Computers and Electrical Engineering 81 (2020) 106530

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Arm fracture detection in X-rays based on improved deep convolutional neural network ☆

Bin Guan [a], Guoshan Zhang [a,*], Jinkun Yao [b], Xinbo Wang [a], Mengxuan Wang [a]

[a] School of Electrical and Information Engineering, Tianjin University, 300072 Tianjin, China
[b] Department of Radiology, Linyi People's Hospital, 276000 Linyi, China

ARTICLE INFO

Article history:
Received 29 November 2018
Revised 4 October 2019
Accepted 26 November 2019
Available online 2 December 2019

Keywords:
Deep learning
Convolutional neural network
Arm fracture detection
Medical image processing
Computer aided detection
X-ray

ABSTRACT

In this paper, a novel deep learning method is proposed and applied to fracture detection in arm bone X-rays. The main improvements include three aspects. First, a new backbone network is established based on feature pyramid architecture to gain more fractural information. Second, an image preprocessing procedure including opening operation and pixel value transformation is developed to enhance the contrast of original images. Third, the receptive field adjustment containing anchor scale reduction and tiny RoIs expansion is exploited to find more fractures. In the experiments, nearly 4000 arm fracture X-ray radiographs collected from MURA dataset are annotated by experienced radiologists. The experiment results show that the proposed deep learning method achieves the state-of-the-art AP in arm fracture detection and it has strong potential application in real clinical environments.

1. Introduction

Musculoskeletal conditions affect more than 1.7 billion people worldwide, and these conditions are the most common cause of severe, long-term pain and disability [1]. The developments of CT scanning technologies extremely improve the diagnosis and treatment of musculoskeletal conditions. However, compared with a great deal of patients, the number of excellent radiologists is too scarce. The workload of the radiologists is beyond their capability. Thus, they are overwhelmed with huge amount of medical image data. As a result, assistive technologies are urgently needed for radiologists in this field.

To reduce the workload of radiologists and help them make more accurate diagnosis, Computer Aided Detection and Diagnosis (CAD) are developed and widely employed as a second reader. It has been pointed out that CAD attempts to achieve the following four goals: improve radiologists' performance, save time, be seamlessly integrated into workflow and have negligible incremental costs [2]. For bone fracture detection in X-ray images, some methods have been proposed. Cao et al. [3] presented a generalized bone fracture detection method called Stacked Random Forests Feature Fusion based on a discriminative learning framework in 2015. This method outperforms other fracture detection frameworks that use local features, single layer random forests, and support vector machine (SVM) classification. Bandyopadhyay et al. [4] proposed digital-geometric techniques to detect long-bone fracture locations and types in 2016. However, the detection accuracy of these methods is far away from the clinical application.

☆ This paper is for CAEE special section SI-mip. Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Li He.

* Corresponding author.
E-mail address: zhanggs@tju.edu.cn (G. Zhang).

https://doi.org/10.1016/j.compeleceng.2019.106530
0045-7906/

With the advent of deep learning methods, CAD algorithms have shown promise to help radiologists in clinical environment. So far, the deep learning method has been applied to the pulmonary nodule detection [5] and wrist fractures detection [6] to name a few. Currently, as for application in bone fracture detection. Kim et al. [7] re-trained the Inception-v3 network by using of lateral wrist radiographs to produce a model which can determine if a new case is fractural. Raghavendra et al. [8] developed a novel CNN classification model to automatically diagnose the thoracolumbar fracture. Combining a deep residual network (ResNet) with long short-term memory (LSTM) network, Tomita et al. [9] studied automatic classification of osteoporotic vertebral fractures. Rajpurkar et al. employed MURA [10], a large dataset of musculoskeletal radiographs containing 40,895 radiographs, to train a simple binary classification model by using of a 169-layer DenseNet. Ebsim et al. [11] trained a CNN to detect wrist fractures from posteroanterior and lateral radiographs. England et al. [12] used deep CNN to detect traumatic pediatric elbow joint effusion. Urakawa et al. detect intertrochanteric hip fractures by using of VGG-16 to analyze whether the proximal femurs cropped from an anterior-view hip radiograph is fractured or non-fractured, and they reported that their fine-tuned model achieves accuracy of 95.5% [13]. Badgeley et al. employ Inception-v3 to predict hip fracture using confounding patient and healthcare variables [14]. Adams et al. use AlexNet and GoogLeNet to detect femoral neck fractures in X-ray and get the highest accuracy of 94.4% [15]. However, to the best of our knowledge, most of mentioned studies focus on classification task rather than real detection task, i.e., these models can diagnose whether a radiograph is fractured or non-fractured, but they cannot show exact locations of the fracture by predicting bounding-boxes.

In fact, real object detection is a complex problem, and it deals with two main tasks. First, the detector is required to solve the recognition problem, i.e., it needs to distinguish foreground objects from background and assign them the proper object class labels. Second, the detector needs to solve the localization problem, i.e., it needs to assign accurate bounding boxes for different objects. The following methods with localization capabilities may be more helpful and valuable to radiologists. Pranata et al. combine a ResNet-50 with speeded-up robust features (SURF) method to demonstrate the feasibility for computer-aided classification and detection of the location of calcaneus fractures in CT images [16]. Gan et al. use Faster R-CNN and Inception-v4 to detect distal radius fractures, and they reported that the proposed network presented a similar diagnostic performance to that of the orthopedists [17]. Lindsey et al. develop an extension of the U-Net architecture to detect and localize wrist fractures in radiographs, a controlled experiment shows that the assistance of the deep learning model significantly improve the diagnostic accuracy of emergency medicine clinicians [6]. Guan et al. [18] design a new convolutional neural network using dilated convolutions to detect and localize the thighbone fractures in X-ray and achieve the state-of-the-art average precision (AP) of 82.1%.

