural network architecture for image segmentation that has demonstrated exceptional performance (as seen in Fig. 4). Each pixel in the input picture and the output segmentation mask is given a unique category name. It is easier to precisely identify object borders when fine-grained features may move around the network thanks to jump connections. Compared to previous segmentation systems, this leads to a better computational efficiency. Interlayer connections significantly improve the effectiveness of parameter use.

In medical imaging, deep learning-based segmentation is used to divide images into distinct sections according to characteristics such as color, texture, form, etc. Deep neural networks are employed in this process to divide organs or tissues. In medical image analysis, it is more typical to use Inception as the foundation of the deep learning model. A convolutional neural network architecture called inception uses a range of filter sizes in the convolutional layer to capture characteristics at different scales, increasing the precision and effectiveness of picture classification. This multi-scale method works well for segmenting medical images. A number of segmentation models, including BAA models, have made use of Inception.

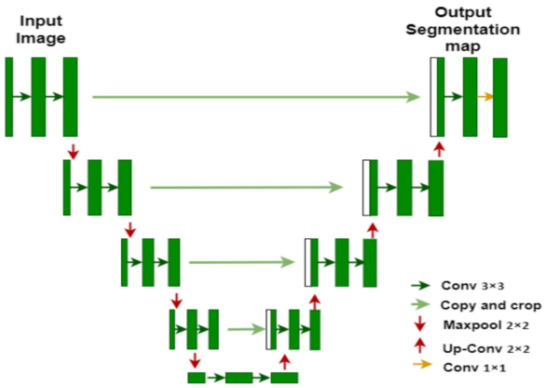

Fig. 3 The typical U-Net architecture.

3.1.3. Application of Deep Learning Models for BAA Segmentation

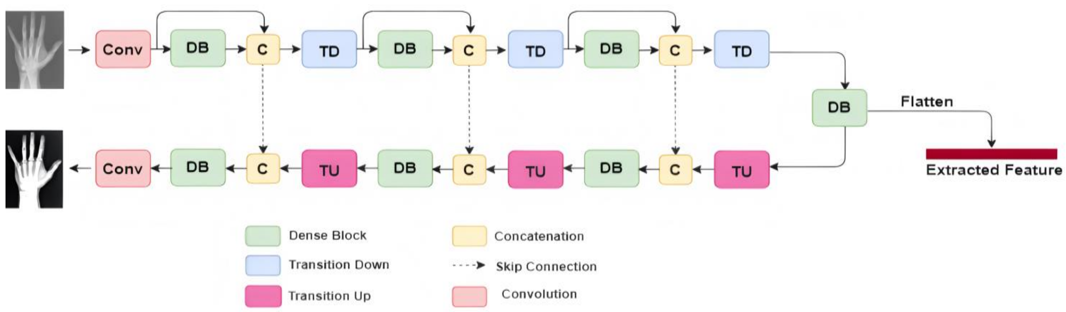

Fig. 4 Refinement of the architecture of FC-DenseNet.

The usage of region-of-interest based assessment methods often includes additional image preprocessing and human a priori knowledge. The solution usually includes the following processing steps, first segmenting the hand from the original X-ray image, then using human a priori knowledge to recognize the skeletal parts, and finally outputting the age assessment results.

Table 1 Application of DL Models for BAA Segmentation.

Year	Researcher	Content of study Specific	Methodology	Experimental results / model characteristics
2016	Liu et al. [8]	Proposing a One-Step Target Detection Algorithm for Multi-Frame Prediction (SSD)	Eliminating the need for RPNs, applying multi-scale features and default boxes	Improve speed and detection accuracy by detecting targets at different scales in different layers, whereas previously the algorithm only detected at the highest layer
2019	Son et al. [9]	Identifying Regions of Interest from Left Hand X-Ray Images with Faster R-CNN and Assessing Bone Age with VGG Network Training	Identify 13 regions of interest using Faster R-CNN and train VGG network with regions of interest extracted from 3344 images.	The mean difference between experimental and expert results was 5.52 months
2020	Hao et al. [10]	Proposing a new multimodal data fusion learning network for BAA	In order to learn nonlinear correlation information of various modal data, a CNN containing a spatial pyramid pooling layer, an attention mechanism module, and textual information is designed.	The former ensures image spatial information integrity and the latter improves tiny feature extraction
2022	Palaniswamy et al. [11]	Introducing a deep learning model based on hyperparameter optimization for automatic BAA and classification	The swarming spider optimization algorithm was utilized to optimize the hyperparameters, a lightweight model based on regional CNN was utilized for feature extraction, and a weighted extreme value learning matrix-based skeletal class classification model and an age assessment model based on the Softmax classifier were used to determine the skeletal age and class labels.	/

