applied sciences

MDPI

Article

High-Performance Scaphoid Fracture Recognition via Effectiveness Assessment of Artificial Neural Networks

Yu-Cheng Tung [1], Ja-Hwung Su [2,*], Yi-Wen Liao [3], Ching-Di Chang [1], Yu-Fan Cheng [1], Wan-Ching Chang [1] and Bo-Hong Chen [3]

[1] Department of Diagnostic Radiology, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung 83347, Taiwan; yctung66@gmail.com (Y.-C.T.); mr6777@cgmh.org.tw (C.-D.C.); prof.chengyufan@gmail.com (Y.-F.C.); o927003551@gmail.com (W.-C.C.)
[2] Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 83347, Taiwan
[3] Department of Information Management, Cheng Shiu University, Kaohsiung 83347, Taiwan; k0632@gcloud.csu.edu.tw (Y.-W.L.); 40711119@gcloud.csu.edu.tw (B.-H.C.)
* Correspondence: bb0820@ms22.hinet.net

Featured Application: This work will be materialized into the Bioimage Diagnostic system in Kaohsiung Branch of Chang Gung Hospital, Taiwan.

Abstract: Image recognition through the use of deep learning (DL) techniques has recently become a hot topic in many fields. Especially for bioimage informatics, DL-based image recognition has been successfully used in several applications, such as cancer and fracture detection. However, few previous studies have focused on detecting scaphoid fractures, and the related effectiveness is also not significant. Aimed at this issue, in this paper, we present a two-stage method for scaphoid fracture recognition by conducting an effectiveness analysis of numerous state-of-the-art artificial neural networks. In the first stage, the scaphoid bone is extracted from the radiograph using object detection techniques. Based on the object extracted, several convolutional neural networks (CNNs), with or without transfer learning, are utilized to recognize the segmented object. Finally, the analytical details on a real data set are given, in terms of various evaluation metrics, including sensitivity, specificity, precision, F1-score, area under the receiver operating curve (AUC), kappa, and accuracy. The experimental results reveal that the CNNs with transfer learning are more effective than those without transfer learning. Moreover, DenseNet201 and ResNet101 are found to be more promising than the other methods, on average. According to the experimental results, DenseNet201 and ResNet101 can be recommended as considerable solutions for scaphoid fracture detection within a bioimage diagnostic system.

Keywords: scaphoid fracture; image recognition; deep learning; artificial intelligence; convolutional neural networks

check for updates

Citation: Tung, Y.-C.; Su, J.-H.; Liao, Y.-W.; Chang, C.-D.; Cheng, Y.-F.; Chang, W.-C.; Chen, B.-H. High-Performance Scaphoid Fracture Recognition via Effectiveness Assessment of Artificial Neural Networks. *Appl. Sci.* **2021**, *11*, 8485. https://doi.org/10.3390/app11188485

Academic Editor: Seokwon Yeom

Received: 3 August 2021
Accepted: 10 September 2021
Published: 13 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

To date, artificial intelligence (AI) has advanced technologies worldwide, successfully being applied in many fields such as industry, commerce, agriculture, smart life, medical informatics, and so on. In these fields, image recognition is widely used to simulate human vision (so-called computer vision). In traditional computer vision, an image can be described by its color, texture, and shape; thereby, the objects in an image can be recognized by machine learning methods. However, the effectiveness of such approaches is limited in incomplete feature analysis. To address such problems, deep learning (DL) has been proposed as a solution, which can be viewed as a set of neural networks based on human neural biology and which can recognize objects by conducting iterative feature filtering. In fact, DL has been shown to be effective for the diagnosis of cancer and fractures. For

example, several online systems that allow the user to submit an image to detect the carcinomas have been developed. Normally, the accuracy can reach around 95%. In biomechanics, deep learning has been successfully used to detect general fractures in bioimages, including those in the axial skeleton, appendicular skeleton, skull, spine, ribs, sternum, and so on. Few previous studies have focused on scaphoid fracture detection. As shown in Figure 1, the scaphoid bone is one of the eight bones of the wrist and is the most common location for carpal fractures (i.e., about 70% of carpal fractures). Scaphoid fractures are mostly caused by the palms supporting the ground when falling, such that the wrists are excessively squeezed. Patients often have the wrist pain on the thumb side; this pain is especially intense during rotation. As the bone structure of the wrist is delicate and complex, it is not easy to discover the deformity from the appearance of the fracture, that is, if the fracture does not have much displacement, traditional X-ray examination often causes misjudgments and, thus, delay of treatment.



Figure 1. Example of the scaphoid bone.

Although some researchers have made attempts to address the issue above, the prediction result is not satisfactory. Therefore, in this paper, we propose a two-stage scaphoid fracture recognizer based on deep learning techniques. Overall, the contribution of this study over those in the literature can be summarized as follows:

- To increase the precision, the proposed method identifies and segments the scaphoid bone first. Based on this segmentation, the recognition space is reduced into a specific area and the computational cost is also reduced.
- To discover the most effective recognizer, we provide a detailed analysis and conduct a set of experiments considering numerous existing well-known CNNs using data augmentation and transfer learning.
- Afterward, a comprehensive empirical study on a real data set is presented, and an insightful analysis is proposed, in order to make a near-optimal recommendation from the testing CNNs.
- Finally, the proposed method will be materialized further in the bioimage diagnostic system in the Kaohsiung Branch of Chang Gung Hospital, Taiwan.

Basically, this paper can be viewed as an evaluation study, and the major intent behind it is to enhance the existing techniques through the use of segmentation, data augmentation, and transfer learning. By conducting an effectiveness assessment, we can confidently recommend AI networks for the desired purpose. The remainder of this paper is laid out as follows. A review of related works is provided in Section 2. In Section 3, the proposed method for scaphoid fracture detection is presented in detail. The empirical study is described in Section 4, and our conclusions are given in Section 5.

2. Related Works

Artificial intelligence has recently shifted bioinformatics toward an automated and effective field. Deep learning plays a critical role in this field, with applications in areas such as petroleum engineering [1], computer engineering [2], electrical engineering [3], biomedical engineering [4], software engineering [5], and energy engineering [6]. Although many diseases can be detected successfully by AI, an effective method for recognizing scaphoid fractures is not easy to pursue. This is the main aim of this paper. In this section, works in the literature related to AI-assisted bioinformatics are reviewed categorically.

2.1. Bioimage Recognition and Deep Learning

To date, bioinformatics has played an important role in risk assessment [7,8], disease detection [9], and healthcare [10], including sources of bioimages [11], biochemical tests [12], diagnostic reports [13], and so on. Therefore, several AI techniques, including data mining and deep learning, have been proposed for this topic. Among them, artificial neural networks (ANNs) have been widely and effectively used with bioimages. Regarding the chest, Tang et al. [14] tested several convolutional neural networks (CNNs) for abnormality classification, including AlexNet, ResNet, Inception, and DenseNet, among others. Furthermore, CheXNet and VGG19 have been used to extract the features from X-ray images for further classification of pediatric pneumonia [15]. Considering lung CT (computed tomography) images, Francis et al. [16] and Javan et al. [17] took advantage of Gated-SCNN and VGG-XGBoost, respectively, to recognize pulmonary lesions. Besides lung and osteosarcoma, deep learning has also been applied to segment the intracerebral hemorrhage in CT images [18–20]. For liver, Xia et al. [15] combined traditional multiclassification cross-entropy loss functions to perform better segmentation. Moreover, Yang et al. [21] constructed FCN-DecNet for coarse segmentation by optimization weighting. Other cancer image recognition (e.g., breast) has been performed using small SE-ResNet module [22].