In fact, the above-mentioned deep learning methods for fracture detection and localization are inspired by the generic object detection algorithm. Recently, generic object detection is popularized by both two-stage and single-stage detectors. Two-stage detectors were first introduced by R-CNN, Gradually derived Fast R-CNN [19] and Faster R-CNN [20] promoted the developments furthermore. Faster R-CNN proposed a region proposal network to improve the efficiency of detectors and allow the detectors to be trained end-to-end. After this meaningful milestone, lots of methods were introduced to enhance Faster R-CNN from different points. For example, FPN [21] alleviated the scale variance via architecture of multi-scale feature pyramid. Cascade R-CNN [22] extended Faster R-CNN to a multi-stage detector. So far, the two-stage detectors have been recognized as the best-performing generic object detection method, and they have been widely utilized in many fields.

However, the application on arm bone fracture detection in radiographs has not been found. At present, arm bone fractures in X-rays are still need to be manually examined by radiologists. Hence, it is necessary to develop a deep learning-based arm fracture detection method. So, in this paper, we propose a novel deep learning framework for arm fracture detection.

The main contributions of the proposed method include following three aspects. First, a new backbone network is proposed based on feature pyramid architecture. Second, an image preprocessing procedure including opening operation and pixel value transformation is developed to enhance the contrast of original images. Third, the receptive filed adjustment containing anchor scale reduction and tiny RoIs expansion is exploited. In the experiments, 3392 X-ray radiographs of arm bone fracture are collected from the MURA dataset [10] to train the algorithms, and the other 612 radiographs are employed as test set to evaluate our algorithm. The bounding-boxes of these arm fracture X-ray radiographs are annotated by invited experienced radiologists. The experiment results show that the proposed method achieves the state-of-the-art detection performance in arm fracture detection.

2. Methodology

2.1. Overview

The proposed method is an improved two-stage R-CNN method, and the overview of the proposed method is shown in Fig. 1. First, all images are preprocessed by opening operation and pixel transformation. Second, inspiring by feature pyramid architecture, a novel backbone network is designed to extracts features from preprocessed image. Third, feature maps {S2; S3; S4; S5; S6} with 5 different scales are fed into Region Proposal Network [21], which provide object proposals at each pixel position. Fourth, the RPN is trained to tell the following Fast R-CNN classifier where to detect by generating 256 region of interests (RoIs). Hereafter, the receptive field expansion is exploited to expand the tiny RoIs for the detection of tiny fractures. Fifth, the RoI pooling layer [20] unifies the size of the cropped features in RoIs into a small feature map with a fixed spatial extent of 7×7. Sixth, the feature map with fixed 7×7 spatial extent is flattened to a feature vector,

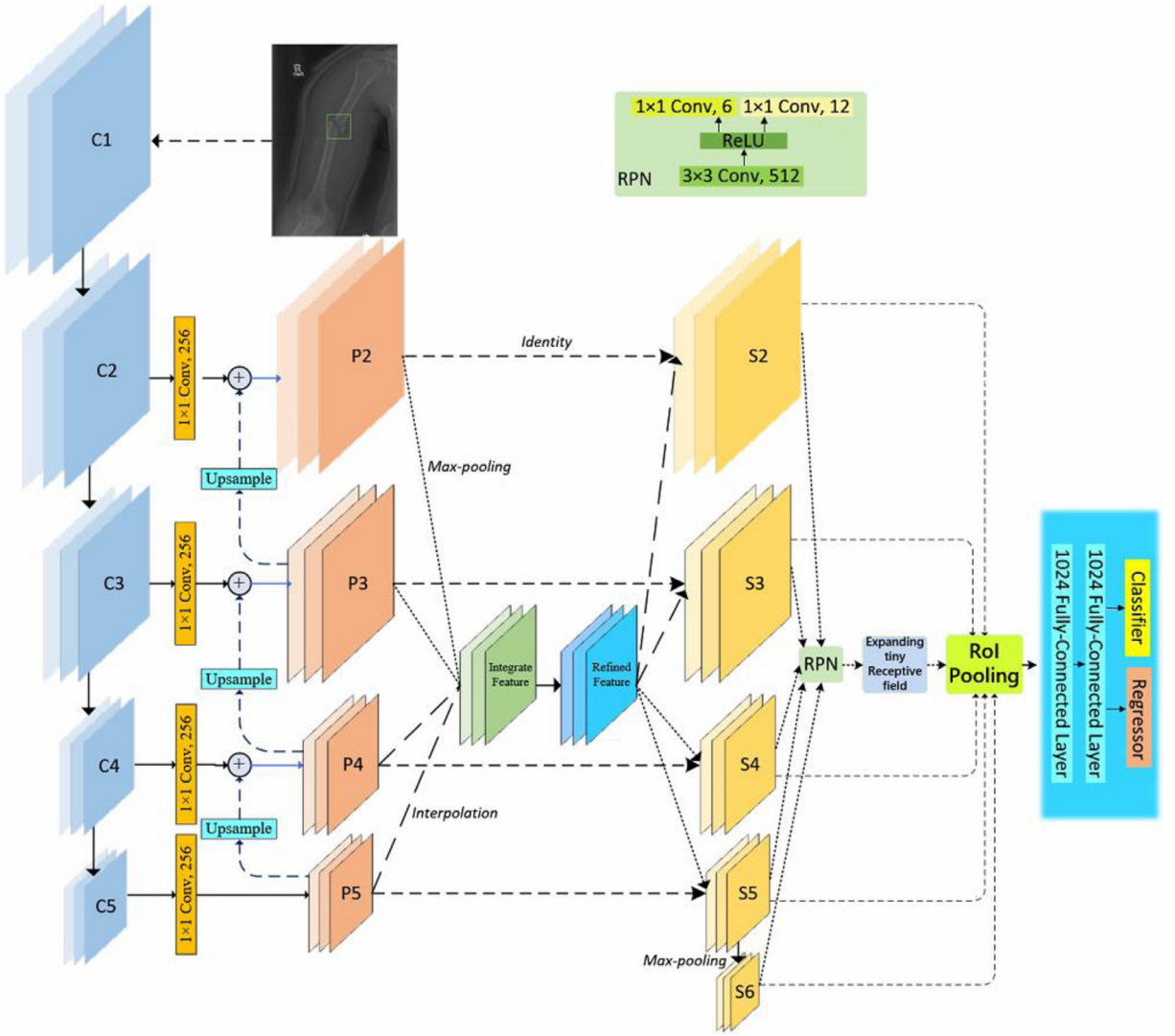


Fig. 1. Outline of the proposed method for arm fracture detection.

which is input into two 1024-way fully-connected layers. Finally, the regressor regresses bounding boxes, and the classifier predicts classes.

