ICIC Express Letters
Volume **14**, Number **2**, February **2020**
 ISSN 1881-803X
pp. **163–170**

ENHANCEMENT FOR AUTOMATIC EXTRACTION OF ROIS FOR BONE AGE ASSESSMENT BASED ON DEEP NEURAL NETWORKS

Chansu Lee and Byoung-Dai Lee*

Department of Computer Science
Kyonggi University
154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Kyonggi-do 16227, Korea
ckstn222@kgu.ac.kr; *Corresponding author: blee@kgu.ac.kr

Received August 2019; accepted November 2019

Abstract. *The Greulich-Pyle (GP) and Tanner-Whitehouse (TW) methods are two major methods for evaluating pediatric growth. Of these methods, the TW3 technique is more commonly used than the GP method in Bone Age Assessment (BAA). As deep learning has been developed recently, research on automating the BAA system using deep learning has been continued with promising results. It is important to extract 13 Regions of Interest (RoIs) in the TW3-based fully automated BAA system. When directly extracting 13 RoIs using the existing object detection technology, there are numerous difficulties in distinguishing RoIs with similar shapes, and as a result, the performance of the system tends to deteriorate. Therefore, in the TW3-based fully automated BAA system, bounding RoIs (bRoIs), including the 13 RoIs, are extracted first. Among the various methods that can be used to extract bRoIs, the bRoI-extracting method that considers angle is proposed in this paper. An X-ray image database with approximately 3855 annotations has been built to evaluate this method. Comparative experiments showed that the proposed method improved the performance of the BAA system by extracting bRoIs more accurately than other existing bRoI-extracting methods.*
Keywords: Deep learning, TW3, Bone age assessment system, Rotated object detection

1. **Introduction.** Bone Age Assessment (BAA) is a key measure in assessing the growth of children. Although there is no standardized method for BAA, the Greulich-Pyle (GP) method [1] and the Tanner-Whitehouse (TW) method [2] are the most commonly used methods in clinical practice. The GP method is based on the atlas proposed by Greulich and Pyle of Stanford University. The atlas consists of a series of standard-grade images of 30 regions of the left hand according to bone maturity grades, and the examiner selects one of the standard-grade images that is similar to the X-ray image to be assessed. The examiner then determines the bone age of the subject based on the corresponding recorded standard age. Given the advantages of this method, such as simplicity and short computation time, it is currently used widely in clinical practice, but determining the exact age of the bone is difficult because the standard-grade images are listed at intervals of six months to one year. In addition, bone age reading errors may occur due to the subjectivity of the examiner. Furthermore, a disadvantage of this method lies in the difficulty of accurately calculating the bone age when there are many bone variations. The TW method is a method of determining bone age by summing up the maturity scores of the bones of each region using an X-ray image of the left wrist. This method has been revised several times and recently, a revised version, the TW3, which calculates only the scores of the Radius, Ulna, and Short bones (RUS), has been used. Although this TW3 method has some disadvantages in that the evaluation process is more complicated and the evaluation time is longer compared to the GP method, it has several advantages in

DOI: 10.24507/icicel.14.02.163

that the bone age table is divided into units of 0.1 years, which enables more precise assessment. Furthermore, the maturity of each bone is scored and evaluated, minimizing the overall discrepancy between each bone and allowing an objective evaluation.

As deep learning has been developed recently, research on automating the BAA system using deep learning has been carried out with promising results. It is crucial to accurately extract 13 Regions of Interest (RoIs) for the TW3-based BAA utilizing deep learning, as shown in Figure 1. At this time, there are various ways to extract 13 RoIs. First, 13 RoIs can be extracted directly from an X-Ray image of the left hand. In this method, it is difficult to distinguish RoIs having similar shapes, so the performance of the BAA system tends to deteriorate. The second method consists of two steps: 1) extraction of the approximate areas, including the 13 RoIs, 2) followed by extraction of the 13 RoIs. This method can improve upon the disadvantages of the first method by reducing the search space for extracting 13 RoIs. The approximate areas are called bounding RoIs (bRoIs). Therefore, in this paper, we propose a method to accurately extract RoIs using the TW3-based BAA system utilizing deep learning, referring to the second method described above. We propose, in particular, a technique for extracting areas while considering rotation at the bRoI-extraction step using object detection technology, rather than the existing rectangular extracting method.

