Article

Scaphoid Fracture Detection by Using Convolutional Neural Network

Tai-Hua Yang [1,2], Ming-Huwi Horng [3], Rong-Shiang Li [4] and Yung-Nien Sun [4,*]

1. Department of Biomedical Engineering, National Cheng Kung University, Tainan 701, Taiwan; dd2006tw@gmail.com
2. Department of Orthopedic Surgery, College of Medicine, National Cheng Kung University Hospital, National Cheng Kung University, Tainan 704, Taiwan
3. Department of Computer Science and Information Engineering, National Pingtung University, Pingtung 912, Taiwan; horng@mail.nptu.edu.tw
4. Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 701, Taiwan; geniuscat1227@gmail.com
* Correspondence: ynsun@mail.ncku.edu.tw; Tel.: +886-6-2757575 (ext. 62526)

Abstract: Scaphoid fractures frequently appear in injury radiograph, but approximately 20% are occult. While there are few studies in the fracture detection of X-ray scaphoid images, their effectiveness is insignificant in detecting the scaphoid fractures. Traditional image processing technology had been applied to segment interesting areas of X-ray images, but it always suffered from the requirements of manual intervention and a large amount of computational time. To date, the models of convolutional neural networks have been widely applied to medical image recognition; thus, this study proposed a two-stage convolutional neural network to detect scaphoid fractures. In the first stage, the scaphoid bone is separated from the X-ray image using the Faster R-CNN network. The second stage uses the ResNet model as the backbone for feature extraction, and uses the feature pyramid network and the convolutional block attention module to develop the detection and classification models for scaphoid fractures. Various metrics such as recall, precision, sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve (AUC) are used to evaluate our proposed method's performance. The scaphoid bone detection achieved an accuracy of 99.70%. The results of scaphoid fracture detection with the rotational bounding box revealed a recall of 0.789, precision of 0.894, accuracy of 0.853, sensitivity of 0.789, specificity of 0.90, and AUC of 0.920. The resulting scaphoid fracture classification had the following performances: recall of 0.735, precision of 0.898, accuracy of 0.829, sensitivity of 0.735, specificity of 0.920, and AUC of 0.917. According to the experimental results, we found that the proposed method can provide effective references for measuring scaphoid fractures. It has a high potential to consider the solution of detection of scaphoid fractures. In the future, the integration of images of the anterior–posterior and lateral views of each participant to develop more powerful convolutional neural networks for fracture detection by X-ray radiograph is probably important to research.

Keywords: convolutional neural network; convolutional block attention module; faster R-CNN; feature pyramid network; scaphoid fractures

1. Introduction

The scaphoid is the largest carpal bone in the human wrist and is close to the carpals and radius, see Figure 1 [1]. Its unique location and shape are prone to fracture as people fall and their palms strike any hard surface. The standard treatment of scaphoid is screw bone surgery because of its good recovery and short treatment time. However, the screw bone surgery needs precise positioning of the scaphoid and its fracture to plan an appropriate angle to implant screws. The small size of the scaphoid and the complex structures of carpals cause accurate screw implantation to be difficult, and thus it is a challenge.
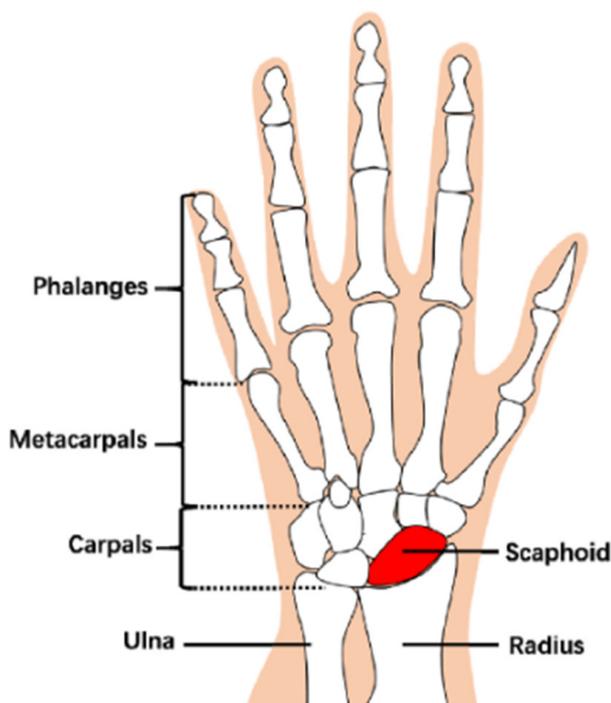
Figure 1. The anatomical structures of the wrist bone, in which the region with red is the scaphoid.

A scaphoid fracture is usually diagnosed by an X-ray of the wrist [2]. However, a break in the bone that cannot yet be seen on X-ray is called an "occult" fracture. The occult fractures may be difficult to detect by eye observations because their occurrence probability has been estimated at 7–21% reported in a recent prospective study [3]. If pain persists, a follow-up exam and X-ray in a week or two can be used to diagnose it. Sometimes, a CT scan or MRI is used to obtain better views of the shape and alignment of the scaphoid and assist with the diagnosis or surgery plans, but it is very expensive. Therefore, it is critical to develop a more precise X-ray diagnosis of scaphoid fractures.

Image processing technologies are widely used to separate the region segmentation methodology of X-ray images, but they always need manual intervention to decide the boundary of scaphoid fractures. To date, the convolutional neural network (CNN) has advanced development worldwide, successfully being applied to several areas of medical diagnosis and robotics [4–6]. Langerhuized et al. [7] used CNN to detect scaphoid fractures on the conventional radiographs. In this paper, two consecutive CNNs are developed. One is used for scaphoid segmentation, while another is for fracture detection. The segmentation CNN localized the scaphoid and then passed them into the detection CNN for fracture detection. The resulting performance is the Area Under the Curve (AUC) of Receiver Operating Characteristic (ROC) of 0.77, 72% accuracy, 0.84 sensitivity, and 0.6 specificity.