2.2. Backbone network

In most of algorithms of object detection, backbone networks are employed to extract the feature maps from original images. The state-of-the-art backbone network is ResNet [23], which is composed of 5 stages, and the feature maps output from last layers of each 5 stages are denoted as C1, C2, C3, C4, C5, respectively.

Feature Pyramid Network (FPN) [21], is a state-of-the-art algorithm in generic object detection proposed by Lin et al. in 2017. The FPN introduces a pyramid-shaped architecture utilizing the multi-scale features to replace the single scale feature in Faster R-CNN. The feature pyramid architecture, which combines low-resolution, semantically strong features with high-resolution, semantically weak features that has rich semantics at all levels. It is built quickly without sacrificing representational power, speed or memory [21]. In the construction of the feature pyramid, the feature maps {C2; C3; C4; C5} are used to create the feature pyramid. C1 is not included in the pyramid due to its large memory footprint. The feature pyramid relies on an architecture including a top-down pathway and lateral connections. The top-down pathway generates higher resolution feature maps by up-sampling lower resolution feature maps. Each lateral connection merges feature maps of the same spatial size from the bottom-up pathway and the top-down pathway by element-wise addition. This process is

Fig. 2. Image preprocessing procedure.

iterated until the finest resolution map is generated. The final feature maps {P2; P3; P4; P5} are of the same spatial sizes with {C2; C3; C4; C5}.

In the proposed method, we establish a novel feature architecture which is improved from feature pyramid {P2; P3; P4; P5}. First, P2, P3, P4, P5 are resize to the same size as P4 through max-pooling and interpolation. Second, integrated features are obtained by average the rescaled {P2; P3; P4; P5}. Third, we use the embedded Gaussian non-local attention [24] module to refine the integrated features. Fourth, the refined features are then rescaled using the same but reverse procedure to strengthen the original features {P2; P3; P4; P5}, namely element-wise adding refine features to {P2; P3; P4; P5}. Finally, the outputs {S2; S3; S4; S5; S6} are used for object detection following the same pipeline in FPN. Here, S6 is max pooled from S5. In this new architecture, each resolution in the feature pyramid gains the same information from other resolutions, balancing the flow of information and making the features more discriminating.

2.3. Image preprocessing

For X-ray images in the original MURA dataset, we are facing two problems. One is the existence of noise, and the other is the dark background of images. So, we need to perform the image preprocessing for these two issues.

To mitigate the effects of noise, we preprocess images by morphological method. The morphological opening operation with a 21×21 kernel is adopted to process the grayscale image. By the opening operation, the isolated noise in the image can be eliminated, and meanwhile, the main area can be identified. Here, the main area of the image refers to the area containing all bones and fractures.

To increase the brightness of the image, we use cumulative distribution function of the normal distribution to perform gray stretch on the original image. Here, the pixel variance of the main area is taken as the variance of the normal distribution. Considering that the fracture area is often the brightest area in the main area, we take the maximum pixel value of main area as the mean of the normal distribution so as to make the transformation sensitive to the fracture area.

Through above two operations, the contrast of the entire image can be improved, and the fracture area in transformed images becomes clearer and brighter. The preprocessing procedure is shown in the Fig. 2.

2.4. Anchor scales reduction

The RPN introduces “anchor” boxes which serve as references at multiple scales and aspect ratios. At each pixel of feature map, RPN simultaneously predicts multiple Region of Interests (RoIs), which are parameterized relative to the corresponding anchors [20]. In training procedure of RPN, if the intersection over union (IoU) between an anchor and ground-truth is beyond 0.7, then the anchor will be marked as a foreground RoI. If the IoU is below 0.3, then the anchor will be marked as a background RoI. In other cases, the anchor will do not participate in the training. After RoIs are selected, RPN further employs Non-maximal suppression (NMS) to screen RoIs by discarding the ones with large overlap area.

In training procedure of arm fracture detection, the proper setting of the anchor scales is vital, for the IoU between anchors and ground truth bounding boxes is too small to propose a foreground RoI by RPN. This will further result in the lack of positive samples for RPN training and lack of useful information for the network. Therefore, it is very important to set appropriate anchor scale for detection tasks.

In the feature pyramid, the original anchor scale with respect to feature maps {P2; P3; P4; P5; P6} is {512; 256; 128; 64; 32}. This configuration fits to the generic object detection task such as car and pedestrian detections. However, our experiments show that it is not well suited for arm fracture detection tasks. We think the reason is that the ground-truth bounding-box of arm fracture is usually smaller than that of generic objects. So, we consider scaling down the anchor for each feature map. The experiment results reveal that {256; 128; 64; 32; 16} is the best choice of anchor scales for {P2; P3; P4; P5; P6}, because it guarantees more foreground RoIs for RPN training.