2.2. Musculoskeletal Image Recognition and Deep Learning

Sato et al. [23] and Huang et al. [24] adopted CNNs to detect musculoskeletal abnormalities and to segment osteosarcoma tumors using radiograph and CT images, respectively. Rayan et al. [25] and England et al. [26] utilized CNNs to detect pediatric elbow fractures and effusions. Olczak et al. [27] used five networks to classify four types of radiographs, including fracture, laterality, body part, and exam view. Kim et al. [28] applied transfer learning from CNNs to enhance the prediction of fractures from radiographs. In addition to the above musculoskeletal image recognition, scaphoid fracture detection is the aim of the proposed paper. At present, scaphoid fractures can be detected manually by radiologists using radiographs and CT images. In terms of radiographs, Nazarova et al. [29] provided a study showing that different angling views can help radiologists to detect scaphoid fractures. In terms of CT images, the cone-beam CT [30] improved the prediction effectiveness significantly, in contrast to radiographs, for a radiologist. Apart from manual detection, Ozkaya et al. [31] compared the performances of automated recognition (CNN), emergency department physicians, and orthopedic specialists using radiographs. The related experimental results showed the experienced specialists performed the best, while the performance of the CNN was close to that of less-experienced specialists.

2.3. Comparative Study

Although numerous forerunners have been devoted to musculoskeletal image recognition, there exists room for further research on scaphoid fracture detection. Here, a comparative study is conducted, as shown in Table 1, which shows the primary points of the previous studies, including the data used, detection by deep learning, number of CNNs, detection on radiographs, transfer learning, data augmentation, number of evaluation metrics, and segmentations, thus indicating the uniqueness of this paper.

Table 1. Comparison of proposed work and related works.

	Considers Scaphoid Fracture	Detection by Deep Learning	Number of CNNs	Detection on Radiographs	Transfer Learning	Data Augmentation	Number of Evaluation Measures	Automated Segmentation
Sato et al. [23]	No	Yes	5	Yes	Yes	Yes	4	No
Huang et al. [24]	No	Yes	4	No	No	Yes	4	Yes
Rayan et al. [25]	No	Yes	2	Yes	No	Yes	1	No
England et al. [26]	No	Yes	1	Yes	No	Yes	1	No
Kim et al. [28]	No	Yes	1	Yes	Yes	Yes	1	No
Ozkaya et al. [31]	Yes	Yes	1	Yes	Yes	No	5	No
Olczak et al. [27]	Yes	Yes	3	Yes	Yes	No	1	No
Nazarova et al. [29]	Yes	No	0	Yes	No	No	1	No
Edlund et al. [30]	Yes	No	0	No	No	No	2	No
Proposed	Yes	Yes	5	Yes	Yes	Yes	7	Yes

3. Methods

3.1. Framework

As shown in Table 1, some past studies related to musculoskeletal images exist. Few studies have provided a comprehensive assessment for existing CNNs in scaphoid fracture recognition. This issue motivated us to conduct a comparative evaluation for numerous CNNs, providing practical insights for radiologists. The framework of the proposed method is shown in Figure 2, including the offline training and online recognition phases.

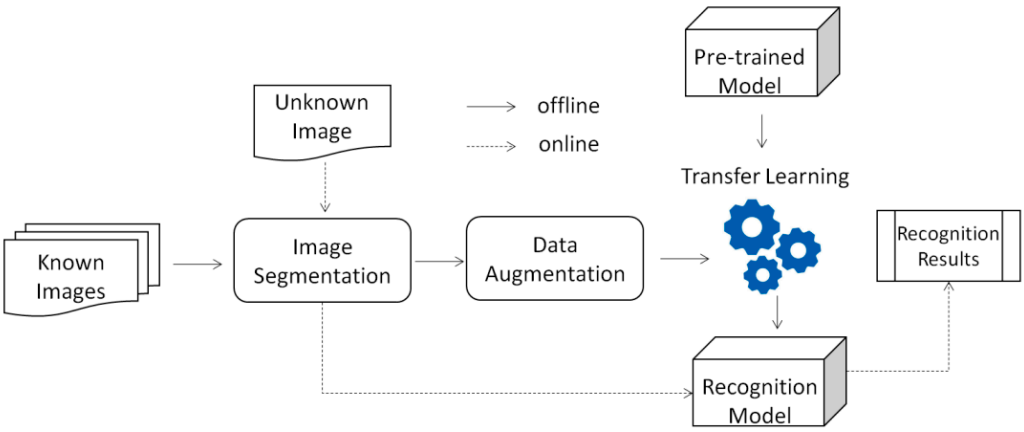


Figure 2. Framework of the proposed method.

- Offline Training

This is a foundational phase that constructs a recognition model through data collection, data preprocessing, and transfer learning. For data preprocessing, the scaphoid bones are extracted from the labeled images. Next, the extracted images are augmented. For transfer learning, the recognition model is constructed by transfer learning from the pretrained model.

- Online Recognition

This phase starts with the input of an unknown image. Then, the scaphoid bone is segmented from the unknown image. Based on the recognition model, this segmented object is classified by convolutional neural networks. In this paper, we test a set of CNNs to reveal their respective effectiveness.

3.2. Offline Training

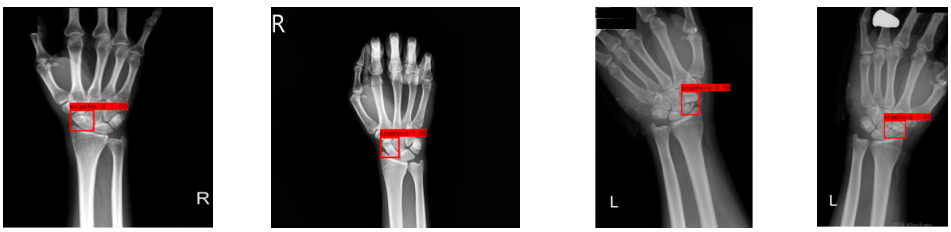
In this section, the data collection, data segmentation, data augmentation, and learning model are discussed.

3.2.1. Data Collection

The experimental data were gathered from Kaohsiung Chang Gung Memorial Hospital, Taiwan, including 154 adult patients (called KCGMH data in Section 4), that is, the aim of this paper is the adult patient fractures. In overall, this data contained types of en-face and en-profile images, and the en-face images were chosen as the filtered data first. In this filtered data, there were 178 clearly positive (fracture) instances and 308 negative (normal) instances, as identified by radiologists. Hence, 178 instances were selected as the examined data, for each of the positive and negative sets, that is, there were 356 images in total. Finally, 70% of positive and negative data were randomly selected as the training data, while the others were taken as the testing data [31]. For CNNs, 10% of training data was split as the validation data. Note that the reason for using these collected data, instead of the former data, can be explained by the following concerns. First, the medical data are not easy to collect, due to privacy concerns. Second, the data sets of former methods are not public. Third, without permission from patients, it is not appropriate to experiment on collected data, from an ethical point of view.

3.2.2. Data Segmentation

Traditionally, a CNN uses a sliding window to filter the features for the whole image. However, in our context, the recognition quality is limited when using the whole image. This is because the scaphoid bone is only a subarea of the image. Considering this issue, in this paper, the scaphoid bone was extracted from the image before training and testing, that is, the scaphoid bone is regarded as an object, which is segmented to narrow the recognition space from the whole image to a subimage. In this paper, the object detector used is YOLO-v4 (You Only Look Once-version 4) [32], which is an efficient and effective method extended from YOLO-v3 [33]. In the related literature, YOLO-v4 is composed of three main parts, namely a backbone, neck, and head, where the backbone is CSPDarknet53 [34], the neck contains SPP [35] and PANet [36], and the head is YOLO-v3. Note that, for the segmentation in this paper, the number of training instances was 408 and the accuracy was around 96%. Figure 3a,b show examples of segmentation results for right hands, while (c) and (d) are those for left hands. All successfully segmented images were considered as the experimental data. In the segmented data, the maximum height and width were 353 and 264, respectively, while 94 and 71 were the minimum height and width, respectively.