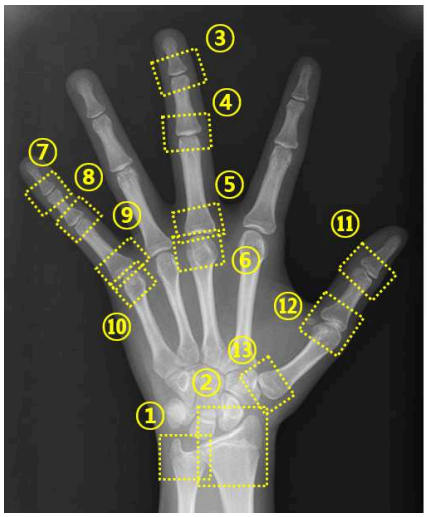


FIGURE 1. RoIs used in the TW3 method

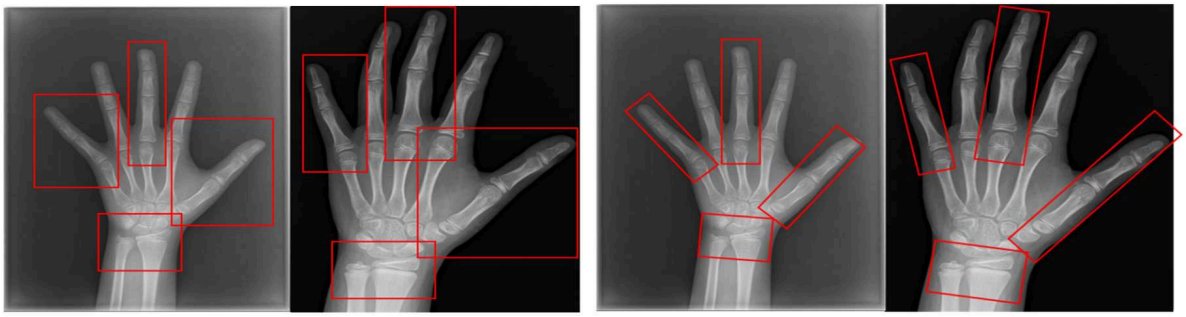
Single Shot multibox Detector (SSD), one of the representative object detection techniques, predicts the category scores and box offsets for fixed default bounding boxes using small convolutional filters that are applied to feature maps [3]. The default bounding boxes composed of different aspect ratios, like the anchor boxes used in the Region Proposal Networks (RPNs) of the Faster Region-based Convolutional Neural Network (Faster R-CNN), are applied to every grid center of each feature map [4]. The key technique of the method proposed in this study is adding the angle parameter to the coordinate information of the default bounding boxes (x, y, w, h), and then letting the system learn the new coordinate information (x, y, w, h, a), including the newly added angle parameter, so that even a rotated object can be accurately detected while considering angle. In this study, three methods were compared through experiments, and the proposed method was proven to be superior to the other methods.

The remainder of this paper is organized as follows. In Section 2, previous studies related to the deep learning-based BAA system and rotated object detection are introduced. Section 3 describes in detail the method proposed in this paper. Section 4 presents the experimental results, and Section 5 summarizes conclusions.

2. Related Works.

2.1. **BAA system using deep learning.** In reference [5], a GP-based fully automated deep learning platform was proposed for the BAA system. The system consists of two main components: a preprocessing engine and a classifier. The preprocessing engine divides the entire 512×512 X-ray image into sample patches (e.g., 32×32 image patches), and the class of each patch is determined by Convolutional Neural Network (CNN). Using the image patch classification results, masks for the hand and the wrist are created and enhanced by image processing. For preprocessed X-ray images, the classifier estimates the bone age using CNN, which are composed of pre-adjusted GoogleNet models [6]. However, it is difficult to accurately determine the bone age by this proposed system, which is a fundamental issue with the GP method.

In reference [7], a TW3-based complete end-to-end BAA system utilizing deep neural networks was proposed. This system takes three steps to extract the 13 RoIs required in the TW3 method, as shown in Figure 2. In the first step, instead of extracting 13 RoIs directly from the X-ray image of the left hand, the concept of bRoIs is applied so that bRoIs, including RoIs, are extracted for each region. In step two, the 13 RoIs are extracted from the corresponding pre-extracted bRoIs. Finally, in step three, the bone age is read for each RoI. In this paper, the necessity of bRoIs was emphasized by comparing the direct extracting method of the 13 RoIs needed in the TW3 method and the two-step method of first extracting the bRoIs, followed by extracting the 13 RoIs. However, it was found that the quality of the image had a significant impact on the bRoI extraction step; there were cases where the bRoIs could not be perfectly extracted when the bRoIs were being extracted from a low-quality X-ray image.



