Smart Media Journal / Vol.10, No.2 / ISSN:2287-1322
https://dx.doi.org/10.30693/SMJ.2021.10.2.22

Region of Interest Localization for Bone Age Estimation Using Whole-Body Bone Scintigraphy

Thanh-Cong Do[1], Hyung Jeong Yang[2], Soo Hyung Kim[2], Guee Sang Lee[2], Sae Ryung Kang[3], Jung Joon Min[4]

Abstract

In the past decade, deep learning has been applied to various medical image analysis tasks. Skeletal bone age estimation is clinically important as it can help prevent age-related illness and pave the way for new anti-aging therapies. Recent research has applied deep learning techniques to the task of bone age assessment and achieved positive results. In this paper, we propose a bone age prediction method using a deep convolutional neural network. Specifically, we first train a classification model that automatically localizes the most discriminative region of an image and crops it from the original image. The regions of interest are then used as input for a regression model to estimate the age of the patient. The experiments are conducted on a whole-body scintigraphy dataset that was collected by Chonnam National University Hwasun Hospital. The experimental results illustrate the potential of our proposed method, which has a mean absolute error of 3.35 years. Our proposed framework can be used as a robust supporting tool for clinicians to prevent age-related diseases.

Keywords : Bone Age Estimation | Deep Learning | Convolutional Neural Network

I. INTRODUCTION

The global population is aging, leading to an increase in the incidence of age-related diseases. A patient's age can be used to evaluate his health and aging status. Much research on aging focuses on people aged 60 or above, who usually already have age-related diseases [1-3]. Chronological age (CA), what we usually call "age," is of limited applicability to estimating one's physiological status, while biological age (BA), which is estimated using certain biomarkers, can more clearly reflect the person's physical condition and aging state [4]. Some studies have shown that BA can be used as an evaluation index of mortality, morbidity, incidence of disease and disability [5,6]. Therefore, age estimation is a key task in the fields of health informatics, forensic science and anthropology.

Some remarkable studies about bone age estimation (BAE) are Greulich-Pyle (GP) [7] and Tanner Whitehouse (TW) [8]. In the former method, bone age is estimated by comparing the whole hand radiograph with a reference atlas of representative ages that is easily applied in clinical practice, while the TW examines 20 specific regions of interest (RoIs) and assigns scores based on the local structural analysis. However, manual methods can be time-consuming and burdensome to radiologists. These disadvantages have led to the establishment of several automatic computer-assisted techniques.

Over the past decade, machine learning (ML) has been at the heart of many advancements in the field of medical image analysis. The BoneXpert model was developed based on conventional ML techniques that have been shown to have good performance for patients in various clinical settings [9]. Recently, deep learning (DL) has produced many promising

1 Student Member, Graduate Student
2 Member, Professor, Dept. of AI Convergence, Chonnam National University
3 Professor, Dept. of Nuclear Medicine, Chonnam National University Hwasun Hospital
4 Professor, Institute for Molecular Imaging and Theranostics, Chonnam National University Medical School
* This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (NRF-2020R1A2B5B01002085) and This study was financially supported by Chonnam National University (Grant number: 2018-3359)

Manuscript : 2021. 04. 02
Revised : 2021. 04. 19
Confirmation of Publication: 2021. 04. 22
Corresponding Author: Hyung-Jeong Yang,
e-mail: hjyang@jnu.ac.kr

results that go far beyond the results of previous methods in many areas. The use of DL in bone age assessment using X-ray images of hands has received a lot of attention. The hand is the body part of choice for X-ray imaging due to its high bone count and relatively low radiation requirement [10]. Other studies have demonstrated the relationship between the aging process and bone degradation as well as age-related bone uptake of Tc-99m-HDP measured by whole-body bone scintigraphy [11,12]. To quantify the BA of adult bone before the first signs of bone degeneration, a (BAE) method with whole-body bone scintigraphy was suggested [13]. Bone scintigraphy is a highly sensitive nuclear medicine diagnostic imaging technique that is most often performed in all radioactive processes [14]. This technique uses radiation to assess the distribution of active bone formation in the skeleton and detect early significant metabolic changes before they become obvious in conventional radiographs.