(**a**) Example of segmentation result for right hand

(**b**) Example of segmentation result for right hand

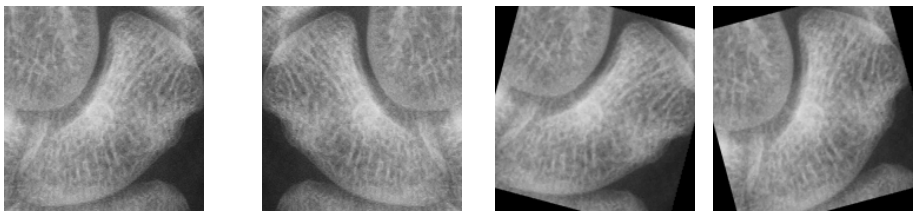
(**c**) Example of segmentation result for left hand

(**d**) Example of segmentation result for left hand

Figure 3. Examples of segmentation results using YOLO-v4.

3.2.3. Data Augmentation

In real applications, it is not easy to collect sufficient data for better recognition quality. This incurs a problem called overfitting, which indicates that a gap exists between the training and prediction performances. To address this issue, data augmentation is widely adopted as a solution, which transforms the images through resizing, flipping, rotating, scaling, and presentation tuning operations, among others. For this concern, in this paper, the training data were augmented by flipping and rotation. Figure 4 shows examples of original, flipped, and rotated images. Finally, the training data size was enlarged to 1136.



(**a**) origin (**b**) flipping horizontally (**c**) rotation by 15° (**d**) rotation by −15°

Figure 4. Examples of data augmentation.

3.2.4. Convolutional Neural Networks

Advanced hardware technology has enabled the rapid growth of artificial neural network technology. They can simulate the human brain to recognize images, as is the case for CNNs. The basic idea is that a CNN performs feature filtering by convolutional computation and pooling. Then, through flattening, full connection, and softmax activation, the prediction probabilities for each class are output. Figure 5 depicts the framework containing convolution sets, maximum pooling sets, flattening, a fully connected layer, and a softmax activation function. Figure 6 illustrates the convolution, which indicates that a convolutional map is generated by sliding dot product computations between an image and a filtering map, while the maximum pooling indicates that a map is generated by maximizing the region features from the convolutional map. In this paper, we compare five NNs: VGG [37], ResNet [38], DenseNet [39], InceptionNet [40], and EfficientNet [41].

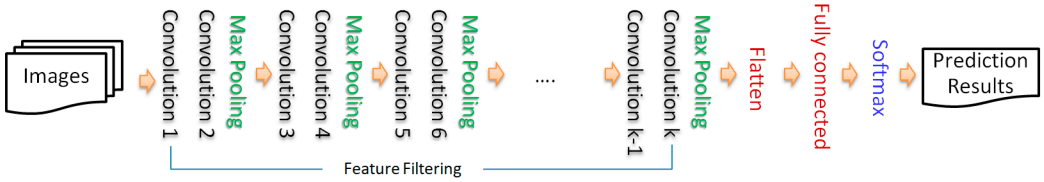


Figure 5. Framework of a general CNN.

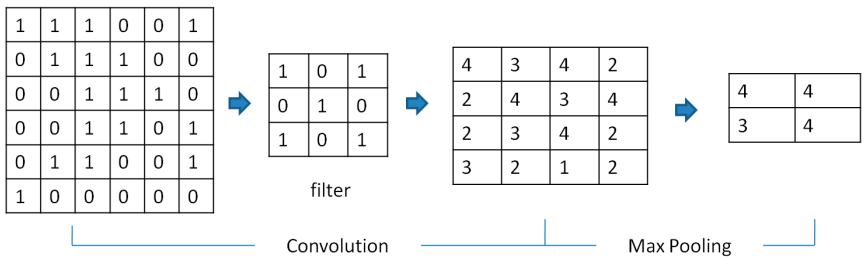


Figure 6. Examples of Convolution and max-pooling.

Table 2 provides the architecture details of the considered CNN models, including the numbers of convolutions and pools, activation functions, and optimization functions. Based on these architectures, our primary intent was to approximate the nearly optimal settings for different CNN models, as depicted in Table 3. The detailed analysis for determining the best settings is shown in Section 4.

Table 2. Architecture details of used CNN models.

	#Convolutions	#Pools	Activation	Optimization Function
VGG16	13	5	ReLu/Softmax	Adam
VGG19	16	5	ReLu/Softmax	Adam
ResNet50	53	1	ReLu/Softmax	Adam
ResNet101	104	1	ReLu/Softmax	Adam
ResNet152	155	1	ReLu/Softmax	Adam
DenseNet121	143	1	ReLu/Softmax	Adam
DenseNet169	168	1	ReLu/Softmax	Adam
DenseNet201	200	1	ReLu/Softmax	Adam
EffB0	81	0	Swish/Softmax	Adam
InceV3	150	3	ReLu/Softmax	Adam

Table 3. Best settings for the considered CNN models.

	Batch Size	Resize	Epochs
VGG16	16	112	100
VGG19	16	168	200
Res50	24	112	120
Res101	32	196	70
Res152	32	140	50
Den121	24	168	100
Den169	32	196	100
Den201	32	196	100
InceV3	8	168	100
EffB0	8	140	100

3.2.5. Transfer Learning

In addition to data augmentation, transfer learning has been proposed to increase the recognition effectiveness when facing insufficient data. The basic notion is to reuse a pre-trained model to enhance the prediction ability. In this paper, two paradigms of transfer learning are utilized, namely layer transfer and fine-tuning. For layer transfer, as shown in Figure 7, the network is initialized by the pretrained model. Next, feature filtering is skipped, and only the prediction layer is trained by the target images. For fine-tuning, as shown in Figure 8, the whole network is initialized by the pretrained model. Then, the feature filtering and prediction layer both keep working with target images. Therefore, the network is tuned based on both the pretrained model and target images. In this paper, the fine-tuning-based transfer learning is performed with VGG networks, while the layer transfer-based learning was performed for the other networks.

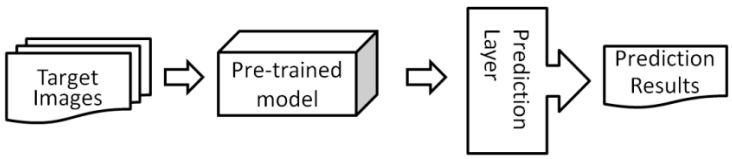


Figure 7. Layer transfer-based transfer learning.

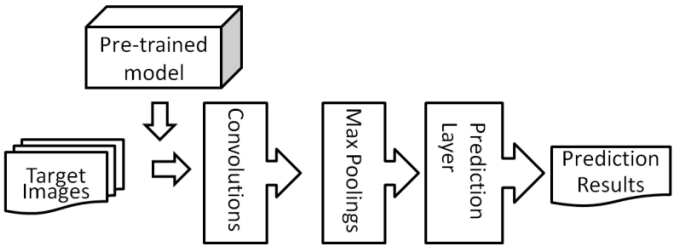


Figure 8. Fine-tuning-based transfer learning.

3.3. Online Recognition

This phase works with an unknown image as input. First, the scaphoid bone is anchored and segmented from the unknown image. Next, the split image is resized. Finally, as shown in Figures 5–8, the split object is predicted through transfer learning from CNN.