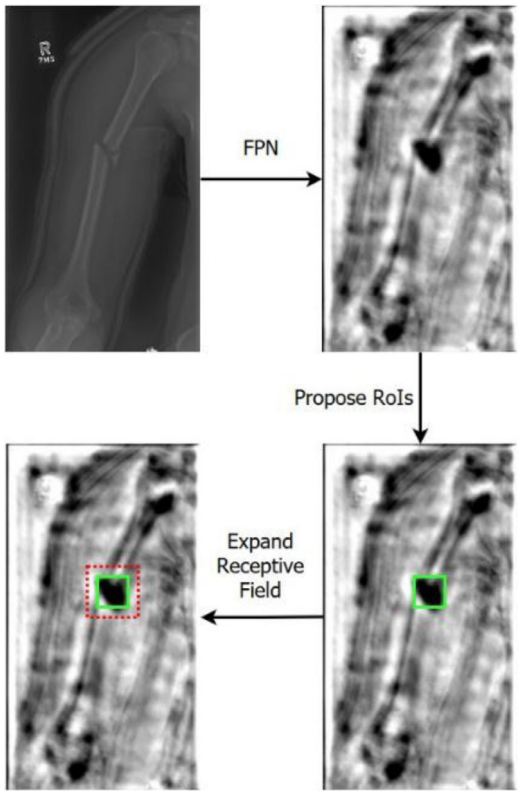


Fig. 3. Expanding receptive field.

2.5. Expanding receptive field to find tiny fracture

Tiny object detection remains a challenge in the field of detection. The detection of tiny fractures in the MURA dataset is the main difficulty in our task. In fact, the tiny RoI with a scale of less than 40×40 pixels contain inadequate information for the network to diagnose whether it contains a fracture. In order to extract useful information from these tiny RoIs, it is necessary to expand the receptive field, just as did in tiny face detection [25]. In tiny face detection, if the face is too tiny, then all locations of hair, shoulders and other adjacent information are taken into account.

The implementation process of expanding receptive field is shown in the Fig. 3. For each RoI, if its width is less than 30 pixels, then 20 pixels are added to its width. If its width is less than 40, then 10 pixels are added to its width. This rule also applies to the length adjustment of the RoIs. The above rule ensures that the receptive fields of tiny fractures can be enlarged.

3. Experiments

The training and testing procedure are shown in Fig. 4. First, the original image is manually annotated by radiologists. Second, the proposed network is trained by the images with annotated ground-truth bounding-boxes. Third, the loss is calculated based on the predictions of temporal model and the ground-truth bounding-boxes, and the weights in the model are updated repeatedly by Stochastic Gradient Descent (SGD). After 12 epochs of training, the testing result of the trained model is optimized, and the training procedure is finished.

3.1. Data and label

3.1.1. Dataset

The MURA dataset, organized by Stanford Machine Learning Group, is employed in our experiments. MURA is one of the largest public radiographic image datasets published in Jan 2018, and it contains 40,895 multi-view radiographic images of the upper extremity including the shoulder, humerus, elbow, forearm, wrist, hand and finger. All images are in 8-bit png format. Our experiments select all 3392 positive images of humerus, elbow and forearm as training data and 612 images as test data. Negative images are not suit for training model of detection. The test dataset including 612 images is already split by MURA organizer. There is no overlap between training dataset and test dataset.

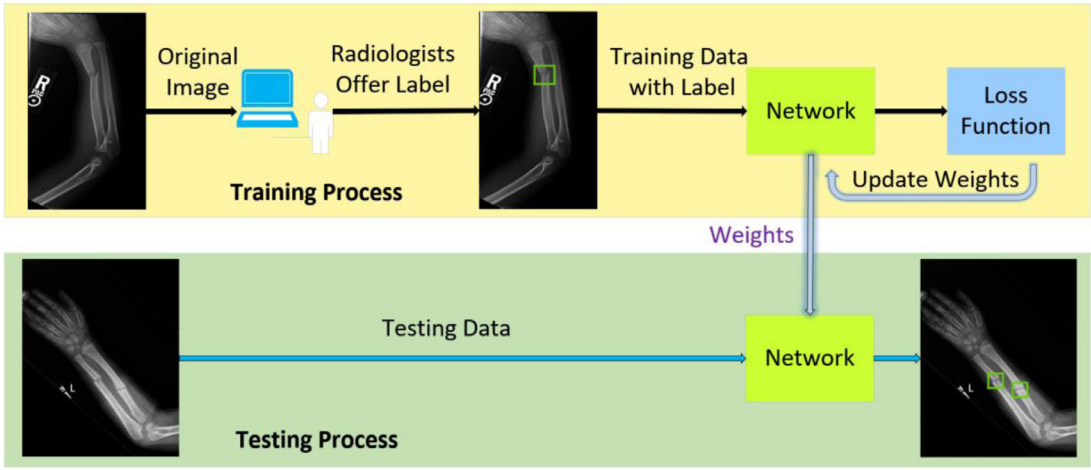


Fig. 4. The pipeline of training and testing procedure.

3.1.2. Label collections

Label collection is vital and complicate in arm fracture detection task. There is no labeled dataset for detection of arm fracture. The annotation in the original MURA dataset only aimed to the binary classification problem, which is unfit to the detection task. So, we invite three radiological experts, who come from the Department of Radiology of Linyi People's Hospital, help us to label these images with bounding boxes, thus we are able to employ the dataset in detection task. The invited radiologists have diagnostic experience of more than 20 years. Each label has been checked twice, and it spends more than one month in completing all the annotations of dataset.

3.1.3. Data augmentation

The dataset including 3392 training images is quite small to fine tune a network which has pre-trained on ImageNet. For training network to generalize to arm fracture X-ray images, data augmentation needs to be considered. We employ horizontal flipping and random rotation in our experiment. The experimental results show that the horizontal flipping can be helpful for training but the random rotation is ineffective.

3.2. Implementation details

The network backbone used in our experiment is pre-trained on ImageNet dataset. The architectures are trained end-to-end. The input image is resized by 800 pixels for shorter side and 1333 pixels for longer side. The model training is adopted to 4 GPU NVIDIA GeForce GTX 1080Ti. A mini-batch involves 4 images. A weight decay of 0.0005 and a momentum of 0.9 are adopted. The learning rate is 0.005 for the first 6 epochs and 0.0005 for the remaining 6 epochs. We employ 5 scale anchors of {16, 32, 64, 128, 256}. In training procedure of RPN, if the intersection over union (IoU) between an anchor and ground-truth is beyond 0.7, then the anchor will be marked as a foreground RoI. If the IoU is below 0.3, then the anchor will be marked as a background RoI. RPN propose 256 RoIs per image for training, criteria are first select as many positive proposals as possible, then randomly select negative proposals to make up 256.