In this paper, we focus on DL approaches for the task of predicting the ages of whole-body bone images. This task presents two main challenges: (1) Raw input images are large (about 900 x 3000 pixels). This size is too big for most convolutional neural network (CNN) models to learn efficiently. A common solution to this problem involves downsizing the original images, but doing so may cause loss of some important information. (2) Not all parts of the body are important for age estimation. Use of whole-body images can make it difficult for the model to focus on the important regions. A recent DL-based method shows improved BAE performance by localizing the RoIs and using these regions for the regression task instead of the raw images [15]. Even though the method significantly improves BAE performance, it suffers from one major limitation: manual specification of RoIs can be expensive, time-consuming and almost impossible without domain knowledge from expert radiologists. Therefore, in this paper, we propose a DL-based approach that automatically localizes the most discriminative region for BAE without requiring any extra annotation. Specifically, we first train a classification model to learn the attention maps of the most informative regions. Guided by those attention maps, we then crop the RoIs from the original images and use them as input for the regression model.

The rest of this paper is organized into four sections. Section II outlines the related work. Our proposed method for localizing the most informative regions and predicting the age are shown in section III. The experimental results are provided in Section IV. Finally, conclusions and future work are presented in Section V.

II. RELATED WORK

In recent years, numerous DL-based image analysis methods have been developed for BAE, which is a fundamental process that is used to evaluate the states of many diseases. Recently, the CNN-based LeNet-5 network was presented, which uses 32 x 32 input images instead of 512 x 512 images to predict the age of bone [16]. Support vector regression (SVR) can be used to aggregate heterogeneous features for the estimation of bone age [17]. An effective CNN and SVR-based model was also developed that gives better BAE performance when the data are in a heterogeneous form. Another popular model is BoNet, which is an ad-hoc CNN for BAE that exploits the deformation layer to address nonrigid deformation of bone [18]. In addition, a CaffeNet-based CNN model is presented in [19], which has low complexity compared to other DL models. It has numerous edges that are connected to its neurons, and fixed neuronal values are used. In a BAE task, CaffeNet-CNN models perform better when the size of the training data is reduced. Furthermore, some studies focus on applying transfer learning for bone age classification. A Google Net network with a depth of 22 layers was used for a classification task in [16]. The model was pre-trained on the ImageNet dataset, and an inception block was used to train the classification model.

BAE is a fine-grained recognition task because ossification patterns are usually contained in specific small regions. Some previous work focuses on localizing or finding the bounding box of the most informative regions for BAE [15,19]. However, the heavy requirement for manual input from domain experts makes it impossible to apply to a large-scale dataset. Some recent studies looked at attention-guided localization, which allows a CNN to focus on some specific regions of the input images. Hyunkwang Lee et al [20]. proposed an automated model that segments the region of interest, standardizes the image, and processes the input radiographs for BAE. In [21], Li et al. propose an