4. Empirical Study

To determine the recognition quality for scaphoid fractures, we conducted a detailed empirical study. In the following, it is presented in two main aspects—overall comparison and detailed analysis—for the compared deep learning techniques based on CNNs.

4.1. Experimental Settings

In the experiments, the adopted CNNs included VGG16, VGG19, ResNet50 (RN50), ResNet101 (RN101), ResNet152 (RN152), DenseNet121 (DN121), DenseNet169 (DN169), DenseNet201 (DN201), Inception-V3 (Inv3), and EfficientNetB0 (ENB0). All adopted CNNs were executed using the Keras API and the whole evaluation was conducted on a personal computer with 16 GB RAM, an Intel 8-Core i7-10700 processor, and an Nvidia GeForce RTX 3070 GPU with 8 GB RAM. The recognition quality was measured through seven metrics: sensitivity, specificity, precision, accuracy, F1-score, AUC, and kappa based on a confusion matrix (shown in Table 4). In this matrix, there are four outcomes: true positive (TP), false positive (FP), false negative (FN), and true negative (TN), where positive (P) and negative (N) indicate the numbers of actual positive and negative cases, respectively, in the data. Further, TP and TN indicate the number of hits and number of correct rejections, respectively, while FP and FN indicate the number of false alarms and number of misses, respectively. The definitions for the sensitivity, specificity, precision, accuracy, F1-score, AUC, and kappa metrics are as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \tag{4}$$

$$\text{F1} - \text{score} = \frac{2\text{TP}}{2{*}\text{TP} + \text{FP} + \text{FN}}. \tag{5}$$

$$\text{Kappa} = \frac{2 * (\text{TP} * \text{TN} - \text{FP} * \text{FN})}{(\text{TP} + \text{FP}) * (\text{FP} + \text{TN}) * (\text{TP} + \text{FN}) * (\text{FN} + \text{TN})} \tag{6}$$

Table 4. Confusion matrix.

		Prediction Result	
		Positive	**Negative**
Ground Truth	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

AUC denotes the area under the receiver operating characteristic curve. In the proceeding experiments, the results are shown in terms of accuracy charts and AUC curves, where the charts and AUC curves were generated by the Microsoft Excel and Python (Matplotlib Pyplot) software, respectively. The major intent of using accuracy as the basic measure was to observe the overall correction rates, including true positives and true negatives simultaneously, which has been widely used in the field of machine learning [42,43]. Based on the

accuracy, the other six metrics were further chosen by referring to the references [23–30]. Moreover, the CNN models were constructed using Keras and tensorflow 2.3. The reasons for selecting the software are that they are popular, cheap, and easy to obtain.

4.2. Comparison of the Compared CNNs without Transfer Learning

In the first evaluation, we examined the compared CNNs without transfer learning. In this evaluation, the training data was augmented, the batch size was 32, and the images were resized to 196×196. Table 5 provides the effectiveness of the compared CNNs without transfer learning for different metrics, delivering a set of observations. First, although the specificity and precision of VGG16 were high, the values of other metrics were low. This indicates that, in the negative ground-truths, the correct rate was high, and most positive prediction results are correct. In contrast, the sensitivity of VGG19 was the best, but the related specificity and kappa are low, that is, the VGG nets were not robust enough. Second, in terms of accuracy, the three DenseNets were better than the others. This indicates that, for both positive and negative predictions, the correct rates were high for DenseNets. Third, overall, DN201 obtained the best F1-score, kappa, and accuracy. This reveals that DN201 can be recommended in a situation without transfer learning. In summary, the performances of all CNNs without transfer learning were not satisfactory. The potential reason for this is the overfitting problem, caused by insufficient data. This motivated us to fuse the transfer learning into the prediction model, as mentioned above.

Table 5. Effectiveness of compared CNNs without transfer learning for different metrics.

Model \ Metric	Sensitivity	Specificity	Precision	F1-Score	AUC	Kappa	Accuracy
VGG16	0.083	* 1.000	* 1.000	0.154	0.610	0.083	0.542
VGG19	* 0.972	0.222	0.556	0.707	0.610	0.194	0.597
RN50	0.194	0.861	0.583	0.292	0.580	0.056	0.528
RN101	0.361	0.833	0.684	0.473	0.660	0.194	0.597
RN152	0.444	0.889	0.800	0.571	0.770	0.333	0.667
DN121	0.722	0.722	0.722	0.722	* 0.810	* 0.444	* 0.722
DN169	0.750	0.694	0.711	0.730	0.790	* 0.444	* 0.722
DN201	0.833	0.611	0.682	* 0.750	0.790	* 0.444	* 0.722
INv3	0.694	0.500	0.581	0.633	0.610	0.194	0.597
ENB0	0.694	0.556	0.610	0.633	0.630	0.250	0.625

Note that * indicates the best performance in each metric (attribute).

4.3. Performance Analysis in Different Settings for the Compared CNNs with Transfer Learning

As the evaluation results without transfer learning were unsatisfactory, from the viewpoint of all metrics, the further models with transfer learning were examined. Before the further examinations, the settings for the further models with transfer learning had to be set. For this purpose, in this section, the performance analysis under different settings is demonstrated, through several evaluations.

4.3.1. Impact of Data Augmentation

The first result to assess, in terms of the problem of insufficient data, is the impact of data augmentation. Figure 9 shows the effectiveness of different models with different data augmentations. In these results, different data augmentation types had different impacts for each model. On average, the performance rank for three augmentation types were rotation 15°, rotation −15°, and flipping. Overall, for any of the individual augmentations, the performance was better than that without augmentation (denoted as "original"). Besides the individual augmentations, further comparison between those with and without augmentations was made. Figure 10 reveals the results in a manner similar to that of Figure 9. Three viewpoints for the results are shown here. First, the accuracies of VGG Nets and EfficientNets did not improve significantly. Second, the best improvement was

for RN101, which reached an improvement rate of approximately 73%. Third, the potential interpretation for the difference in effectiveness of augmentation on different models is that the model architectures were sensitive to data augmentation, considering the pre-trained model. In detail, the residual idea works well for ResNets and DenseNets. This is also the main contribution to discerning the difference among AI Nets. Fourth, on the whole, most models were enhanced through data augmentation. Note that this evaluation was carried out based on the settings presented in Table 3.

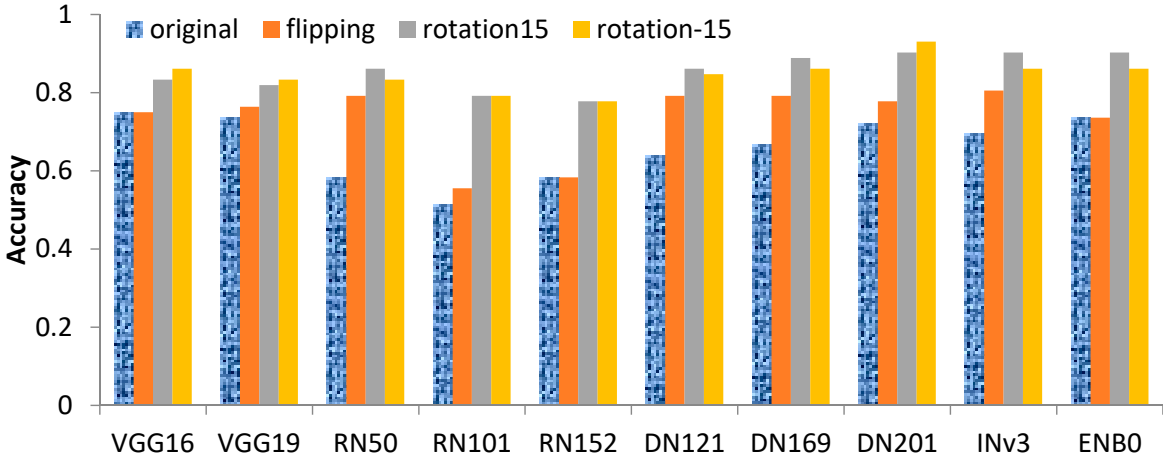


Figure 9. Accuracies of different models with different data augmentations.