3.3. Data preprocessing

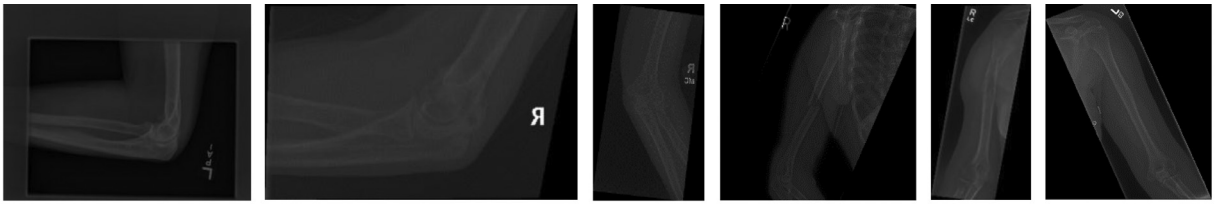
Before training, all images are preprocessed by morphological opening operation and pixel value transformation to make images clearer. The comparison of effect between original images and preprocessed images is shown in Fig. 5.

3.4. Arm fracture detection results

We train the proposed network with 3392 annotated images, and test the model by 612 images. The predicted boxes, whose confidence scores are above 0.4, would be regard as areas containing fractures. The test results are illustrated in Fig. 6. Here, the green boxes are ground-truth boxes annotated by our invited radiologists, and the blue boxes are predicted by trained model.

We also encountered failure cases during experiment. Some failure cases are illustrated in Fig. 7. These failure cases may give us insight on limitation of the proposed algorithm and motivation on future research and development. As for the reason of the failure, we explained it in Section 4.

For the evaluation of the detection performance, the standard metric of precision used in image classification problems cannot be directly applied in the detection problem. Hence, we evaluate the effect of our model by two factors, precision and



(a) Original images.

(b) Preprocessed images.

Fig. 5. Comparison between original images and preprocessed images.

recall, as did in PASCAL Visual Object Classes (VOC) challenge [26]. Here, the predicted box is regarded as correct predicted if the IoU between predicted box and ground-truth boxes is above the fixed IoU threshold. We achieve the average precision (AP) of the detection is 62.04% with IoU = 0.5.

To evaluate the proposed method, we compare the detection results of our method with other state-of-the-art methods proposed in recent years. To be fair, all of these methods have been fine-tuned and show the best results. The experiment results listed in the Table 1 show that, our method could achieve state-of-the-art AP on testing dataset and it remarkably outperforms other state-of-the-art deep learning methods. To further show the effects of the proposed method, we conduct more ablation studies as following.

3.5. Ablation experiments of arm fracture detection

For fair comparisons of the effectiveness of each improvement in the proposed framework, we run 8 ablation experiments respectively with different settings as illustrated in Table 2. The experimental results show that both three improvements can improve the detection performance and combining these three improvements gains the highest AP. From ID 2 vs. ID 3 and ID 4, we can see that the new backbone network is the major improvement.

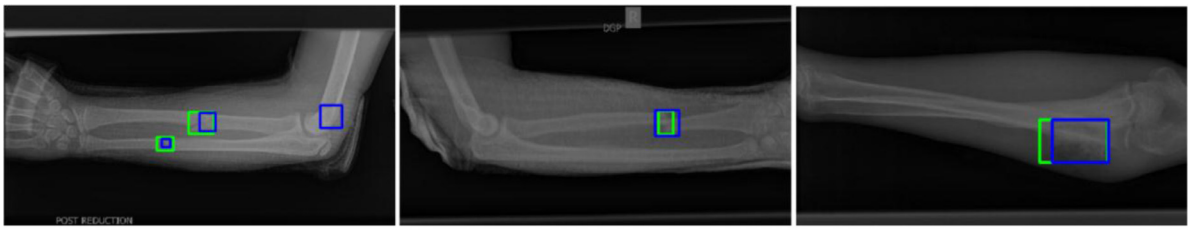
3.6. Impact of feature maps combination

In order to analyze the effect of different combinations of feature maps output from last residual block of each stage in ResNet101, we have compared the detection results of several combinations of feature maps, and list the experiment results in Table 3. Here, the result of each combination is obtained under the condition that IoU is equal to 0.5. We can see that the combination of {C2; C3; C4; C5} is the best combination for the detection. So, in our method, the feature maps {C2; C3; C4; C5} are used to create the feature pyramid. Moreover, we have some other meaningful findings through these experiments. By comparing the detection results based on {C2; C3; C4} and {C3; C4; C5}, we find that the shallow feature maps contain more fracture information than the deep feature maps. Furthermore, although shallow feature maps contain more fracture information, the deep features maps are still useful.