Yoon et al. [8] isolated the scaphoid area in a bounding box by using the cascade RCNN model [9] and then fed them to the EfficientNetB3 neural network [10] to determine whether the scaphoid was fractured. The results of the cascade R-CNN model achieved an overall sensitivity and specificity of 87.1% and 92.1% (AUC = 0.995). The second EfficientNetB3 obtained an overall sensitivity of 79.0% and specificity of 71.6% with an AUC of 0.810.

Hendrix et al. [11] proposed two consecutive CNNs including a segmentation CNN for scaphoid segmentation and a detection CNN for detecting the fractures. The segmentation CNN localized and then cropped scaphoid area. The cropping area was resized into a fixed size and normalized its contrast. The detection CNN was based on the DenseNet-121 model. A class activation map was then calculated using the smooth Grad-CAM++ algorithm [12]. The resulting segmentation CNN achieved a Dice of Coefficient (DSC) of 97.4% and symmetric Hausdroff Distance (HD) of 1.31 mm. The detection CNN had an overall sensitivity of 78.0% and specificity of 84% with an AUC of 0.87.

Tung [13] also used two CNNs to detect scaphoid fractures using the YOLO-v4 CNN model for scaphoid area detection and classification CNN to determine whether the detected scaphoid has fractured. The different backbones such as the VGG, ResNet [14], DenseNet [15], InceptionNet [16], and EfficientNet used to construct classification CNN. The experimental results showed that DenseNet 201 and ResNet 101 are more promising selections. The reported performances of the DenseNet 201 backbone are sensitivity of 0.833, specificity of 0.611, precision of 0.682, F1-score of 0.750, AUC of 0.444, and accuracy of 0.722.

Even though many articles have attempted to assess the detection issues of fractures in the X-ray images, the detection results of those articles do not meet the requirement of clinical diagnosis. The paper proposes a two-stage fracture detection of scaphoid images based on the convolutional neural networks. The contribution of this study can be described as follows:

- To increase detection accuracy, the proposed method consists of two CNNs: one is to identify the scaphoid area, and another is to detect the fractures of scaphoid.
- Identifying scaphoid areas can reduce the searching space and computational times for consequent fracture detection.
- A powerful fracture detection CNN consists of ResNet, spatial feature pyramid (SSP), and convolutional block attention module (CBAM). Experimental results showed that the proposed CNN achieved high detection performance.
- The rest of this paper is organized as follows. The material data set and methods are proposed in detail in Section 2. Section 3 then presents the experimental results and empirical discussion. Finally, Section 4 provides the conclusions of this study.

2. Material and Method

This section describes the data collections, data augmentations, scaphoid bone detection, and fracture detection.

2.1. Data Collections and Implementation Environment

In this study, all experimental X-ray images were collected from the National Cheng Kung University Hospital (NCKUH) in Taiwan, including 280 adult patients. Specifically, 178 fracture instances of surgical verification were selected as the positive images, and 102 normal instances were considered the negative images. All programs were implemented in a personal computer with an i9 processing unit, 16 GB of main memory and a GeForce RTX 3090 GPU using the Windows 10 operating system. The performance of our proposed system is measured by 5-fold cross-validation method.

2.2. Methods

The proposed CNNs include the scaphoid detection CNN and the fractures detection CNN. The first one is based on the original Faster R-CNN model [17], and the other fractures detection is based on the rotation-decoupled anchors [18] to finely detect the fracture areas shown in Figure 2. The two CNNs are described as follows.
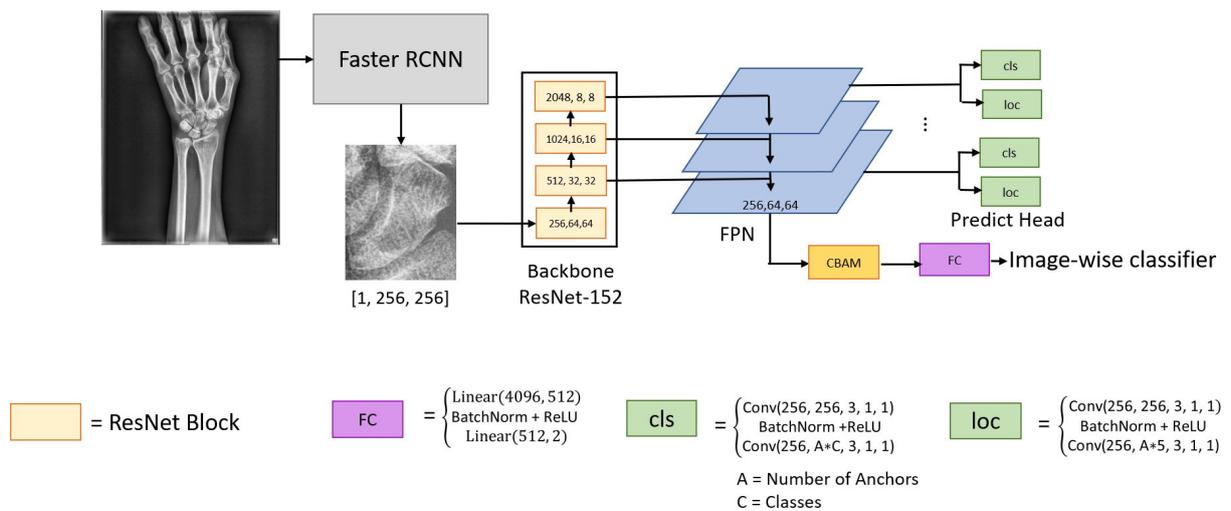
Figure 2. The CNNs used include scaphoid detection and fracture detection.