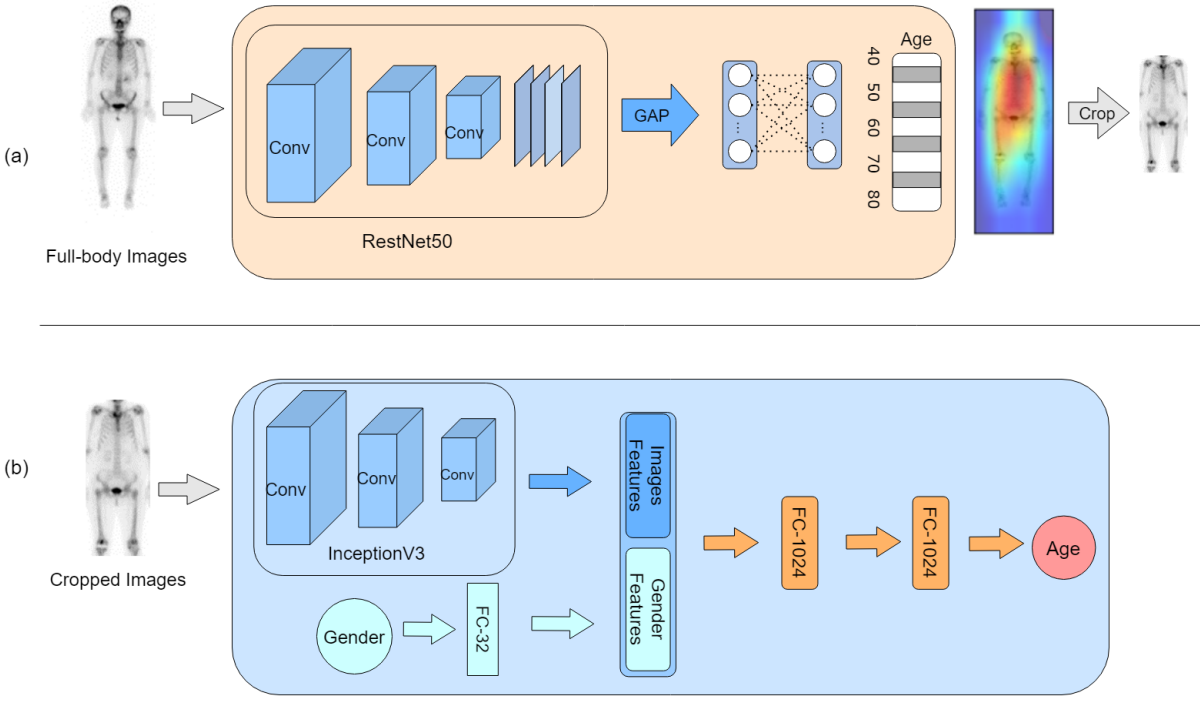

Fig. 1. Our proposed method includes 2 networks. (a) Localization network. (b) Regression network

attention-based multiple instances learning model for weakly-supervised RoI detection. Fu et al [22]. proposed RA-CNN, a model that learns discriminative region attention and region-based feature representation at multi-scales image recognition. Another method is class activation mapping (CAM) [23] that utilize a global average pooling layer with CNN in order to understand which parts of an input image were important for a classification decision. Furthermore, Grad-CAM [24], a generalization of CAM which is applicable to a significantly broader range of CNN model families.

III. PROPOSED METHOD

In this section, we present our proposed method for prediction of bone age. As shown in Figure 1, our proposal includes two phases: RoI localization and age estimation. In the localization step, we first train a classification model to generate attention heat maps that are used to find the most discriminative regions. Guided by these heat maps, we then crop the RoIs from the original images. These cropped patches, along with gender information, are used as input for the regression network to predict patient age.

RoIs localization techniques are widely used in many images analysis research [19-24]. Inspired by these techniques, we propose a method that can automatically identify highly discriminative local patches for BAE. Our patients are divided into four groups by age as in Figure 1(a). For the image features extraction task, we apply ResNet50, a CNN model that was designed to ease the training of deep networks, which can easily enjoy accuracy gains from greatly increasing depth [25]. Let $F \in R^{UxVxK}$ denote K feature maps (with width U and height V) of the last convolutional layer and Y_c is the output for class c. The feature maps are then fed into a global average pooling (GAP) layer, followed by a fully connected (FC) layer. First, we compute the gradient of Y_c with respect to the feature maps F_k (k = 0, 1, ..., $K-1$) as: $\frac{\partial Y_c}{\partial F_k}$. In next step, we apply GAP to the gradients over the width dimension (indexed by i) and the height dimension (indexed by j) as in Eq. (1):

$$w_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial Y_c}{\partial F_{ij}^k} \quad (1)$$

The w_k^c value in for class c and feature map k is used as the weight applied to the corresponding feature

map F_k, and calculate the final heatmap using the weighted sum of feature maps as in Eq. (2):

$$H^c_{(i,j)} = \sum_{k=0}^{K-1} w^c_k F^k_{(i,j)} \quad (2)$$

The final heatmap now has size U x V, similar to the size of the final convolutional feature maps. After obtaining the heat map H, we resize it to the original input images and design a binary mask M to define the most informative regions of the input image:

$$M^c_{(i,j)} = \begin{cases} 1 & H^c_{(i,j)} \geq \gamma \\ 0 & H^c_{(i,j)} < \gamma \end{cases} \quad (3)$$

where γ is the threshold that specifies the size of the cropped regions. The higher value of γ lead to the smaller size of RoIs, and vice versa. Based on Eq. (3), we can crop the most discriminative regions for BAE. In this work, we train the classification model with the original images which have been resized to 300x1000, and the value of γ is set to 150 empirically.