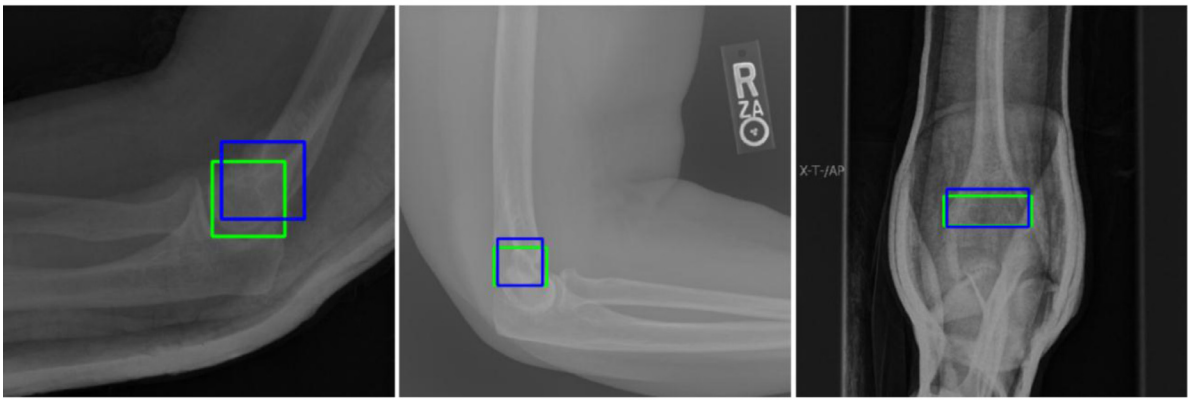
4. Discussions

The basic concept of CAD is to provide a computer output as a second opinion to assist radiologists' image interpretation by improving the accuracy and consistency of radiological diagnosis and also by reducing the image reading time. From the perspective of radiologists, they desire CAD system can provide them more accurate fracture location information, so as to save their diagnosis time and decrease misdiagnosis. As for the CAD of arm fracture, previous methods were far from clinical applications because their APs were less than 30% [3]. In this paper, we improve the state-of-the-art deep learning method and applied to arm fracture detection. The AP of our model reaches to 62.04%, and this ensures that our model can be applied in the clinical environment.

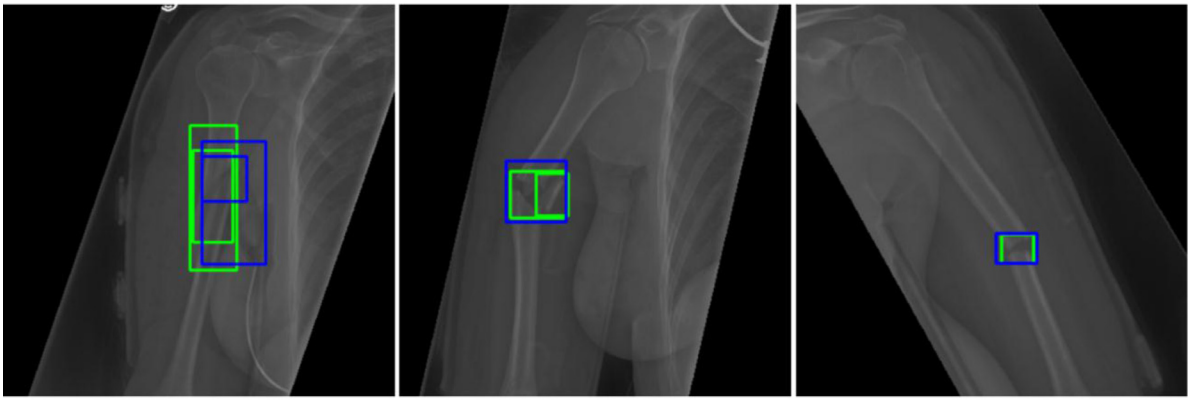
The radiologists think that our model can help them a lot in arm fracture diagnosis. From radiologist's point of view, the existence of some false positive predictions is not a major problem, since in this case they can quickly estimate whether



(a) Fracture in forearm.



(b) Fracture in elbow.



(c) Fracture in humerus.

Fig. 6. Arm fracture detection results (green boxes: ground-truth boxes, blue boxes: predicted boxes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the prediction is wrong. In contrast, what they really concern is whether the real fracture is undetected. In the radiology department of hospital, more than 90% of the medical disputes are caused by omissions in fracture detection.

It should be pointed out that there are still some fractures cannot be accurately detected by our model, although our proposed method could offer help to radiologists to some extent. We think that the limitations of model performance are caused by two factors. One is the quality and number of images in dataset, and the other is the special shape of the fracture. We will explain separately as follows. Firstly, the deep learning method itself is a data-driven method, so the model is very sensitive to the quality of the dataset. In the MUAR dataset, all images are collected from 2001 to 2012, and the resolution of almost all images is less than 600 $\times$ 600. Such a low resolution makes it is very difficult to detect tiny fractures. In fact, for a 600$\times$600 original image, the resolution of the feature map C2 generated after the first few layers of the backbone network ResNet101 is only 150 $\times$ 150, and this makes the area of the tiny fracture only occupy a few pixels in C2. Thus, the model is powerless to detect a tiny fracture. The radiologists we invited tell us that, X-ray images in their hospital's database collected by the newest Digital Radiography (DR) technology have at least 3000 $\times$ 1500 resolution, which is much higher than that of the MURA dataset. Therefore, we believe that the performance of our model will be much better if the model is trained by X-ray images generated by the DR technology. Secondly, as for the number of images in the dataset, 3392 training