2.2.1. Scaphoid Area Detection

The traditional CNN uses a sliding kernel to convolute the whole X-ray image to generate the feature maps for further object detection or segmentation. Yet, the scaphoid area must be limited to a small area of an X-ray image. In this paper, the scaphoid area detection used is the Faster R-CNN, which is faster and more precise than Fast R-CNN [19,20]. The Faster R-CNN consists of three components: feature map convolution network, region proposal network (RPN), and Fast R-CNN as shown in Figure 3.
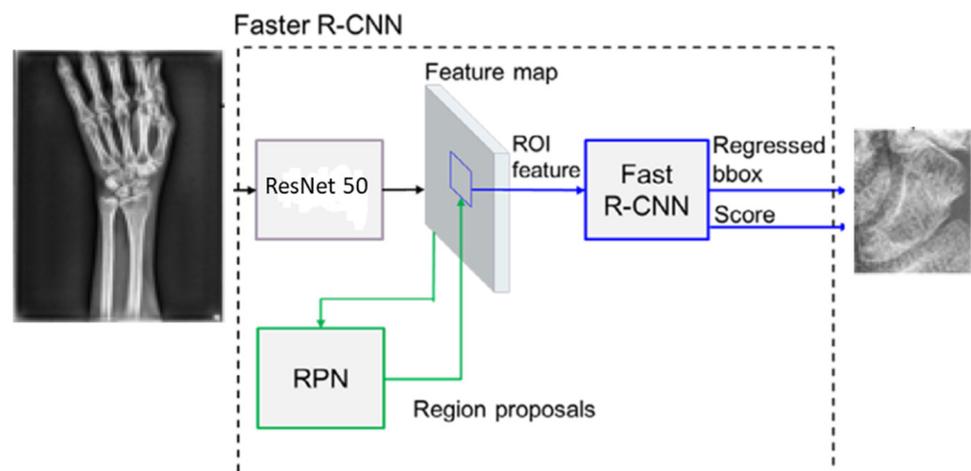
Figure 3. The structure of the Faster R-CNN model.

The feature map convolution network was implemented by the ResNet 50 that consists of five stages. Each stage includes max pooling, the residual block, and convolutional + batch normalization + ReLU block were applied as the backbone for generating feature maps. The original X-ray image is resized into 1600×1200 pixels and the last output feature map was 75×100 with 1024 channels. In addition, the RPN and the Fast R-CNN shared this feature map. Typically, the RPN referred to box regressions and confidence scores of searching objects to generate several bounding boxes with different sizes. It then predicted objects by these bounding boxes and integrated them into some proposals. The corresponding region of interest (ROI) on the feature maps of these proposals was transferred to the Fast R-CNN for further use. The Fast R-CNN received proposals from the RPN with the corresponding ROI features from the shared feature maps. Different sizes of ROI features were max-pooled to a 7×7 feature map. The fixed-size feature map was

fed into a sequence of fully connected layers and then connected into two sibling layers for classification and bounding box regression. The classification gave the detection confidence score, and the regression gave the position regression of the bounding box. The used loss function includes the classification loss and the regression loss in the training stage of the Faster R-CNN model.

2.2.2. Fracture Area Detection

In this paper, the fractures detection CNN consists of the ResNet 152 backbone, the feature pyramid network [21,22] (FPN, for generating the multiscale feature maps) and the convolutional block attention module [23] (CBAM; for determining whether it is fractured). The last three blocks comprise the feature pyramid network. The number of layers used by the FPN is three rather than the original four because of better performances in experiments. The resulting feature maps of FPN are fed into the predict head for detecting the fracture areas. The predict head has two branches: the *cls* branch (detecting positive/negative bounding box) and the *loc* branch (regressing the position of fracture bounding box). The two branches are implemented in Conv + BN + ReLU + Conv, and the shapes of the output of cls are [B, A × C] and loc is [B, A × 5] (where B, A, C are denoted to the numbers of the batch, anchor and class).

The detailed structure of FPN is shown in Figure 4. The FPN passes the deep features layer by layer and directly combines or concatenates the deep and shallow features to generate high-resolution feature maps and desirable semantics.
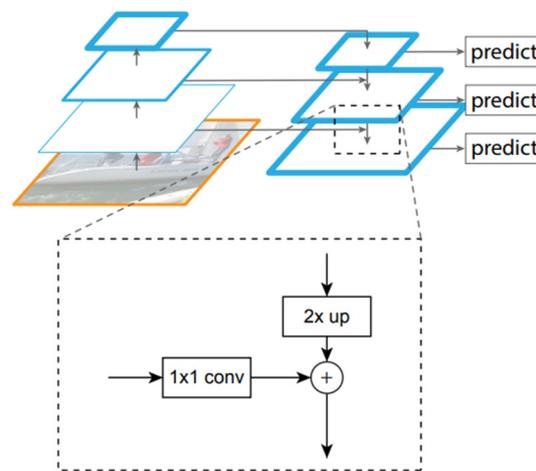

Figure 4. The structure of the feature pyramid network; ⊕ is the elementwise sum.