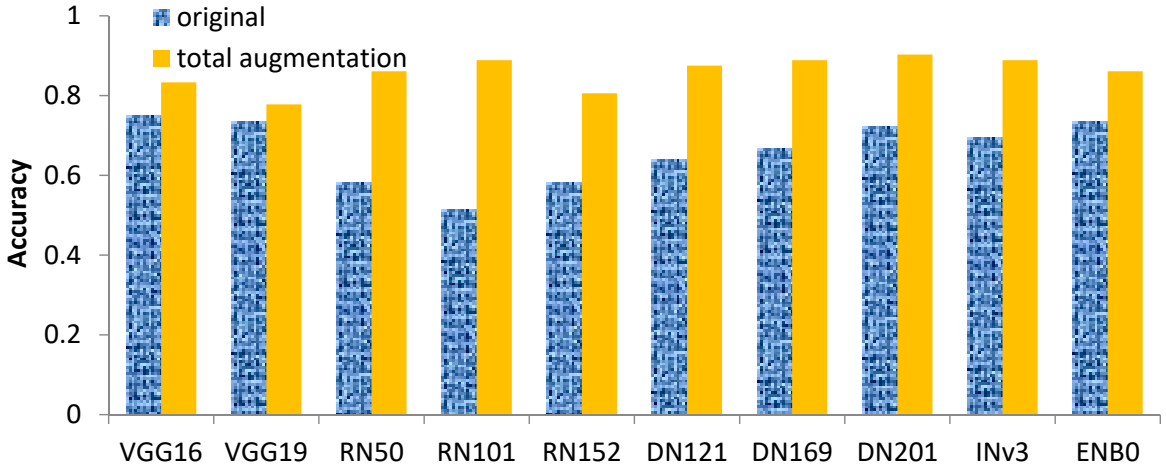


Figure 10. Accuracies of different models with or without data augmentations.

4.3.2. Impact of Batch Size

The next concern to clarify is the impact of different batch sizes. For this, we conducted a further evaluation of all models, using different batch sizes. Table 6 shows the experimental results for different batch sizes, which can be interpreted from two aspects. First, for each model, different models performed best with different batch sizes. Some differences between the best and worst are obvious, while the others are not. For example, for RN152, the accuracy difference was the biggest, while that for INv3 was the smallest, that is, the impact of batch size on RN152 was significant, which reached 0.292, while that for INv3 was not clear. Overall, the average difference of all models was 0.114. Second, considering the batch size, the accuracy of INv3 can reach 0.861 only when using a batch size of 8. This indicates that INv3 does not need much cost to perform well. Additionally, the batch size ranking, in terms of average accuracies, was 8, 24, 16, and 32. However, the best batch size differed for each model. Note that this evaluation was made using the resize to 196 $\times$ 196.

Table 6. Accuracies of compared CNNs for different batch sizes.

Model \ Batch Size	8	16	24	32
VGG16	0.792	0.819	0.722	0.806
VGG19	0.792	0.819	0.792	0.764
RN50	0.764	0.792	0.806	0.792
RN101	0.694	0.750	0.764	0.889
RN152	0.500	0.681	0.681	0.792
DN121	0.806	0.833	0.889	0.847
DN169	0.722	0.847	0.833	0.889
DN201	0.764	0.875	0.875	0.903
INv3	0.861	0.833	0.847	0.847
ENB0	0.847	0.847	0.833	0.806

4.3.3. Impact of Resizing

The final setting for the overall comparisons was the parameter resize. This parameter indicates the image size formatted normally for all models. This was used as the input image sizes were not equal. Table 7 depicts the evaluation results for different resizes. From the results, we can see that the differences between the best and worst results for each model were not clear. It can be interpreted that the impact of image size is not obvious, where the maximum difference was just 0.139. Although the best performance occurred when resizing to 196 × 196, the best setting for each model differed. Note that this evaluation was made using a batch size of 32.

Table 7. Accuracies of compared CNNs for different image resizes.

Model \ Resize	112 × 112	140 × 140	168 × 168	196 × 196
VGG16	0.819	0.778	0.778	0.806
VGG19	0.708	0.806	0.847	0.764
RN50	0.861	0.806	0.792	0.792
RN101	0.861	0.764	0.778	0.889
RN152	0.778	0.833	0.764	0.792
DN121	0.861	0.833	0.903	0.847
DN169	0.833	0.861	0.833	0.889
DN201	0.847	0.806	0.875	0.903
INv3	0.875	0.847	0.875	0.847
ENB0	0.819	0.833	0.819	0.806

4.4. Overall Comparisons for the Compared CNNs with Transfer Learning

In the above experiments, the best settings for data augmentation, batch size, and resizing were determined. In this subsection, based on these settings, a comparative study is detailed, with evaluations of all models using transfer learning. Table 8 shows the effectiveness of the compared CNNs with transfer learning for different metrics, which can be explained from numerous viewpoints. First, the ranking of the five Nets was DenseNet, InceptionNet, EfficientNet, ResNet, and VGG, with average accuracies of 0.889, 0.889, 0.861, 0.852, and 0.806, respectively. Second, the top three individual models were DN201, DN169, and RN101, in accordance with their accuracies. Third, overall, the best performances for each metric consistently occurred in two networks, namely DN201 and RN101. From this aspect, DN201 and RN101 were the two most reliable models. Fourth, from an average point of view, the top three individual models were DN201, RN101, and INv3, for which the averages of all metric results were 0.886, 0.882, and 0.879, respectively. This can be regarded as an echo of the third point that DN201 and RN101 had a balanced performance for all metrics.

Table 8. Effectiveness of compared CNNs with transfer learning for different metrics.

Model \ Metric	Sensitivity	Specificity	Precision	F1-Score	AUC	Kappa	Accuracy
VGG16	0.861	0.806	0.816	0.838	0.860	0.667	0.833
VGG19	0.833	0.722	0.750	0.789	0.870	0.556	0.778
RN50	0.889	0.833	0.842	0.865	0.910	0.722	0.861
RN101	0.889	* 0.889	* 0.889	0.889	* 0.950	0.778	0.889
RN152	0.806	0.806	0.806	0.806	0.880	0.611	0.806
DN121	0.917	0.833	0.846	0.880	0.930	0.750	0.875
DN169	0.917	0.861	0.868	0.892	0.890	0.778	0.889
DN201	* 0.944	0.861	0.872	* 0.907	0.910	* 0.806	* 0.903
INv3	0.889	* 0.889	* 0.889	0.889	0.930	0.778	0.889
ENB0	* 0.944	0.778	0.810	0.872	0.920	0.722	0.861

Note that * indicates the best performance in each metric (attribute).

4.5. Experimental Discussion

To realize the recognition quality of the well-known CNNs, in this paper, a comprehensive evaluation was proposed, as detailed above. In the following, an overall empirical discussion is further given, for detailed analysis.