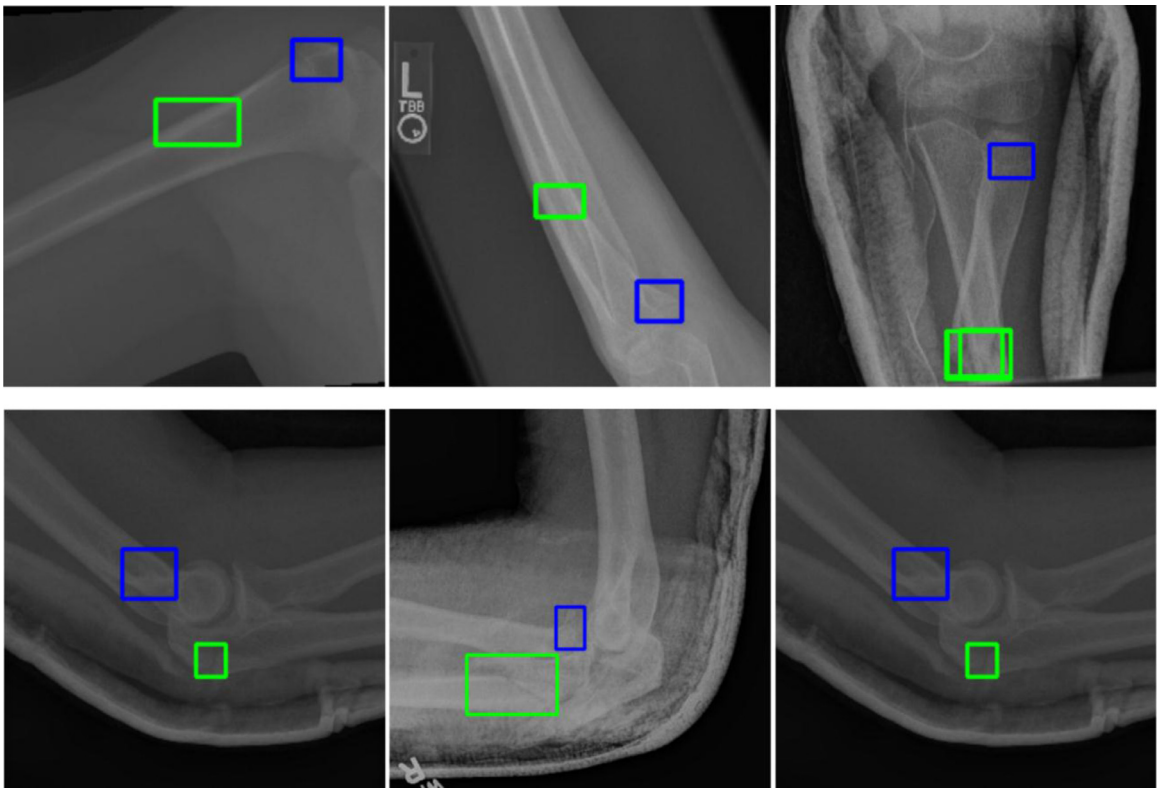


Fig. 7. Failure cases (green boxes: ground-truth boxes, blue boxes: predicted boxes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison of the proposed method with other state-of-the-art methods.

Algorithm	Backbone network	AP (%)
Faster R-CNN [20]	Res101	32.72
FPN [21]	Res50-FPN	52.34
FPN [21]	Res101-FPN	54.28
Cascade R-CNN [22]	Res50-FPN	54.64
FCOS [27]	Res101-FPN	54.86
GCNet [28]	Res50-FPN	54.88
DCN v2 [29]	Res50-FPN	57.68
DCN v2	ResX101-FPN	58.62
Guided Anchoring [30]	ResX101-FPN	59.52
Proposed method	**Proposed backbone**	62.04

Table 2
Ablation experiments.

ID	New backbone network	Image preprocessing	Receptive filed adjustment	AP (%)
1				54.28
2		√		55.34
3			√	56.36
4	√			58.96
5		√	√	58.29
6	√	√		59.56
7	√		√	60.86
8	√	√	√	62.04

Table 3
Comparison of results for different combinations of feature maps.

Feature maps combination	AP (%)
C2, C3	24.82
C3, C4	33.32
C4, C5	6.42
C2, C3, C4	56.28
C3, C4, C5	32.62
C2, C3, C4, C5	62.04

images are undoubtedly small for the training of deep learning models. If the number of training images exceeds 10,000, the model can be trained by more kinds of shapes of fractures, and the performance of model should be further improved. For these reasons, we are currently working with doctors to organize larger, higher quality dataset.

After analyzing our experiment results, the invited radiologists think that some of fractures with special shapes may be missed in the detection. The cases of missing detection of fractures are listed as follows:

(a) Fracture line is hidden.
(b) Fracture area is too tiny.
(c) Gypsum fixed on the arm.
(d) Overlapping fracture.

To solve these questions, one approach is to further improve the deep learning method, and the other is to manually collect more images with special shapes for training dataset.

Computer-aided diagnosis (CAD) algorithms have shown promise to help radiologists detect fractures, but it is well known that the image features that support their predictions are difficult to understand. Further research is needed to clarify the deep learning decision-making processes so that computers and clinicians can effectively cooperate. Although our research is focus on the fracture of the arm bone, the proposed framework is not limited to learn from the X-rays of the arm bone. Given model enough training images and moderately fine-tune the framework, the model can learn to detect any analogous X-rays that human clinicians can recognize. To facilitate the research on the fracture detection, we plan to make the annotations of the dataset available in the future.

5. Conclusion

In this work, we proposed a new deep learning method for arm fractures detection in X-rays. The experiment results show that, even for MURA dataset whose images are in low quality, the proposed method could achieve the state-of-the-art average precision of 62.04% on arm fracture detection and it remarkably outperforms other state-of-the-art deep learning methods. Considering that the experiment results of our model are based on a smaller dataset with low quality images, the fracture detection results can be further enhanced for a larger dataset with higher quality images. Hence, we think the improved deep learning method has strong potential application in real clinical environments.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgments

The authors would like to thank Doctor Wanquan Liu from Curtin University for his valuable comments to improve this paper. The authors also thank Doctor Jinliang Wang and Fuzhou Li for their hard work in the annotation of nearly 4000 radiographs, and thank the radiologists in the Department of Radiology of Linyi People's Hospital for their kindly help in the analysis of our experiment results. This work is supported in part by the National Natural Science Foundation of China under Grants 61473202.