The CBAM enhances the input feature F of FPN into refined feature F' by inferring the one-dimensional (1D) channel attention map $N_c \in R^{1 \times 1 \times C}$, and the 2D spatial attention map $N_s \in R^{H \times W \times 1}$ as shown in Figure 5 [24,25]. The entire attention process can be summarized as follows:

$$F' = N_s(N_c(F) \odot F) \odot N_c(F) \oplus F \tag{1}$$

where $\odot$ and $\oplus$ denote the elementwise product and sum. The following describes the two attention modules as shown in Figure 6.
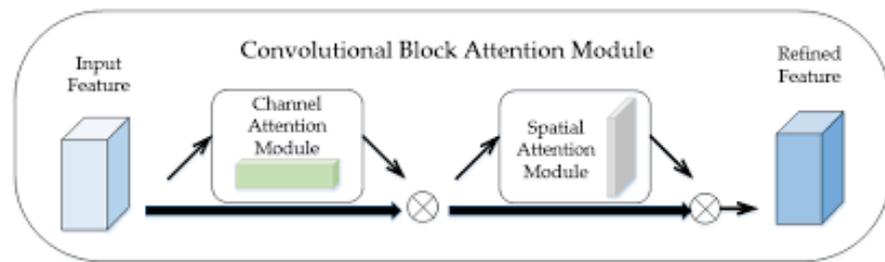
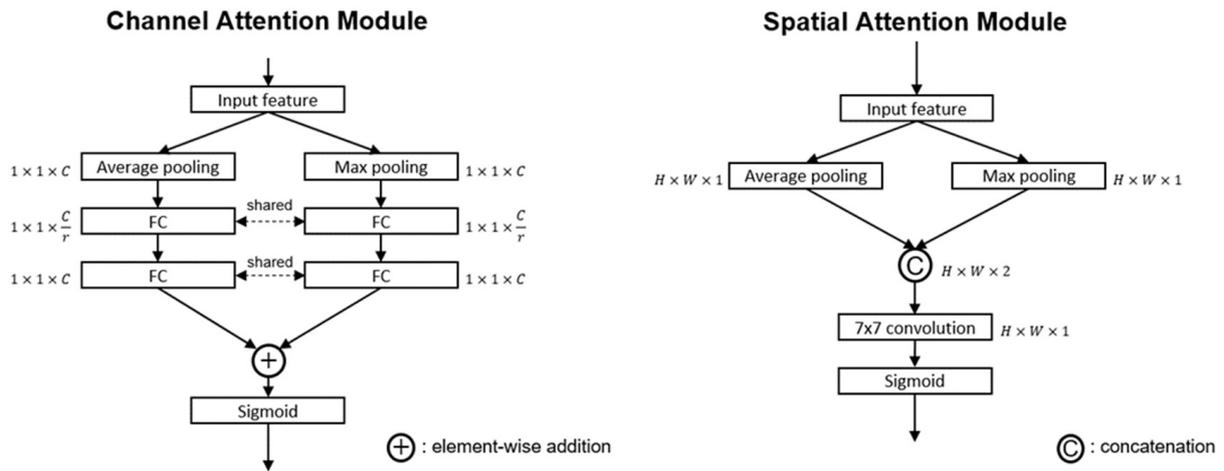**Figure 5.** The structure of the convolutional clock attention module.

Figure 6. The structure of the attention module, *H*, *W*, and *C* are the height, width, and channel of feature map.

- Channel attention module. The aggregated feature map is combined through average-pooling and max-pooling operations. The two resulting feature maps are represented as F_{ave}^c and F_{max}^c with the sizes of $1 \times 1 \times c$. These two feature maps are independently built from two fully connected layers and then integrate them into a feature map $N_c(F)$ by elementwise addition and sigmoid operations.
- Spatial attention module. The channel attention feature map $N_c(F)$ and original feature map F are aggregated into the new feature map, and down-sampled by using the average-pooling and max-pooling operation. The results generate two different feature maps F_{ave}^s and F_{max}^s, with sizes of $H \times W \times 1$. The two feature maps are concatenated into larger maps and then convoluted and activated by a $7 \times 7 \times 2$ kernel and sigmoid function.

The image-wise classifier is established with a full connected network block. The binary classification cross-entropy loss used is defined as follows.

$$L_{image}BCE(y, p) = \begin{cases} -\log(p) & if\ y = 1 \\ -\log(1 - p) & if\ y \neq 1 \end{cases} \tag{2}$$

where y represents a scaphoid image and $y = 1$ means the scaphoid image y has fractured, and the $y \neq 1$ means no fracture occurs.

A rotation-decoupled detector (RDD) of the predicted head of this paper is used to detect oriented bounding box (OBB) with additional rotation angle because the fractured area of scaphoid always emerges with arbitrary oriented angles, dense distribution and a larger aspect ratio. The horizontal bounding box (HBB) is usually represented by (x, y, w, h), where (x, y) is the center, and w and h are the lengths of the X and Y axes. The OBB is represented by (x, y, w, h, θ) where θ is the angle of the bounding box. To combine the advantages of HBB and OBB, the detecting rotational bounding box is redefined as the