In the second phase, we perform the bone age regression task with the cropped RoIs. As shown in Figure 1(b), we utilize the InceptionV3 model for feature extraction. Gender information is also applied in this phase, which is crucial for the accuracy of the age prediction, as male and female bone structures are naturally different [26]. We make the gender network that takes a binary input (0 for female or 1 for male) and fed it to a FC layer. The output was concatenated with image features before fed to two additional FC layers and a final layer of a single neuron with linear activation to predict the age. The loss function for our regression model is the mean square error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y_i} - y_i)^2 \quad (4)$$

In Eq. (4), N is the batch size. $\hat{y_i}$ and y_i are the predicted age and actual age. Mean absolute error (MAE) is the metric used to evaluate our model as in Eq. (5):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y_i} - y_i| \quad (5)$$

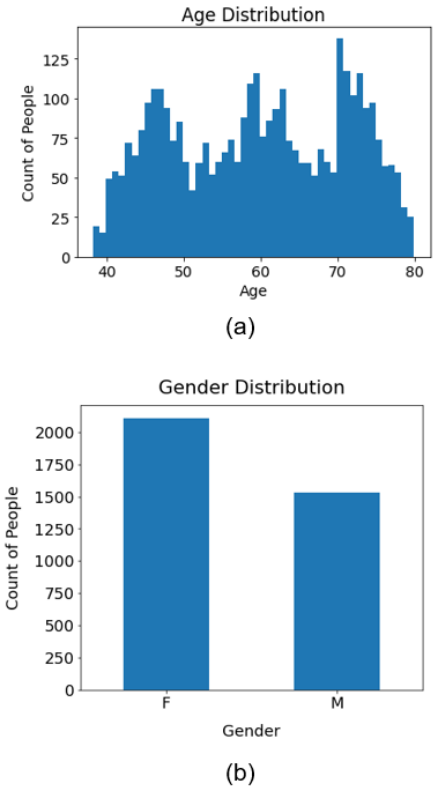

Fig. 2. (a) age distribution, (b) gender distribution

IV. EXPERIMENT AND RESULTS

The dataset used in this paper includes whole-body bone scintigraphy images from 3636 subjects, along with their age (which ranges from 40 to 80) and gender information. The distribution of age and gender is shown in Figure 2. It is shown that the patient distribution in each class is almost even. For the classification network, we apply ResNet50 for feature extraction, then add a GAP layer followed by a final FC layer with 4 output nodes. Some samples of the heat maps and RoIs generated by our localization model are shown in Figure 3. We found that the most discriminative region is usually around the center of the body. In the regression phase, the cropped RoIs are taken as input for an InceptionV3 followed by a GAP layer. The binary gender input is fed through a 32-neuron FC layer before being concatenated with the image features. The concatenated layer then is fed through two 1024-neuron FC layers with relu activation and a dropout of 0.2 after each before serving as input for the last single-neuron layer for age prediction. Both of the