In order to increase speed by doing away with the necessity for a region suggestion network and to increase detection accuracy by utilizing multiscale characteristics and default boxes, Liu et al. [8] presented Single ShotMultiBox Detector (SSD), a one-step target detection technique for multibox prediction. In order to determine bone age, Son et al. [9] trained a VGG network using FasterR- CNN; the average difference between expert and experimental results was 5.52 months. 2022 A hyper-parameter optimization-based deep learning model for automatic BAA and classification was presented by Palaniswamy et al. [11]. YOLO versions have been modified in the subsequent years, and they are now YOLOv2, YOLOv3, YOLOv4, and YOLOv5. After every batch of training data is processed by the data loader, the most widely used one at the present, YOLOv5, enhances the training data. The data loader uses three types of data augmentation: picture scaling, color space alteration, and mosaic enhancement to improve the detection of specific small targets. All of YOLO's earlier iterations used predetermined anchor frames during training, and YOLOv5 anchor frames are automatically learned from the training set. By utilizing slicing operations and the cross-stage local network structure, the approach improves the network's feature fusion capabilities during the feature extraction phase, allowing for the extraction of finer-grained features.

To improve target detection, the network filters the target frames using non-maximal suppression techniques and the CIOU-Loss [12] loss function. As a result, YOLOv5 performs exceptionally well in terms of accuracy and detection speed. The model is used to produce the bone age result from the input X-ray picture, although BAA currently just employs the target identification approach. Because the unique difference information of bone age images and the uneven distribution of bone age image samples are ignored, the BAA findings generated are insufficiently accurate. As a result, more study and instruction are needed for deep learning-based BAA.

3.2. Regression

3.2.1. Development of modeling based on deep learning

Although neural network regression analysis was employed in 1995 following manual selection of useful feature regions in wrist bones, the generalization capability of these features remained limited. A classifier based on the attention mechanism was created by Rucci et al. [13], however because of the short sample size, it did not perform well.

In 2008, the particle swarm optimization technique was utilized to identify and segment images from the carpal and RUS dual bone regions. Subsequently, proprietary classifiers were developed to optimize bias performance. In 2016, Spampinato et al. [4] developed the BoNet BAA technique, which used a convolutional neural network to extract information from hand X-ray images. With an average age inaccuracy of 0.8 years, the validity and robustness are good for samples of different racial, age, and gender groups.

Results from pertinent studies have been coming out since 2018. With an inaccuracy of 0.98 years, Zhou Wenxiang [14] suggested the ST-ResNet BAA model, which is based on the spatial transformer and ResNet network. With 90% ulna recognition accuracy, Wang Yongcan et al. [15] presented a convolutional neural network-based bone age detection technique to identify the ulna and radius end regions. Son et al. [9] provided a complete end-to-end BAA system to use the TW3 technique of BAA in 2019 and created an annotated dataset with a mean absolute error of 0.46 years and a root mean square error of 0.62 years.

3.2.2. Application of Deep Learning Models in BAA Regression

Table 2. Development and Current Status of DL in Regression for BAA.