horizontal bound box $HBB_{h/v}$ and an angle θ ranged in $[-\pi/4, \pi/4]$, where the detected OBB has the same center, width, and length as the $HBB_{h/v}$ horizontal bound box. To accelerate the training and interfacing process, only horizontal anchors are used. In the anchor matching, the rotational ground truth box is decoupled to a $HBB_{h/v}^T$ and an acute angle θ_T for matching. The computation of intersection of union (IOU) between OBB and ground truth is disregard of their angles, as shown in Figure 7 [25].
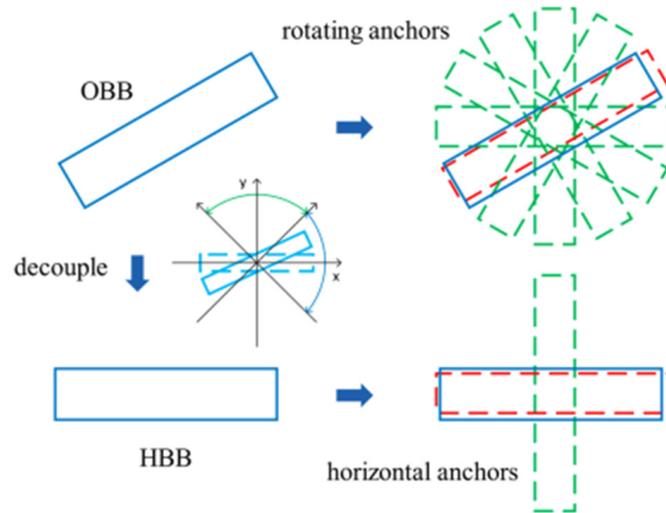
Figure 7. Matching methods of the ground truth bounding box and the rotational bounding box.

After matching, most of the bound boxes are classified as positive objects or negative backgrounds based on the IOU thresholds set to 0.5 and 0.4. In other words, if the IOU of the bounding box is more than 0.5 or less than 0.4, it will be assigned the positive or negative case, and the remaining one is considered the ignored bounding box. In general, the number of negative bounding boxes is much larger than the positive case. The positive-negative samples imbalance problem always results in inaccurate classification. The popular balance strategy is Focal Loss [26]; the Focal Loss method assigned the categorical labels of positive, negative, and ignored bounding to 1, 0, and −1. The corresponding binary cross-entropy (BCE) is defined as follows:

$$BCE(y, p) = \begin{cases} \log(p) & if \ y = 1 \\ -\log(1-p) & if \ y = 0 \\ 0 & if \ y = -1 \end{cases} \tag{3}$$

where y is the class label of a bounding box and p is its predicted probability.

The classification Loss L_{cls} is defined as follows.

$$L_{cls} = \frac{1}{\#(postive \ bboxes)} \sum_{i=1}^{N} BCE(y_i, p_i) \tag{4}$$

where the N and the #(*positive boxes*) denote the number of all bounding boxes and the number of positive bounding box.

The ground truth box and prediction box are represented as $v = (t_x, t_y, t_w, t_h, t_\theta)$ and $v^* = \left(t_x^*, t_y^*, t_w^*, t_h^*, t_\theta^*\right)$ for position regression. If a corresponding anchor is expressed as (x_a, y_a, w_a, h_a), the v^* is defined in the following equations:

$$
\begin{aligned}
t_x &= x - x_a/w_a, \ t_y = y - y_a/h_a \\
t_w &= \log(w/w_a), \ t_h = \log(h/h_a) \\
t_\theta &= 4\theta/\pi \\
t_x^* &= x^* - x_a/w_a, \ t_y^* = y^* - y_a/h_a \\
t_w^* &= \log(w^*/w_a), \ t_h^* = \log(h^*/h_a) \\
t_\theta^* &= \tan h(\theta^*)
\end{aligned}
\tag{5}
$$

The $smooth_{L1}$ loss for the rotation bounding box regression is used for the object anchors.

$$
L_{reg} = \frac{1}{\#(postive \ bboxes)} \sum_{i=1}^{\#(postive \ bboxes)} smooth_{L1}(v_i^* - v_i)
\tag{6}
$$

where,

$$
smooth_{L1}(x) = \begin{cases} 0.5x^2 & if \ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}
\tag{7}
$$

Finally, the multi-task loss is defined as:

$$
L = L_{cls} + \alpha L_{reg} + \frac{1}{\beta} L_{image} BCE(y, \ p)
\tag{8}
$$

where β is the batch size.

2.2.3. Performance Evaluation

In this section, the performance metrics of the proposed system were evaluated [27–30]. There are four performance metrics, including the *Accuracy*, *Recall*, *Precision*, and F *scores* are defined as:

$$
Accuarcy(AC) = \frac{TP + TN}{TP + TN + FP + FN}
\tag{9}
$$

$$
Sensitivity(SE) = \frac{TP}{TP + FN}
\tag{10}
$$

$$
Specificity(SP) = \frac{TN}{TN + FP}
\tag{11}
$$

$$
Recall(R) = \frac{TP}{TP + FN}
\tag{12}
$$

$$
Precision = \frac{TP}{TP + FP}
\tag{13}
$$

$$
F - scores = \frac{2 \times Precision \times Recall}{Precision + Recall}
\tag{14}
$$

The *TP*, *TN*, *FP*, and *FN* denote the number of true positive, true negative, false positive, and false negative.

In addition, the Receiver Operating Characteristic (ROC) curve [31,32] is a graphical plot that illustrates the capability of a binary classification system. The ROC curve is created by plotting the true positive rate against the false-positive rate at different thresholds. The Area Under the Curve (AUC) is a powerful measure of the ability of a classifier to distinguish between classes. The AUC measures the ability of a classifier to distinguish classes and is used as a summary of the ROC curve. The AUC ranges from 0 to 1. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes.

3. Results

This section describes the experimental results of the scaphoid and fracture detection.