- The medical data in real applications is not easy to collect. To make the experiment more reliable, the experimental data were gathered from Kaohsiung Chang Gung Memorial Hospital, instead of crawling the Web. Therefore, due to the problem of insufficient data, overfitting occurred and heavily impacted the recognition quality. To cope with this problem, two operations were adopted: data augmentation and transfer learning. Here, an issue to clarify is the performance comparisons among learning by original data, learning by data augmentation, learning by transfer learning, and learning by fusing data augmentation and transfer learning. This comparison can be explained by Figure 11, summarizing Tables 5 and 8 and Figure 10. From this comparison, we can determine that, whatever the model, without the fusion of transfer learning and data augmentation, the best accuracy cannot be achieved.
- In addition to the four observations mentioned in Section 4.4, another point to show is the performance comparison in terms of the AUC. Figure 12 reveals the AUC comparisons of models with accuracies larger than 0.9, namely RN50, RN101, DN121, DN201, INv3, and ENB0. Basically, it can be divided into two observation spaces, by the line where the false positive rate (FPR) equals 0.1. Before FPR 0.1, DN121 is better than the others. To the contrary, the performances of RN101 and DN201 are relatively close but higher than those of the other models, after FPR 0.1. Overall, RN101, DN201, and INv3 are the highly considerable models, in terms of the AUC.
- Besides the AUC, a further comparison to show the interdiscrimination ability of TP, TN, FP, and FN for each model is the hybrid validation, which indicates the average of the F1-score, AUC, and kappa metrics. Figure 13 depicts the validation result, showing that the top three models were DN201, DN169, and RN101, although their values were close. This result is consistent with those given above. However, by considering the overall performances in Table 8, the best two models for recognizing scaphoid fractures are DN201 and RN101.
- In addition to the effectiveness evaluation, another important issue is the efficiency. Considering this issue, Figure 14 shows the execution time of compared models. The comparison result revealed that DenseNet, InceptionNet, and EfficientNet had higher costs than the other models. By integrating the results of Figures 13 and 14, three further viewpoints are given here. First, from the effectiveness point of view, DN201, DN169, and RN101 are three candidate models. Second, from the efficiency point of view, VGG16, RN50, and RN152 are three candidate models. Third, from a balanced viewpoint, the top three models are VGG16, RN50, and RN152 because the referred balances between test performance and training cost are higher than that of the other

models. Overall, DN201 and RN101 are the considerable recommendations, due to three aspects. First, the training operation would not be performed frequently. Second, the testing quality is more important than the training cost in the field of bioinformatics. Third, in the experiments, the recognition for each model could be carried out within 1 s. Hence, the testing quality was our major consideration in determining the recommendations.

- In the above experiments, the evaluation results using the KCGMH data were shown in great detail. However, to make the experiments more robust, we further verified the above models using another data set named RSNA [44], which was proposed for a challenge to predict bone age from pediatric hand radiograph. The major intent behind this verification is to investigate the performance variances of above recognition models if using different data sources. In fact, the data set RSNA contains bones of different ages, and 538 images were selected as the testing data. Because the bones in this data set are all normal, the specificity (TN rate) is the aimed measure. Figure 15 shows the specificity comparisons of using RSNA data and KCGMH data for the recognition models. The specificity differences are small (within 5%), even though the data sets are generated by different radiograph capturing devices, that is, the constructed models are stable in detecting the scaphoid fractures.
- In summary, the above experimental results provide evidence that DenseNet and ResNet perform better than the other three networks. From these two networks, DN201 and RN101 are further selected as the recommended recognizers because the overall performances are better than the others in these two networks.

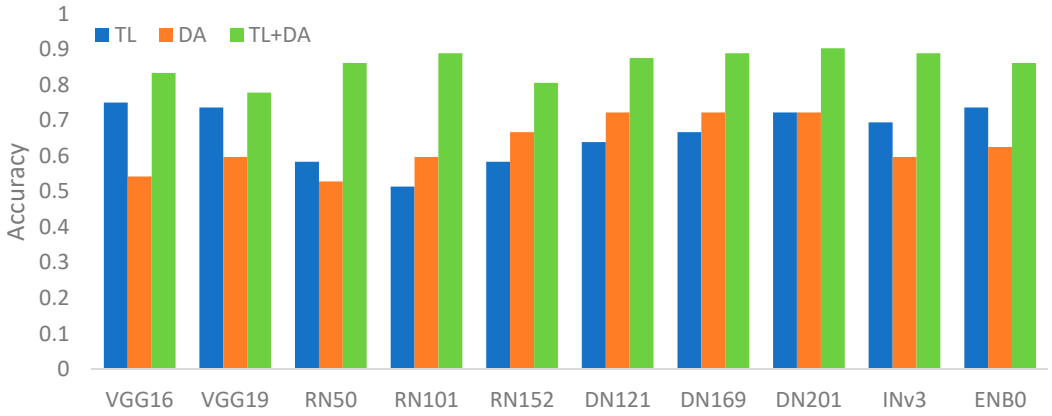


Figure 11. Accuracies of compared CNNs by considering transfer learning (termed TL), data augmentation (termed DA), and fusion of TL and DA (termed TL+DA).

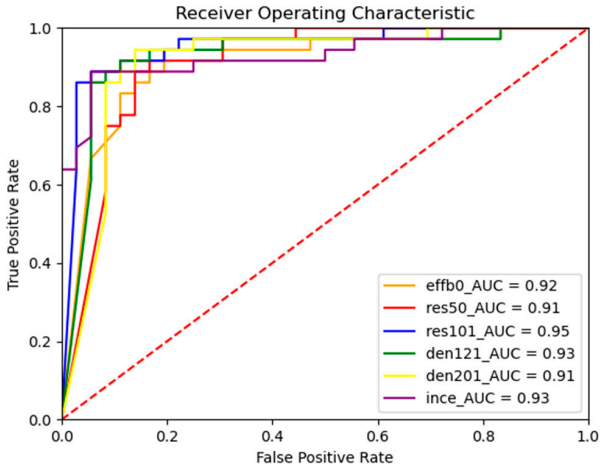


Figure 12. AUCs of models with accuracies larger than 0.9.

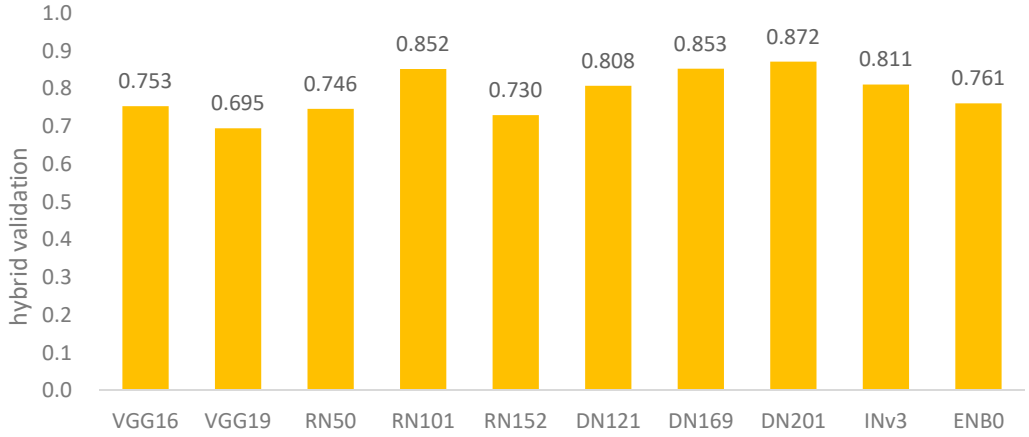


Figure 13. Hybrid evaluations by averaging F1-score, AUC, and kappa metrics.