Author	Year	Data's Source	Database	Sample Size	Age Range (years)	Utilized Algorithm	Performance (years)
Kashif et al. [16]	2016	The USA	DHA of the USC	1,101	0-18	SVM	MAE: 0.61
Larson et al. [5]	2018	The USA	Local dataset	14,036	8-20	CNNs	RMS: 0.63 MAD: 0.5
Beheshtian et al. [17]	2023	The USA	RSNA and DHA of the USC	15,238	0-18	DL model	MAD: RSNA: 0.57 ; DHA: 0.58
Liu et al. [18]	2024	China	RSNA, RHPE and a local dataset	21,951	0-18	DL model	MAD: RSNA: 0.34 ; RHPE: 0.52 ; local dataset: 0.47

Many researchers have started working on integrating advanced machine learning technology into the BAA system as a result of the widespread use of computers and the quick development of machine learning knowledge systems. Their goals are to reduce subjective errors, expedite the assessment of bone age by utilizing computers' powerful capabilities, and achieve automated bone age assessment (As shown in Table 2).

4. Shortcomings and Prospects of Current Research

4.1. Problems

The lack of thorough research on local bone features makes the assessment results insufficiently accurate; the straightforward application of target detection algorithms for BAA overlooks the problem of fine-grained variations in bone age images and unequal sample distribution; and current deep learning-based BAA techniques pay insufficient attention to image detail information and some of them rely on subjective and costly manual annotations.

4.2. Future Prospects

Future studies can concentrate on how to better use the detailed information in images to increase BAA's accuracy; investigate techniques like weakly supervised learning to lessen the need for manual annotations and cut costs; investigate local bone features in-depth to maximize feature extraction and model building; and integrate multimodal data to further enhance BAA's performance to better serve radiologists' and clinics' needs.

5. Conclusion

In BAA, deep learning has advanced significantly, and numerous models and algorithms have been developed and used in real-world scenarios. There are still a lot of obstacles to overcome, though, and in order to encourage the broader use and advancement of deep learning in BAA, we must carry out comprehensive research on raising the model's generalization power, lowering costs, and increasing accuracy.

References

[1] Mughal A M, Hassan N, Ahmed A. Bone age assessment methods: a critical review. Pakistan journal of medical sciences, 2014, 30(1): 211-215.

[2] Creo A L, Schwenk W F. Bone age: a handy tool for pediatric providers. Pediatrics, 2017, 140(6).

[3] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[4] Spampinato C, Palazzo S, Giordano D, et al. Deep learning for automated skeletal bone age assessment in X-ray images. Medical image analysis, 2017, 36: 41-51.

[5] Larson D B, Chen M C, Lungren M P, et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. Radiology, 2018, 287(1): 313-322.

[6] He J, Jiang D. Fully automatic model based on se-resnet for bone age assessment. IEEE access, 2021, 9: 62460-62466.

[7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer international publishing, 2015: 234-241.

[8] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[9] Son S J, Song Y, Kim N, et al. TW3-based fully automated bone age assessment system using deep neural networks. IEEE Access, 2019, 7: 33346-33358.

[10] Hao P, Ye T, Xie X, et al. Radiographs and texts fusion learning based deep networks for skeletal bone age assessment. Multimedia Tools and Applications, 2021, 80: 16347-16366.

[11] Palaniswamy T. Hyperparameter optimization based deep convolution neural network model for automated bone age assessment and classification. Displays, 2022, 73: 102206-102216.

[12] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[13] Rucci M, Coppini G, Nicoletti I, et al. Automatic analysis of hand radiographs for the assessment of skeletal age: a subsymbolic approach. Computers and Biomedical Research, 1995, 28(3): 239-256.

[14] Zhou W. Automatic Evaluation of Bone Age Based on X-ray Images. SiChuan: University of Electronic Science and Technology of China, 2018, 1-68.

[15] Wang Y, Hu Y, Shen Y, et al. Bone Age Recognition Using Convolution Neural Network. Journal of Image and Signal Processing, 2018, (1): 1-15.

[16] Kashif M, Deserno T M, Haak D, et al. Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment. Computers in biology and medicine, 2016, 68: 67-75.

[17] Beheshtian E, Putman K, Santomartino S M, et al. Generalizability and bias in a deep learning pediatric bone age prediction model using hand radiographs. Radiology, 2022, 306(2): e220505.

[18] Liu Q, Wang H, Wangjiu C, et al. An artificial intelligence-based bone age assessment model for Han and Tibetan children. Frontiers in Physiology, 2024, 15: 1329145.