3.1. Results of Scaphoid Detection

A total of 361 X-ray scaphoid radiographs were used including the 167 fractured samples and 194 normal samples. Approximately, there were 18.56% (i.e., 31/167) occult fractured samples in the 167 fractured samples. Figure 8 shows the two examples of the X-ray images with fracture and without fracture. All sample images were verified using the 5-fold cross-validation method. Three different data augmentation methods were used; (1) contrast limited adaptive histogram equalization (CLAHE) [33], (2) random horizontal flip with 50% probability, and (3) random contrast with 50% probability. The used CNN is the Faster R-CNN model with a backbone of ResNet 50. The trained strategy is the learning rate of 0.001, optimizer is SGD, batch size of 1 and epochs of 10,000. In total, the number of parameters of the CNN model of scaphoid detection is 41,755,286 and the parameters are pre-trained by using the PASCAL-VOC [34]. The training time of CNN is about 61 min and the prediction time of each test image is 0.267 s. The classification accuracy tested in the 361 samples based on the five times of 5-fold cross-validations is 0.997 in which only one fracture scaphoid image failed to detect. Similar result of classification accuracy of 0.994 was also found in Tung's work [13].
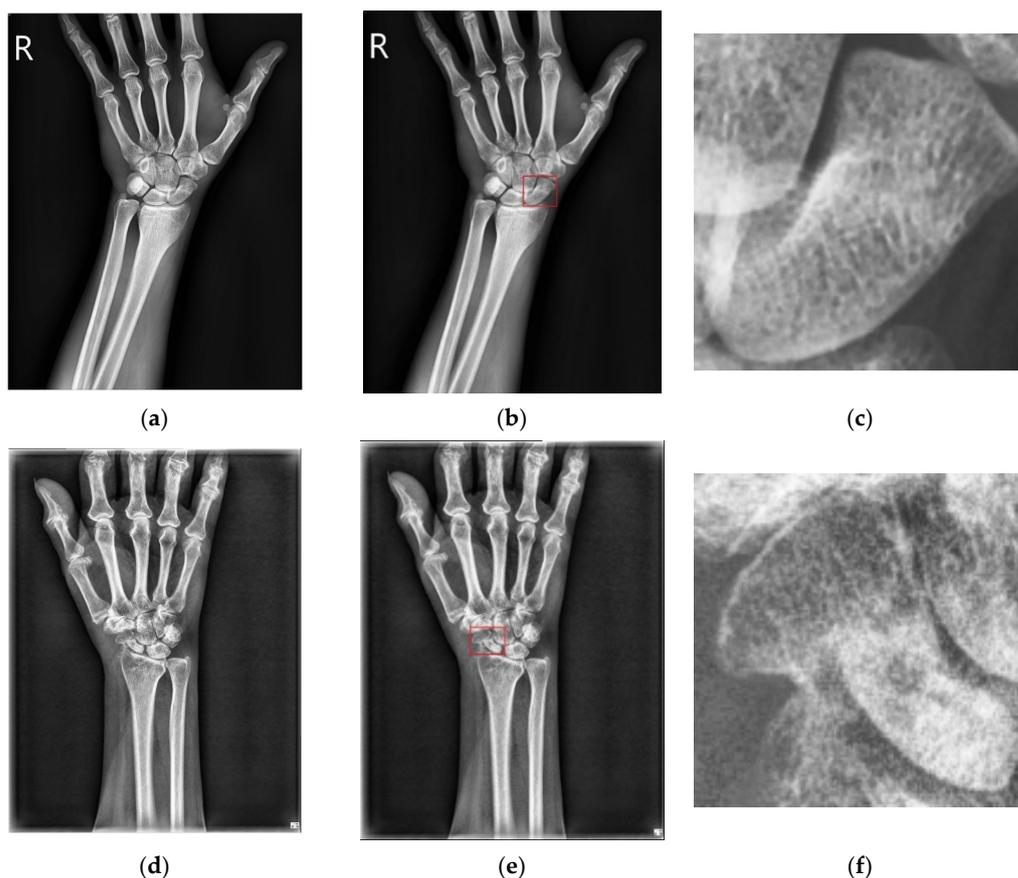

Figure 8. Two image samples with and without fractures. (**a**) Normal case. (**b**) Scaphoid area marked red of (**a**). (**c**) Enlarged scaphoid area of (**b**). (**d**) Fractured case. (**e**) Scaphoid area of (**a**). (**f**) Enlarged scaphoid area of (**b**).

3.2. Fracture Detection and Classification

The detected scaphoid images also include the 166 fractured samples (previous one scaphoid image is discarded) and 194 normal samples. Figure 9 shows the two examples of the detected scaphoid images with/without fractures. The data augmentation and

5-fold cross-validation methods are used the same as the scaphoid detection. The trained strategy is the learning rate of 0.001, optimizer is SGD, batch size of 1, and epochs of 10,000. As shown in Tables 1 and 2, we compute the averages of recall, precision, sensitivity, specificity, accuracy, and AUC using the five times of 5-fold cross-validations methods. In total, the number of parameters of the CNN model of fracture detection is 66,913,876 and the parameters are pre-trained by using the PASCAL-VOC [34]. The training time of CNN is about 37 min and prediction time of each test image is 0.028 s.
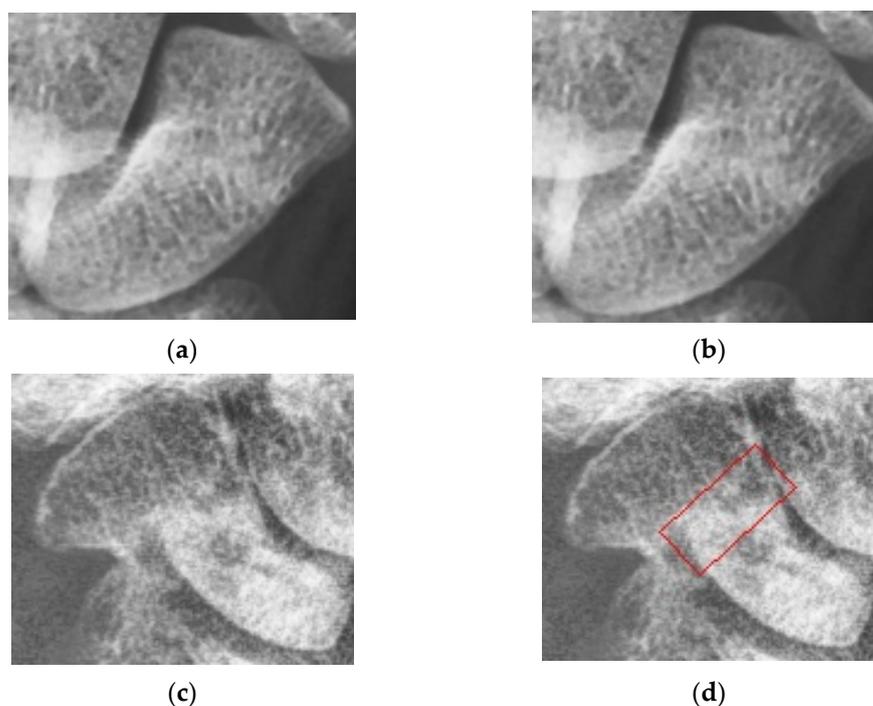
Figure 9. Two scaphoid cases with/without fractures. (**a**) Scaphoid case without fracture. (**b**) No fracture bounding box. (**c**) Scaphoid case with fracture. (**d**) Fracture bounding box.