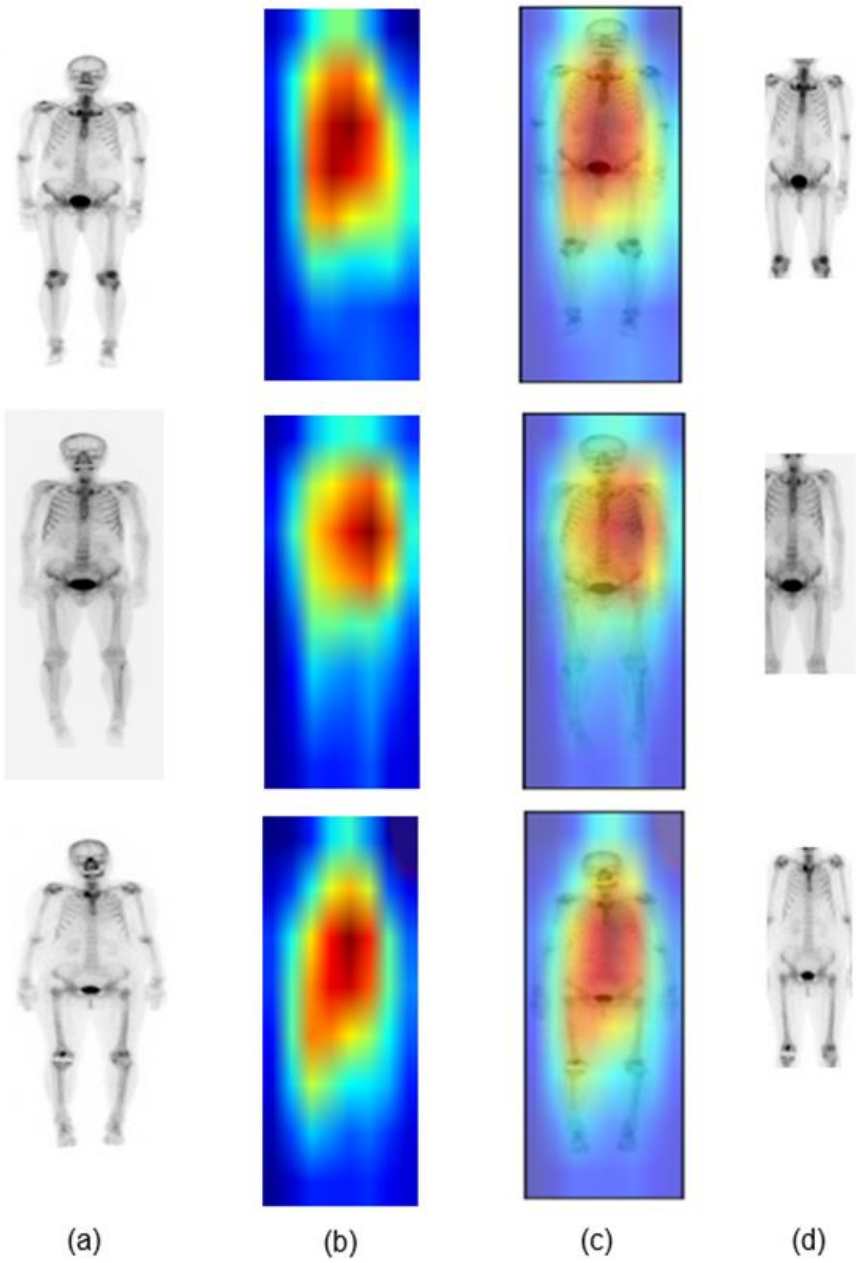

Fig. 3. Some samples of RoIs localization. (a) Original images, (b) Heat maps generated by the localization network, (c) heat maps overlay on original images, (d) cropped patches

CNN models are pre-trained on the ImageNet dataset. The implementation is conducted in Tensorflow with Adam optimizer, a batch size of 32 and a learning rate of 0.0001. Figure 4 illustrates the prediction result of our proposed framework on the test dataset. For comparison purposes, we trained some widely-used CNN architecture for BAE task, including VGG19, Xception, ResNet50, and the model of multiple inputs VGG16 that was proposed in recent research [13]. All input images were resized to 250x750 for these models. Gender information is known as an important factor in BAE because of the physical differences between men and women. Table 1 demonstrates the performance of our model in comparison with other networks, which are also pre-trained with the ImageNet dataset. Clearly, our

proposed method outperforms the others, with an MAE of 3.35 years. To evaluate the impact of gender information on the BAE task, we implement all the models under two conditions: with and without gender input. Inclusion of gender information improves the performance of every model, confirming that it is an important factor for BAE.

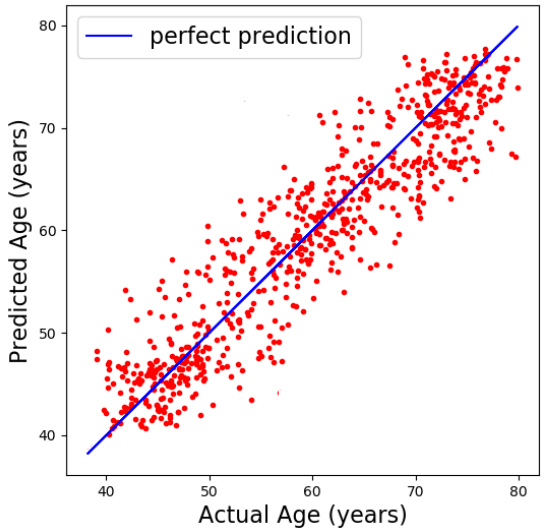

Fig. 4. Prediction result on test dataset

Table 1. Performance of different models

Methods	MAE (years)	
	With gender	Without gender
VGG19	8.02	8.31
ResNet50	7.42	7.81
Xception	4.49	5.24
Multiple inputs VGG16 [13]	3.40	3.75
Proposed method	**3.35**	**3.41**