(a) Without considering angle (b) Considering angle

FIGURE 2. Examples of extraction bRoIs

2.2. **Object detection for rotated object.** In reference [8], a method for detecting a rotated object was proposed, in which a Rotation Anchor (R-Anchor) is added to the RPNs of the Faster R-CNN for the detection of a rotated object. The key technique of this method is adding the angle parameter to the existing Bounding Box (BBox) and adding the Rotation RoI Pooling (RRoI Pooling) Layer to pool the rotated bbox. In the RRoI Pooling Layer, each sub-region is rotated at right angles until the corresponding area in the feature map becomes the pool size, then the feature value is stored, and a Max Pooling is performed. The above technique may be applicable to all methods that have a domain proposal step in the object detection technique. Therefore, in this paper, this technique was applied to SSD at the bRoI extraction step prior to extracting the 13 RoIs.

3. Methodology.

3.1. **Extraction of RoIs in BAA system.** There are a variety of methods for extracting the 13 RoIs using the TW3-based BAA system. However, as mentioned above, extracting the 13 RoIs directly from an X-Ray image of the left hand is likely to cause performance degradation of the BAA system [7]. Therefore, it is ideal to perform the extraction of the 13 RoIs in two steps. In this method, approximate areas containing the RoIs are extracted in step one, and then the RoIs included in each approximate area are extracted in step two.

3.2. **Proposed method for extraction of rotated object detection.** Traditional object detection methods use Bounding Box (BBox) to find the target object in the image. BBox is a parameterized square with four variables: center point position (x, y), width (w), and height (h). The red BBoxes in Figure 2(a) are four bRoIs, including the 13 RoIs. In this situation, the BBoxes cannot provide the exact size of the object, and the unnecessary surrounding information is also included in these BBoxes. When extracting the bRoIs to extract the 13 RoIs using the BAA system, the corresponding bRoIs cannot be extracted vertically, which can lead to significant performance degradation.

In this paper, to overcome the above difficulties, rotated default bounding boxes are defined. Rotated default bounding boxes are rectangles with an angle parameter that defines the direction. Using this method, bRoIs are extracted while considering angle using the rotated default bounding boxes as shown in Figure 2(b), so the disadvantage of existing bRoIs including surrounding information can be overcome. Therefore, we propose that applying rotated default boxes in the BAA system is a better choice when extracting the 13 RoIs. The SSD modified according to bRoIs while considering each angle suggests several rotated default bounding boxes, through which positive samples for each rotated default bounding box are selected via Intersection-over-Union (IoU) during training. Repeated predictions in detection are suppressed. The IoU between two rotated default bounding boxes A and B is defined as follows:

$$\text{IoU } (A, B) = \frac{\text{area } (A \cap B)}{\text{area } (A \cup B)} \tag{1}$$

where $\cup$ and $\cap$ are Boolean operations between two rotated default bounding boxes. To calculate the IoU of the two rotated default bounding boxes, the IoU computation algorithm described in reference [8] was used in this study. The IoU calculated by the algorithm of reference [8] is used as a selection criterion for positive samples of rotated default bounding boxes during training to help correctly regress the corresponding rotated default bounding boxes. The technique that adds angle to the BBox, such as rotated default bounding boxes, can be also applied to various object detection methods that draw anticipated bounding boxes at places where an object is likely to exist, like the anchor box in Region Proposal Networks (RPNs).

3.3. **Training.** We expanded the training procedure of SSD to include angular estimation for the rotated bounding RoIs learning process. The rotated default boxes with different aspect ratios and angles suggested in each feature map are calculated as individual boxes using the IoU computation and then classified into positive and negative samples using 0.5 as a reference boundary. Boxes determined to be positive samples are responsible for generating the losses of location and angle regression. The overall objective loss function is as follows:

$$L(x, l, g, p) = \frac{1}{N} \left(L_{cls}(p) + L_{loc}(x, l, g) \right) \tag{2}$$

where N is the number of matched prior rotated default boxes. Classification loss $L_{cls}(p)$ is the focal loss [9] where p is predicted probability value over multiple classes:

$$L_{cls}(p_t) = -(1-p_t)^{\gamma}\log(p_t) \tag{3}$$

The localization loss is similar to SSD, where we calculate the Smooth L1 loss used in Faster R-CNN [4] between the predicted box (l) and the ground truth box (g) parameters:

$$L_{loc}(x,l,g) = \sum_{i\in Pos}\sum_{j}\sum_{m\in\{cx,cy,w,h,a\}} x_{ij}\,\mathrm{smooth}_{L1}\left(\hat{l}_i^m - \hat{g}_j^m\right) \tag{4}$$

where $x_{ij} \in \{1,0\}$ is an indicator for matching the ith prior rotated default boxes to the jth ground truth rotated default boxes. $\hat{l}$ and $\hat{g}$ are defined as follows, which are offsets of the parameters in l and g with their corresponding prior rotated default boxes p, respectively:

$$\hat{t}^{cx} = (t^{cx} - p^{cx})/p^w,\ \ \hat{t}^{cy} = (t^{cy} - p^{cy})/p^h; \tag{5a}$$

$$\hat{t}^{w} = \log(t^w/p^w),\ \ \hat{t}^{h} = \log\left(t^h/p^h\right); \tag{5b}$$

$$\hat{t}^{a} = (t^a - p^a) \tag{5c}$$

Equations (5a), (5b), and (5c) correspond to the location regression term, the size regression term, and the angle regression term, respectively. In the angle regression section, the angle parameter was determined simply by subtraction. This minimizes the angle regression term, approximating the angle ground truth during training.

4. **Experimental Results.** In this section, the data set used in the experiment is introduced and the three methods for extracting the 13 RoIs are compared based on their success rates of RoI extraction.

4.1. **Dataset.** Because there is no publicly available sample data for TW3-based machine learning, we created a dataset using 3,855 X-ray images of the left hand. Table 1 shows how this data set is applied to the three bRoI-extraction methods.

Table 1. Using datasets in three methods

	METHOD 1	METHOD 2	METHOD 3
Training	X	21893	21893
Validation	X	2432	2432
Test	380	380	380

In Method 1, the bRoIs are extracted through image processing used in [7]. Since this method does not use a deep neural network, training and validation sets are not needed; thus, a set of 380 images was used for the test. In Method 2, the bRoIs are extracted using BBoxes that do not include angle, while in Method 3, the bRoIs are extracted while considering angle using BBoxes that include angle. In the last two methods, rotation-mediated data augmentation was performed on 3,475 images out of the existing 3,855 X-ray images of the left hand, except for the images included in the random selected 380 test set. A data set consisting of a total of 24,325 images, including the original images and the images rotated by ±10 degrees, ±20 degrees, and ±30 degrees, was prepared. For use in the learning process, 21,893 and 2,432 images were used for the training and validation sets, respectively.

4.2. **Training details.** The experiments in Methods 2 and 3 were conducted using SSD. The training was carried out using the stochastic gradient descent algorithm with momentum. The mini-batch size was set to 16, and the momentum was set to 0.9. The training was regularized by weight decay, which was set to 0.0001. The learning rate was initially set at 1e-3 and gradually decreased by 10 times at every 50 specified step. The epoch size was set to 30. Feature Pyramid Network (FPN) [10] was set as the Backbone Network. The SSD used in Method 3 was modified to allow even a rotated object to be