Table 1. The results of fracture detection of detected scaphoid.

Methods	Recall	Precision	Accuracy	Sensitivity	Specificity	AUC
Our proposed method	0.789	0.894	0.853	0.789	0.900	0.920

Table 2. The results fracture classification of detected scaphoid.

Methods	Recall	Precision	Accuracy	Sensitivity	Specificity	AUC
Our proposed method	0.735	0.898	0.829	0.735	0.920	0.917

Compared to other works of fracture detections, Yoon [6] presented the sensitivity of 0.79, the specificity of 0.716, and the AUC of 0.81. From Table 1, the sensitivity of 0.789, the specificity of 0.900, and the AUC of 0.92 are found; these performance measures are superior to the results of Yoon, especially as the specificity of our proposed method markedly precedes the work of Yoon.

The works of the fracture classifications in the literature showed that Tung [13] presented a sensitivity of 0.833, specificity of 0.611, and AUC of 0.79. Yoon [8] presented a sensitivity of 0.87, specificity of 0.92, and AUC of 0.96. Langerhuizen [7] presented a sensitivity of 0.84, specificity of 0.6, and AUC of 0.81. Hendrix [11] presented a sensitivity of 0.78, specificity of 0.84, and AUC of 0.87. While our result is worse than the work of Yoon, it is superior to other three methods.

However, the recalls of our proposed results for the fracture detection and classification are only 0.789 and 0.735. The results mean that a higher false-negative rate occurs, meaning there may exist a large number of misclassifications of positive images with occult fractures. To understand the results of the 31 occult fracture samples, their detection and classification results are shown in Table 3. The results showed that the correct detection and classification is only 50%.

Table 3. The results of fracture detection and classification of 31 occult samples.

Methods	Accurate Samples	Inaccurate Sample
Detection	15	16
Classification	16	15

4. Conclusions

This study delivers evidence supporting the CNN approaches that accurately detect the fracture areas and classify the status of scaphoid fractures using X-ray scaphoid radiographs. The proposed model includes two consecutive CNNs to detect the scaphoid and classify scaphoid fractures. Experimental results reveal that the detection of the scaphoid area performs well, but the recalls of the detection and classification of scaphoid fractures are only 0.789 and 0.735. It therefore implies high false-negative rates and warrants further exploring the results of 31 occult samples. We found that only 50% correct fracture classification and detection were reached in testing of the 31 occult samples.

In the future, the improvement of the classification capability of occult samples is still an important research topic. The integration of the anterior-poster and lateral view images of each participant is a probable approach to developing more powerful convolutional neural networks for fracture detection with X-ray radiographs.

References

1. Chen, C.; Liu, B.; Zhou, K.; He, W.; Yan, F.; Wang, Z. CSR-Net: Cross-Scale Residual Network for multi-objective scaphoid fracture segmentation. *Comput. Biol. Med.* **2021**, *137*, 104776. [CrossRef] [PubMed]
2. Schmitt, R.; Rosenthal, H. Imaging of scaphoid fractures according to the new S3 guidelines. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*; Georg Thieme Verlag KG: Stuttgart, Germany, 2016; Volume 188, pp. 459–469.