V. CONCLUSION

In this paper, we proposed an approach to improve the performance of BAE by automatically localize the most discriminative patches in the full-body images, using them as input for regression model. Traditional methods usually resize input images to a smaller resolution before training, which can lead to the missing of valuable information. Localizing the RoIs instead of resizing the images not only reduces the computational burdens but also specifies the most informative regions and keep them for training. The experimental result shows the efficiency of our method with the MAE of 3.35 years. Our research also indicates the key role of gender information for the task of bone age prediction. Furthermore, finding the RoIs in whole-body images is a complex problem that need more research and guarantees. For the future work, we plan to localize multiple RoIs to ensure no valuable information is missing.

REFERENCES

[1] D.W. Belsky, A. Caspi, R. Houts, H.J. Cohen, D.L. Corcoran, A. Danese, et al., "Quantification of Biological Aging in Young Adults," *Proceedings of the National Academy of Sciences*, Vol. 112, No. 30, pp. E4104-E4110, The Rockefeller University, USA, Jul. 2015

[2] N. Barzilai, et al., "The place of genetics in ageing research," *Nature Reviews Genetics,* Vol. 13, No. 8, pp. 589-594, Jul. 2012

[3] H. Heyn, N. Li, et al., "Distinct DNA methylomes of newborns and centenarians," *Proc Natl Acad Sci USA*, Vol. 109, No. 26, pp. 10522-10527, University of Southern California, USA, Jun. 2012

[4] Young Gon Kang, Eunkyung Suh, Jae-woo Lee, Dong Wook Kim, Kyung Hee Cho, Chul-Young Bae, "Biological age as a health index for mortality and major age-related disease incidence in Koreans: National Health Insurance Service - Health screening 11-year follow-up study," *Clinical interventions in aging*, Vol. 13, pp. 429-436, Jan. 2018

[5] G.A. Borkan, A.H. Norris, "Assessment of biological age using a profile of physical parameters," *Journal of Gerontology*, Vol. 35, No. 2, pp.177-184, Mar. 1980

[6] M. Uttley, M.H. Crawford, "Efficacy of a composite biological age score to predict