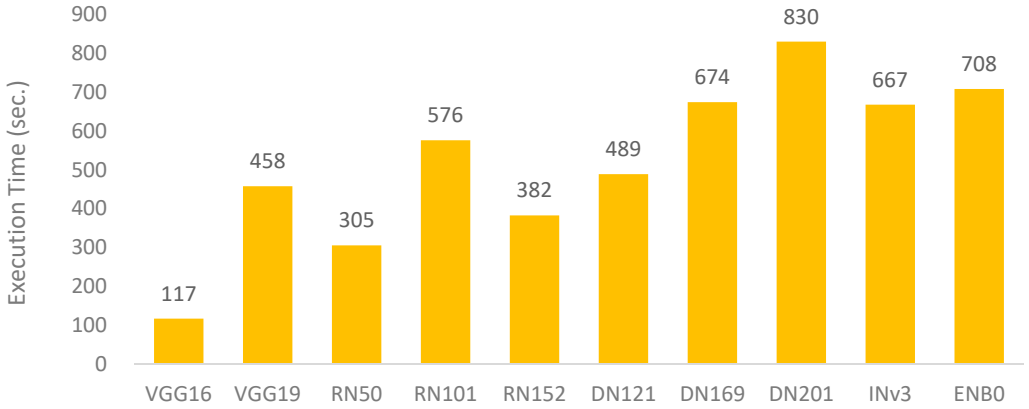


Figure 14. Training time of the recognition models.

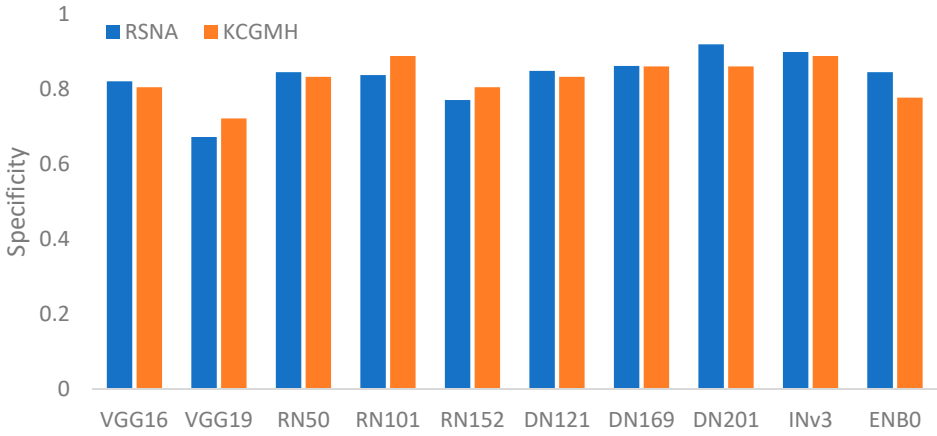


Figure 15. Specificity comparisons of using RSNA data and KCGMH data for the recognition models.

5. Research Limitations

In this paper, we detail how we studied numerous CNN models as an important determinant of scaphoid fracture recognition. However, there are several limitations in this study that need to be declared here. First, the recognition models are constructed by the same data source. Second, the data type is limited in the adult type because this paper is just a beginning of scaphoid fracture recognition. More data types will be tested in the future. Third, the data size is not large because the real data are not easy to gather. Fourth, the data came only from patients in Kaohsiung Chang Gung Memorial Hospital, Taiwan. Fifth, the models of radiograph capturing devices are the same. Sixth, the resolution

qualities are close because the images are generated from the same radiograph-capturing devices. More related suggestions for future research are listed in the proceeding section.

6. Conclusions and Future Works

Over the past few years, deep learning methods have been successfully used in the field of bioimage recognition. However, few studies have focused on the detection of scaphoid fractures. In this paper, the related works are reviewed and compared comprehensively. From the review, the lack of related works inspired us to propose a two-stage method based on state-of-the-art CNNs. In the first stage, the scaphoid bone is segmented from the target image. Next, the segmentation is augmented by geometric operations. Then, the recognition model is constructed using transfer learning from pretrained CNNs. Finally, a set of CNNs were evaluated, in terms of seven metrics, and a detailed analysis was provided. In the following, the main contributions of our work are discussed.

- In this paper, one of our major intentions was to narrow the recognition focus to the scaphoid bone, instead of the whole image. For this, the scaphoid bone was viewed as an object to be extracted from the target image. Therefore, the prediction cost can be reduced, and the effectiveness can be significantly increased.
- Based on the segmentation, several CNNs were employed for solution of the problem. Data augmentation and transfer learning were fused to enhance the prediction quality. The experimental results revealed that the enhancement due to this fusion could tackle the overfitting problem.
- For insight into the recognition results obtained for the models used, in this paper, numerous detailed evaluations were conducted. Through the overall analysis, insightful and comprehensive perspectives were obtained. Finally, based on the empirical study, the best two models could be recommended, namely DenseNet201 and ResNet101.

Although we presented a set of effective enhancements for the detection of scaphoid fractures, the following issues remain to be addressed in the future:

- First, with more data, the effectiveness can be increased. However, scaphoid fracture data are not easy to collect. For this purpose, in the future, we will attempt to gather more data from the other branches of Chang Gung Memorial Hospitals in Taiwan, including from pediatric patients, that is, the pediatric images will be added to enhance the coverage.
- Second, in this paper, the adopted models operated individually. In the future, an ensemble model will be constructed by optimizing the combinations. Moreover, a progressive learning model will also be tested in order to increase the prediction quality.
- Third, in addition to radiographs, CT images will be used as test data, utilizing the ideas proposed in this paper.

Finally, several suggestions about future related studies with similar focus are listed:

- Other bioimage recognition methods have also been studied by the authors' team, such as those for the detection of liver and lung tumors. It has been suggested that, regardless of the method, segmentation is useful in decreasing the computational cost and increasing the accuracy.
- In real applications, a huge imbalance between positives and negatives exists, especially for cancers and noncancers. Consequently, it has been suggested that data augmentation and transfer learning are necessary.
- For fracture recognition, although the experimental results indicated that most scaphoid fractures can be recognized, other symptoms, such as hydrops, joint effusion, nonunion, delayed union, avascular necrosis, arthritis, and so on also need attention, as automated recognition of those symptoms is helpful to doctors in determining the required treatments.

Author Contributions: Conceptualization, Y.-C.T., J.-H.S., Y.-W.L., C.-D.C. and Y.-F.C.; data curation, Y.-C.T., W.-C.C., C.-D.C. and Y.-F.C.; methodology, J.-H.S. and Y.-W.L.; validation, B.-H.C.; writing—original draft, J.-H.S.; writing—review and editing, J.-H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Ministry of Science and Technology and Chang Gung Memorial Hospital, Taiwan, R.O.C. under grant no. MOST 109-2511-H-230-001-MY2, MOST 110-2221-E-390-015 and CORPG8L0251, respectively.

Institutional Review Board Statement: The data were approved by Kaohsiung Chang Gung Memorial Hospital, Taiwan, and all operations in this paper were executed according to the ethical standards of the Institutional Review Board, Taiwan.