3. Gibney, B.; Smith, M.; Moughty, A.; Kavangh, E.C.; Hynes, D.; MacMahon, P.J. Incorporating cone-beam CT into the diagnostic algorithm for suspected radiocarpal fractures: A new standard of care? *AJR Am. J. Roentgenol.* **2019**, *213*, 1117–1123. [CrossRef] [PubMed]
4. Rainey, C.; McConnell, J.; Hughes, C.; Bond, R.; McFadden, S. Artificial intelligence for diagnosis of fractures on plain radiographs: A scoping review of current literature. *Intell.-Based Med.* **2021**, *5*, 100033. [CrossRef]
5. Mathappan, N.; Soundariya, R.S.; Natarajan, A.; Gopalan, S.K. Bio-medical analysis of breast cancer risk detection based on deep neural network. *Int. J. Med. Eng. Inform.* **2020**, *12*, 529–541. [CrossRef]
6. Chen, A.I.; Balter, M.L.; Maguire, T.J.; Yarmush, M.L. Deep learning robotic guideline for autonomous vascular access. *Nat. Mach. Intell.* **2020**, *2*, 104–115. [CrossRef]
7. Langerhuizen, D.W.G.; Bulstra, A.E.J.; Janssen, S.J.; Ring, D.; Kerkhoffs, G.M.M.J.; Jaarsma, R.; Doornberg, J.N. Is deep learning on par with human observers for detection of radiographically visible and occult fractures of the scaphoid? *Clin. Orthop. Relat. Res.* **2020**, *478*, 2653–2659. [CrossRef] [PubMed]
8. Yoon, A.P.; Lee, Y.L.; Kane, R.L.; Kuo, C.F.; Lin, C.; Chung, K.C. Development and validation of a deep learning model using convolutional neural networks to identify scaphoid fractures in radiographs. *JAMA Netw. Open* **2021**, *4*, e216096. [CrossRef] [PubMed]
9. Zhaowei, C.; Nuno, V. Cascade R-CNN: Delving into high quality object detection. *arXiv* **2017**, arXiv:1712.00726v1.
10. Mingxing, T.; Quoc, V.L. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946v5.
11. Hendrix, N.; Scholten, E.; Vernhout, B.; Bruijnen, S.; Maresch, B.; de Jong, M.; Diepstraten, S.; Bollen, S.; Schalekamp, S.; de Rooij, M.; et al. Development and validation of a convolutional neural network for automated detection of scaphoid fractures on conventional radiographs. *Radiol. Artif. Intell.* **2021**, *3*, e200260. [CrossRef] [PubMed]
12. Omeiza, D.; Speakman, S.; Cintas, C.; Weldermariam, K. Smooth grad-CAM++: An enhanced inference level visualization technique for deep convolutional neural network models. *arXiv* **2019**, arXiv:1908.01224.
13. Tung, Y.C.; Su, J.H.; Liao, Y.E.W.; Chang, C.D.; Cheng, Y.F.; Chang, W.C.; Chen, B.H. High-performance scaphoid fracture recognition via effectiveness assessment artificial neural networks. *Appl. Sci.* **2011**, *11*, 8485. [CrossRef]
14. He, K.; Zhang, X.; Ren, S.; Su, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03385.
15. Huang, G.; Liu, Z.; Maaten, L.V.D.; Kilian, Q.; Weinberger, K.Q. Densely connected convolutional networks. *arXiv* **2018**, arXiv:1608.06993v5.
16. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, R. Going deeper with convolutions. *arXiv* **2014**, arXiv:1409.4842v1.
17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-CNN: Towards real-time object detection with region proposal Networks. *arXiv* **2015**, arXiv:1506.01497v3. [CrossRef]
18. Yang, X.; Yan, J.; Ming, Q.; Wang, W.; Zhang, X.; Tian, Q. Rethinking rotated object detection with Gaussian wasserstein distance loss. *arXiv* **2021**, arXiv:2101.11952v3.
19. Girshick, R. Fast r-cnn. *arXiv* **2015**, arXiv:1504.08083.
20. Kuok, C.-P.; Horng, M.-H.; Liao, Y.-M.; Chow, N.-H.; Sun, Y.-N. An effective and accurate identification system of Mycobacterium tuberculosis using convolution neural networks. *Microsc. Res. Tech.* **2019**, *82*, 709–719. [CrossRef] [PubMed]
21. Lin, T.Y.; Dollar, P.; He, K.; Hariharan, B.; Elomgie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
22. Gu, Y.; Qin, X.; Peng, Y.; Li, L. Content-augmented feature pyramid network with light linear spatial transformers for object detection. *arXiv* **2021**, arXiv:2105.09464v2.
23. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. *arXiv* **2018**, arXiv:1807.06521.
24. Xiong, S.; Wu, X.; Chen, H.; Qing, L.; Chen, T.; He, X. Bi-directional skip connection feature pyramid network and sub-pixel convolution for high-quality object detection. *Neurocomputing* **2020**, *440*, 185–196. [CrossRef]
25. Zhong, B.; Ao, K. Single-stage rotation-decoupled detector for an oriented object. *Remote Sens.* **2020**, *12*, 3262. [CrossRef]
26. Oksuz, K.; Cam, B.C.; Kalkan, S.; Akbas, E. Imbalance problems in object detection: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3388–3415. [CrossRef] [PubMed]
27. Gupta, V.; Mittal, M. R-peak detection for improved analysis in health informatics. *Int. J. Med. Eng. Inform.* **2021**, *13*, 213–223.
28. Gupta, V.; Mittal, M.; Mittal, V. FrWT-PPCA-Based R-peak Detection for Improved Management of Healthcare System. *IETE J. Res.* **2021**. [CrossRef]
29. Gupta, V.; Mittal, M.; Mittal, V.; Saxena, N.K. A Critical Review of Feature Extraction Techniques for ECG Signal Analysis. *J. Inst. Eng.* **2021**, *102*, 1049–1060. [CrossRef]
30. Gautam, D.D.; Giri, V.K. A Critical Review of Various Peak Detection Techniques of ECG Signals. *i-Manag. J. Digit. Signal Process.* **2016**, *4*, 27–36.
31. Zou, K.H.; O'Malley, A.J.; Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **2007**, *115*, 654–657. [CrossRef] [PubMed]
32. Brown, C.D.; Davis, H.T. Receiver operating characteristic curves and related decision measures: A tutorial. *Chemom. Intell. Lab. Syst.* **2006**, *80*, 24–38. [CrossRef]

33. Agustin, T.; Utami, E.; Fatta, H.A. Implementation of data augmentation to improve performance CNN method for detecting diabetic retinopathy. In Proceedings of the International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 November 2020.

34. Li, H.; Singh, B.; Nsjibi, M.; Wu, Z.; Davis, L.S. An Analysis of Pre-Training on Object Detection. *arXiv* **2019**, arXiv:1904.05871.