ten-year survival among Kansas and Nebraska Mennonites," *Human Biology,* Vol. 66, No. 1, pp. 121-144, Feb. 1994
[7] W.W. Greulich, S.I. Pyle, *Radiographic atlas of skeletal development of hand wrist, Stanford University Press,* California, 1959
[8] J.M. Tanner, MJR. Healy, H. Goldstein, N. Cameron, *Assessment of skeletal maturity and prediction of adult height (TW3 method) 3rd edition,* Saunders Company, 2001
[9] H.H. Thodberg, S. Kreiborg, A. Juul, K.D. Pedersen, "The BoneXpert method for automated determination of skeletal maturity," *IEEE Transactions on Medical Imaging*, IEEE, Vol. 28, No. 1, pp. 52-66, Jan. 2009
[10] J.H. Lee, Y.J. Kim, K.G. Kim, "Bone age estimation using deep learning and hand X-ray images," *Biomedical Engineering Letters*, Vol. 10, pp. 323-331, Mar. 2020
[11] W. Brenner, N. Sieweke, K.H Bohuslavizki, W.U. Kampen, M. Zuhayra, M. Clausen, E. Henze, "Age and sex-related bone uptake of Tc-99m-HDP measured by whole-body bone scanning," *Nuclear medicine*, Vol. 39, No. 5, pp. 127-132, Feb. 2000
[12] V.D. Kakhki, S.R. Zakavi, "Age-related normal variants of sternal uptake on bone scintigraphy" *Clinical Nuclear Medicine,* Vol. 31, No. 2, pp. 63-67, Feb. 2006
[13] P.D.C. Nguyen, E.T. Baek, H.J. Yang, S.H. Kim, S.R. Kang, J.J. Min, "Multiple Inputs Deep Neural Networks for Bone Age Estimation Using Whole-Body Bone Scintigraphy," *Journal of Korea Multimedia Society*, Vol. 22, No. 12, pp. 1376-1384, 2019
[14] T.V.D. Wyngaert, K. Strobel, W.U. Kampen, T. Kuwert, W.V.D. Bruggen, H.K. Mohan, et al., "The EANM Practice Guidelines for Bone Scintigraphy," *European J ournal of Nuclear Medicine and Molecular Imaging*, Vol. 43, No. 9, pp. 1723-1738, Jun. 2016
[15] M. Escobar, C. Gonzalez, F. Torres, L. Daza, G. Triana, P. Arbelaez, "Hand pose estimation for pediatric bone age assessment," *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 531-539, Shenzhen, China, Oct. 2019
[16] Y. Wang, Q. Zhang, J. Han, Y. Jia, "Application of Deep learning in Bone age assessment," *IOP Conf. Series: Earth Environmental Science*, Vol. 199, No. 3, 2018
[17] M.W. Nadeem, H.G. Goh, A. Ali, M. Hussain, M.A. Khan, V.a. Ponnusamy, "Bone Age Assessment Empowered with Deep Learning: A Survey, Open Research Challenges and Future Directions. Diagnostics," *Diagnostics (Basel)*, Vol. 10, No.781, Oct. 2020
[18] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in x-ray images," *Medical image analysis*, Vol. 36, pp. 41-51, Feb. 2017
[19] V. I. Iglovikov, A. Rakhlin, A.A. Kalinin, A.A. Shvets, "Paediatric bone age assessment using deep convolutional neural networks," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Vol. 11045, pp. 300-308, 2018
[20] H. Lee, S. Tajmir, J. Lee, M. Zissen, B.A. Yeshiwas, T.K. Alkasab, G. Choy, S. Do, "Fully Automated Deep Learning System for Bone Age Assessment," *Journal of Digital Imaging*, Vol. 30, No. 4, pp. 427-441, Aug. 2017
[21] J. Li, W. Li, A. Gertych, B.S. Knudsen, W. Speier, C.W. Arnold, "An attention-based multi-resolution model for prostate whole slide image classification and localization," arXiv:1905.13208, May, 2019
[22] J. Fu, H. Zheng, T. Mei, "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4438-4446, Honolulu, USA, Jul. 2017
[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, "Learning deep features for discriminative localization," *Proceedings of the IEEE conference on computer vision and pattern recognition,* Vol. 1, pp. 2921-2929, Las Vegas, USA, Jun. 2016
[24] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Proceedings of the IEEE international conference on computer vision*, pp. 618-626, Venice, Italy, Oct. 2017
[25] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 770-778, Dec. 2016
[26] M.A. Zulkifley, S.R. Abdani, N.H. Zulkifley, "Automated Bone Age Assessment with Image Registration Using Hand X-ray Images," *Appl. Sci.* Vol. 10, No. 7233, Oct. 2020

Authors

Thanh-Cong Do

He received his B.S. degree in Information Technology from University of Engineering and technology, Vietnam National University (UET-VNU) in 2019. He is currently pursuing the M.S. degree in the School of Artificial Intelligence Convergence at Chonnam National University, South Korea. His research interest includes Reinforcement Learning, Deep Learning, Computer Vision and Bioinformatics.

Hyung-Jeong Yang

She received her B.S., M.S. and Ph.D. from Chonbuk National University, Korea. She is a professor at the Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, Korea. Her main research interests include multimedia data mining, Medical data analysis, Social Network Service data mining and Video data understanding.

Guee- Sang Lee

He received a B.S. degree in Electrical Engineering and a M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received a Ph.D. degree in Computer Science from Pennsylvania State University in 1991. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video technology.

Soo-Hyung Kim

He received a B.S. degree in Computer Engineering from Seoul National University in 1986. He received a M.S. degree and a Ph.D. degree in Computer Science from Korea Advanced Institute of Science and Technology. He is currently a professor at the Department of Artificial Intelligence Convergence, Chonnam National University, Gwangju, Korea. His research interest includes pattern recognition, document-image processing, medical-image processing, and ubiquitous computing.

Sae-Ryung Kang

She received her M.D., M.S. and Ph.D. degree from Chonnam National University, South Korea. She is currently an Assistant Professor of the Department of Nuclear Medicine, Chonnam National University Hwasun Hospital. Her research interests include nuclear medicine and molecular imaging.

Jung-Joon Min

He received her M.D., M.S. and Ph.D. degree from Chonnam National University Medical School. He is currently a professor and Head of Nuclear Medicine at Chonnam National University Medical School and Hwasun Hospital, Director of Institute for Molecular Imaging and Theranostics, Chonnam National University Medical School.