Conflicts of Interest: The authors would like to declare that no conflicts of interest exist in this paper. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Roshani, M.; Sattari, M.A.; Ali, P.J.M.; Roshani, G.H.; Nazemi, B.; Corniani, E.; Nazemi, E. Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Flow Meas. Instrum.* **2020**, *75*, 101804. [CrossRef]
2. Arab, F.; Karimi, M.; Safavi, S.M. Analysis of QoS parameters for video traffic in homeplug AV standard using NS-3. In Proceedings of the 2016 Smart Grids Conference, Kerman, Iran, 20–21 December 2016; pp. 1–6.
3. Fathabadi, F.R.; Molavi, A. Black-box identification and validation of an induction motor in an experimental application. *Eur. J. Electr. Eng.* **2019**, *21*, 255–263. [CrossRef]
4. Tavakoli, S.; Yooseph, S. Algorithms for inferring multiple microbial networks. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine, San Diego, CA, USA, 18–21 November 2019; pp. 223–227.
5. Nisar, M.U.; Voghoei, S.; Ramaswamy, L. Caching for pattern matching queries in time evolving graphs: Challenges and approaches. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing System, Atlanta, GA, USA, 5–8 June 2017; pp. 2352–2357.
6. Bahramian, F.; Akbari, A.; Nabavi, M.; Esfandi, S.; Naeiji, E.; Issakhov, A. Design and tri-objective optimization of an energy plant integrated with near-zero energy building including energy storage: An application of dynamic simulation. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101419.
7. Li, Z.; Lin, Y.; Elofsson, A.; Yao, Y. Protein contact map prediction based on ResNet and DenseNet. *BioMed Res. Int.* **2020**, *2020*, 7584968. [CrossRef] [PubMed]
8. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 1–48. [CrossRef]
9. Tong, Y.; Lu, W.; Yu, Y.; Shen, Y. Application of machine learning in ophthalmic imaging modalities. *Eye Vis.* **2020**, *7*, 1–15. [CrossRef] [PubMed]
10. Khraief, C.; Benzarti, F.; Amiri, H. Convolutional neural network based on dynamic motion and shape variations for elderly fall detection. *Int. J. Mach. Learn. Comput.* **2019**, *9*, 814–820. [CrossRef]
11. Lenchik, L.; Heacock, L.; Weaver, A.A.; Boutin, R.D.; Cook, T.S.; Itri, J.; Filippi, C.G.; Gullapalli, R.P.; Lee, J.; Zagurovskaya, M.; et al. Automated segmentation of tissues using CT and MRI: A systematic review. *Acad. Radiol.* **2019**, *26*, 1695–1706. [CrossRef]
12. De Haan, K.; Rivenson, Y.; Wu, Y.; Ozcan, A. Deep-learning-based image reconstruction and enhancement in optical microscopy. *Proc. IEEE* **2019**, *108*, 30–50. [CrossRef]
13. Ghoneim, S. Accuracy, Recall, Precision, F-Score & Specificity, Which to Optimize on? Based on Your Project, Which Performance Metric to Improve on? Towards Data Science. Available online: https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124 (accessed on 23 July 2021).
14. Tang, Y.X.; Tang, Y.B.; Peng, Y.; Yan, K.; Bagheri, M.; Redd, B.A.; Brandon, C.J.; Lu, Z.; Han, M.; Xiao, J.; et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit. Med.* **2020**, *3*, 1–8. [CrossRef]
15. Xia, K.; Yin, H.; Qian, P.; Jiang, Y.; Wang, S. Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images. *IEEE Access* **2019**, *7*, 96349–96358. [CrossRef]
16. Francis, N.S.; Francis, N.J.; Xu, Y.; Saqib, M.; Aljasar, S.A. Identify Cancer in Affected Bronchopulmonary Lung Segments Using Gated-SCNN Modelled with RPN. In Proceedings of the 2020 IEEE 6th International Conference on Control Science and Systems Engineering, Beijing, China, 17–19 July 2020; pp. 5–9.
17. Javan, N.A.; Jebreili, A.; Mozafari, B.; Hosseinioun, M. Classification and Segmentation of Pulmonary Lesions in CT images using a combined VGG-XGBoost method, and an integrated Fuzzy Clustering-Level Set technique. *arXiv* **2021**, arXiv:2101.00948.
18. He, X.; Chen, K.; Hu, K.; Chen, Z.; Li, X.; Gao, X. HMOE-Net: Hybrid Multi-scale Object Equalization Network for Intracerebral Hemorrhage Segmentation in CT Images. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine, Seoul, Korea, 16–19 December 2020; pp. 1006–1009.

19. Li, L.; Wei, M.; Liu, B.; Atchaneeyasakul, K.; Zhou, F.; Pan, Z.; Kumar, S.A.; Zhang, J.Y.; Pu, Y.; Liebeskind, D.S.; et al. Deep learning for hemorrhagic lesion detection and segmentation on brain ct images. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 1646–1659. [CrossRef] [PubMed]
20. Wang, Z.; Wu, L.; Ji, X. An Interpretable Deep Learning System for Automatic Intracranial Hemorrhage Diagnosis with CT Image. In Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing, Harbin, China, 22–24 January 2021; pp. 338–357.
21. Yang, Y.; Jiang, H.; Sun, Q. A multiorgan segmentation model for CT volumes via full convolution-deconvolution network. *BioMed Res. Int.* **2017**, *2017*, 6941306. [CrossRef] [PubMed]
22. Zhou, Z.; Zhang, B.; Yu, X. Infrared Handprint Classification Using Deep Convolution Neural Network. *Neural Process. Lett.* **2021**, *53*, 1065–1079. [CrossRef]
23. Sato, G.T.S.; Da Silva Segundo, L.B.; Dias, Z. Classification of Musculoskeletal Abnormalities with Convolutional Neural Networks. In *Advances in Bioinformatics and Computational Biology. BSB 2020. Lecture Notes in Computer Science*; Setubal, J.C., Silva, W.M., Eds.; Springer: Cham, Switzerland, 2020; Volume 12558, pp. 69–80. [CrossRef]
24. Huang, L.; Xia, W.; Zhang, B.; Qiu, B.; Gao, X. MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images. *Comput. Methods Programs Biomed.* **2017**, *143*, 67–74. [CrossRef]
25. Rayan, J.C.; Reddy, N.; Kan, J.H.; Zhang, W.; Annapragada, A. Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. *Radiol. Artif. Intell.* **2019**, *1*, e180015. [CrossRef] [PubMed]
26. England, J.R.; Gross, J.S.; White, E.A.; Patel, D.B.; England, J.T.; Cheng, P.M. Detection of traumatic pediatric elbow joint effusion using a deep convolutional neural network. *Am. J. Roentgenol.* **2018**, *211*, 1361–1368. [CrossRef]
27. Olczak, J.; Fahlberg, N.; Maki, A.; Razavian, A.S.; Jilert, A.; Stark, A.; Sköldenberg, O.; Gordon, M. Artificial intelligence for analyzing orthopedic trauma radiographs: Deep learning algorithms—are they on par with humans for diagnosing fractures? *Acta Orthop.* **2017**, *88*, 581–586. [CrossRef] [PubMed]
28. Kim, D.H.; MacKinnon, T. Artificial intelligence in fracture detection: Transfer learning from deep convolutional neural networks. *Clin. Radiol.* **2018**, *73*, 439–445. [CrossRef]
29. Nazarova, G.; Sinitsyn, V.; Mershina, E.A.; Filippov, V.V. Special x-ray projections in assessment of scaphoid bone fractures. *Eur. Congr. Radiol.* **2019**, *80*, 82–90.
30. Edlund, R.; Skorpil, M.; Lapidus, G.; Bäcklund, J. Cone-beam CT in diagnosis of scaphoid fractures. *Skelet. Radiol.* **2016**, *45*, 197–204. [CrossRef] [PubMed]
31. Ozkaya, E.; Topal, F.E.; Bulut, T.; Gursoy, M.; Ozuysal, M.; Karakaya, Z. Evaluation of an artificial intelligence system for diagnosing scaphoid fracture on direct radiography. *Eur. J. Trauma Emerg. Surg.* **2020**. [CrossRef] [PubMed]
32. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
33. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
34. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
36. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
38. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
39. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
40. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
41. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
42. Su, J.-H.; Chou, C.-L.; Lin, C.-Y.; Tseng, V.S. Effective Semantic Annotation by Image-to-Concept Distribution Model. *IEEE Trans. Multimed.* **2011**, *13*, 530–538. [CrossRef]
43. Su, J.-H.; Huang, W.-J.; Yu, P.S.; Tseng, V.S. Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 360–372. [CrossRef]
44. RSNA Pediatric Bone Age Challenge (2017). Available online: https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pediatric-bone-age-challenge-2017 (accessed on 9 September 2021).