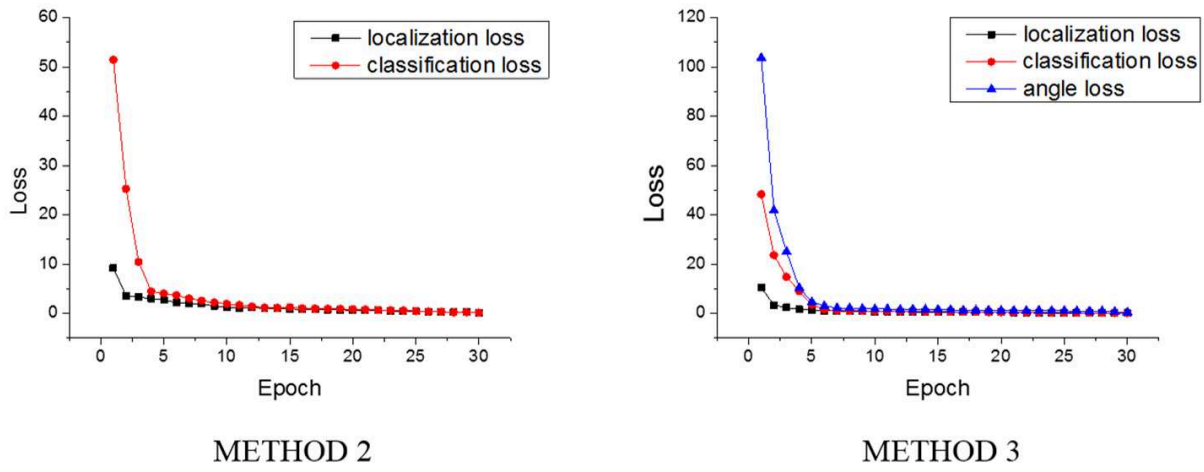


FIGURE 3. Loss curves of the training process

detected by adding the angle parameter to the default boxes. Figure 3 shows the losses according to epoch in Methods 2 and 3.

4.3. **Evaluation.** Table 2 shows the success rates of the RoI extraction performed via the three methods. The success rate of Method 3 proposed in this paper was about 97.63%, which was higher than that of the other methods to which it was compared.

TABLE 2. Success rate of RoIs extraction

bRoIs Extraction METHOD	Success Rate	Total	Extraction Success
METHOD 1	91.84%	380	349
METHOD 2	93.42%	380	355
METHOD 3	97.63%	380	371

Figure 4 shows examples of the failure to extract RoIs using the three methods. CASE 1 shows an example where the RoIs could not be extracted because the bRoIs could not be extracted using Method 1. CASE 2 shows an example where the bRoIs of a middle finger also included the RoIs of an index finger because the bRoIs determined to be a middle finger also included the information of the index finger.

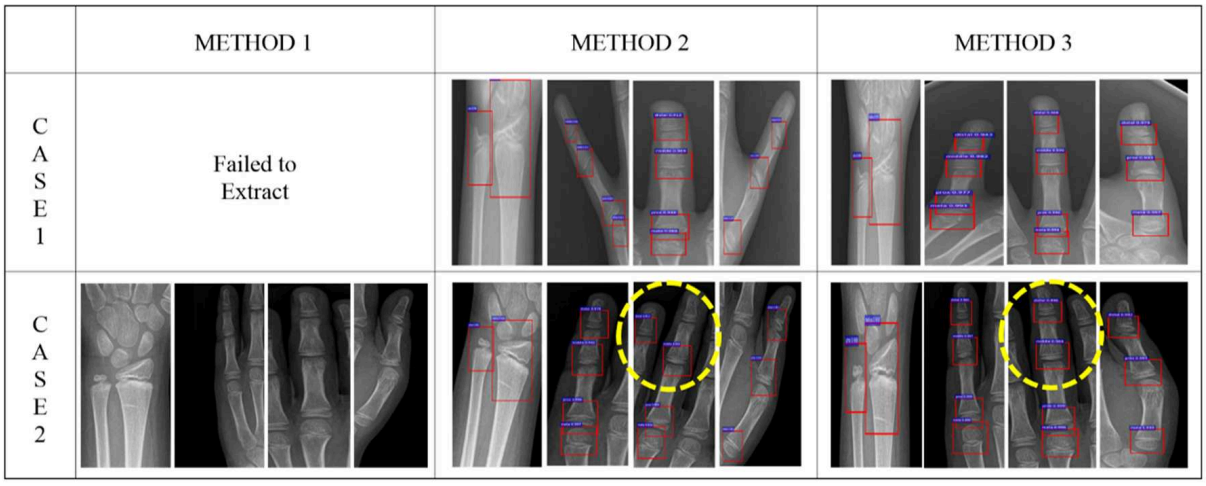


FIGURE 4. Examples of the failure of extracting the 13 RoIs in three methods

According to this experiment, in Method 1, sometimes the bRoIs could not be extracted depending on the quality of the image. As shown in Figure 5, when the RoIs were extracted from the contrast-enhanced images of the original image, the extraction performed by Method 1 was sensitive to the contrast value, and the extraction results were affected accordingly. In Methods 2 and 3, trained bRoIs were extracted regardless of the quality of the image. However, in Method 2, unnecessary surrounding information was also included; in addition to the intended RoIs to be extracted, the images possibly included other areas having shapes similar to the RoIs of other fingers. In addition, due to the lack of angle concept, in some cases, both the middle finger and the index finger could be mistaken for the middle finger in Method 2, making this method difficult to use in the BAA system. These results suggest that Method 3 is superior to the other bRoI-extraction methods.

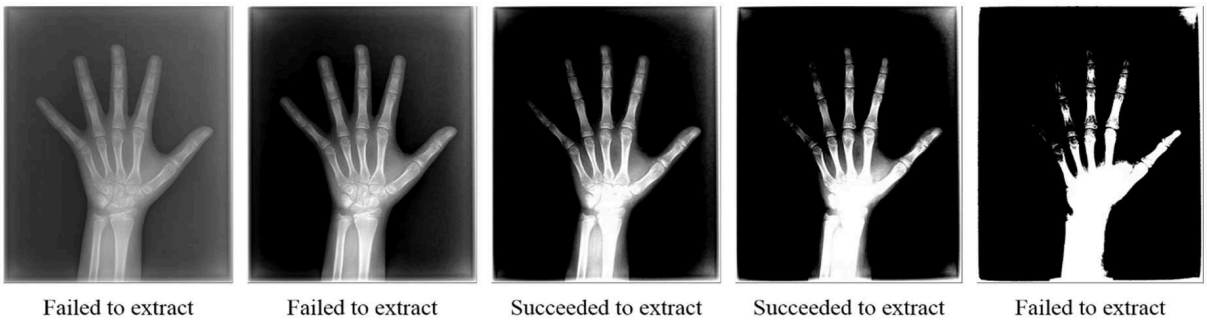


FIGURE 5. Extraction of bRoIs according to the change of contrast in Method 1

5. **Conclusions.** In this paper, we introduced a method for extracting rotated bRoIs, which is one of the methods used to extract bRoIs prior to extracting RoIs in the TW3-based BAA system. The key point of the proposed method is that bRoIs can be extracted regardless of the image quality, excluding the unnecessary information. The experimental results of the present study demonstrated that the proposed method employing bRoIs while considering angle is superior to the other methods to which it was compared. According to the results, the success rates of extracting the 13 RoIs in Method 1, which is based on image processing, was 91.84%. In Method 2, where the bRoIs are extracted using the BBox without considering angle, the success rate was 93.42%. Finally in Method 3 proposed in this paper, where the rotated bRoIs are extracted using the BBox while considering angle was 97.63%.

Acknowledgment. This work is supported by Kyonggi University Research Grant 2017.

REFERENCES

[1] W. W. Greulich and S. I. Pyle, Radiographic atlas of skeletal development of the hand and wrist, *Amer. J. Med. Sci.*, vol.238, no.3, p.393, 1959.

[2] H. Goldstein, J. M. Tanner, M. Healy and N. Cameron, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3 Method)*, Harcour, New York, NY, USA, 2001.

[3] W. Liu, D. Anguelov, D. Erhan, S. Christian, S. Reed, C.-Y. Fu and A. C. Berg, SSD: Single shot multibox detector, *Proc. of European Conference on Computer Vision (ECCV)*, vol.9905, pp.21-37, 2016.

[4] S. Ren, K. He, R. Girshick and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *Proc. of Neural Information Processing Systems (NIPS)*, pp.91-99, 2015.

[5] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy and S. Do, Fully automated deep learning system for bone age assessment, *Journal of Digital Image*, vol.30, no.4, pp.427-441, 2017.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anfuelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, *Proc. of IEEE Conference on Computer Vison & Pattern Recognition (CVPR)*, pp.1-9, 2015.

[7] S. J. Son, Y. M. Song, N. K. Kim, Y. H. Do, N. J. Kwak, M. S. Lee and B. D. Lee, TW3-Based fully automated bone age assessment system using deep neural networks, *IEEE Access*, vol.7, pp.33346-33358, 2019.

[8] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng and X. Xue, Arbitrary-Oriented scene text detection via rotation proposals, *IEEE Trans. Multimedia*, vol.20, no.11, pp.3111-3122, 2018.

[9] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollar, Focal loss for dense object detection, *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp.2980-2988, 2017.

[10] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, Feature pyramid networks for object detection, *Proc. of IEEE Conference on Computer Vison & Pattern Recognition (CVPR)*, pp.2117-2125, 2017.