References

[1] 2014 Report | BMUS: the burden of musculoskeletal diseases in the United States, (n.d.). http://www.boneandjointburden.org/2014-report (accessed September 15, 2018).
[2] Mayo RC, Parikh JR. Breast imaging: the face of imaging 3.0. J Am Coll Radiol 2016;13:1003–7. doi:10.1016/J.JACR.2016.03.010.
[3] Cao Y, Wang H, Moradi M, Prasanna P, Syeda-Mahmood TF. Fracture detection in x-ray images through stacked random forests feature fusion. In: Proc. - int. symp. biomed. imaging, IEEE; 2015. p. 801–5. doi:10.1109/ISBI.2015.7163993.
[4] Bandyopadhyay O, Biswas A, Bhattacharya BB. Long-bone fracture detection in digital X-ray images based on digital-geometric techniques. Comput Methods Programs Biomed 2016;123:2–14. doi:10.1016/j.cmpb.2015.09.013.
[5] Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in ct images using deep convolutional neural networks. Pattern Recognit 2019;85:109–19. doi:10.1016/J.PATCOG.2018.07.031.

[6] Hanel D, Daluiski A, Lachapelle A, Gupta A, Chopra S, Hotchkiss R, Gardner M, Potter H, Sicular S, Lindsey R, Mozer M, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, Hanel D, Gardner M, Gupta A, Hotchkiss R, Potter H. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci 2018;115:11591–6. doi:10.1073/pnas.1806905115.
[7] Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 2018;73:439–45. doi:10.1016/j.crad.2017.11.015.
[8] Raghavendra U, Bhat NS, Gudigar A, Acharya UR. Automated system for the detection of thoracolumbar fractures using a cnn architecture. Futur Gener Comput Syst 2018;85:184–9. doi:10.1016/J.FUTURE.2018.03.023.
[9] Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on ct scans. Comput Biol Med 2018;98:8–15. doi:10.1016/J.COMPBIOMED.2018.05.011.
[10] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R.L. Ball, C. Langlotz, K. Shpanskaya, M.P. Lungren, A.Y. Ng, MURA: large dataset for abnormality detection in musculoskeletal radiographs, (2018) 1–10. http://arxiv.org/abs/1712.06957v4 (accessed September 15, 2018).
[11] Ebsim R, Naqvi J, Cootes TF. Automatic detection of wrist fractures from posteroanterior and lateral radiographs: a deep learning-based approach. In: Vrtovec T, Yao J, Zheng G, Pozo JM, editors. Comput. methods clin. appl. musculoskelet. imaging. Cham: Springer International Publishing; 2019. p. 114–25.
[12] England JR, Gross JS, White EA, Patel DB, England JT, Cheng PM. Detection of traumatic pediatric elbow joint effusion using a deep convolutional neural network. Am J Roentgenol 2018;211:1361–8. doi:10.2214/AJR.18.19974.
[13] Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 2019;48:239–44. doi:10.1007/s00256-018-3016-3.
[14] Badgeley MA, Zech JR, Oakden-Rayner L, Glicksberg BS, Liu M, Gale W, McConnell MV, Percha B, Snyder TM, Dudley JT. Deep learning predicts hip fracture using confounding patient and healthcare variables. Npj Digit Med 2018. doi:10.1038/s41746-019-0105-1.
[15] Adams M, Chen W, Holcdorf D, McCusker MW, Howe PDL, Gaillard F. Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. J Med Imaging Radiat Oncol 2019;63:27–32. doi:10.1111/1754-9485.12828.
[16] Pranata YD, Wang KC, Wang JC, Idram I, Lai JY, Liu JW, Hsieh IH. Deep learning and surf for automated classification and detection of calcaneus fractures in ct images. Comput Methods Programs Biomed 2019;171:27–37. doi:10.1016/j.cmpb.2019.02.006.
[17] Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, Zhou K, Bi M, Pan L, Wu W, Liu Y. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop 2019:3674. doi:10.1080/17453674.2019.1600125.
[18] Guan B, Yao J, Zhang G, Wang X. Thigh fracture detection using deep learning method based on new dilated convolutional feature pyramid network. Pattern Recognit Lett 2019;125:521–6. doi:10.1016/j.patrec.2019.06.015.
[19] Girshick R. Fast r-cnn. IEEE int. conf. comput. vis.; 2015.
[20] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137–49. doi:10.1109/TPAMI.2016.2577031.
[21] Lin T, Doll P, Girshick R, He K, Hariharan B, Belongie S, Ai F, Tech C. Feature pyramid networks for object detection. Cvpr. 2017. doi:10.1109/CVPR.2017.106.
[22] Cai Z, Vasconcelos N. Cascade R-CNN: delving into high quality object detection. In: 2018 IEEE/CVF conf. comput. vis. pattern recognit. IEEE; 2018. p. 6154–62. doi:10.1109/CVPR.2018.00644.
[23] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conf. comput. vis. pattern recognit. IEEE; 2016. p. 770–8. doi:10.1109/CVPR.2016.90.
[24] Wang X, Girshick R, Gupta A, He K. Non-local neural networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2018:7794–803. doi:10.1109/CVPR.2018.00813.
[25] Hu P, Ramanan D. Finding tiny faces. In: Proc. - 30th IEEE conf. comput. vis. pattern recognition, CVPR 2017. IEEE; 2017. p. 1522–30. doi:10.1109/CVPR.2017.166.
[26] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. Int J Comput Vis 2010;88:303–38. doi:10.1007/s11263-009-0275-4.
[27] Tian Z, Shen C, Chen H, He T. FCOS: fully convolutional one-stage object detection, (2019). http://arxiv.org/abs/1904.01355 (accessed June 29, 2019).
[28] Cao Y, Xu J, Lin S, Wei F, Hu H. GCNet: non-local networks meet squeeze-excitation networks and beyond, (2019). http://arxiv.org/abs/1904.11492 (accessed June 29, 2019).
[29] Zhu X, Hu H, Lin S, Dai J. Deformable convnets v2: More deformable, better results, arXiv:1811.11168v1, (2018).
[30] Wang J, Chen K, Yang S, Loy CC, Lin D. Region proposal by guided anchoring. arXiv:1901.03278v1, (2019).

Bin Guan is currently working toward Ph.D. degree in School of Electrical and Information Engineering, Tianjin University, China. He received his B.S degree in Information and Computer Science from University of Jinan, China in 2017. His research interests include object detection, fracture detection, medical image processing and deep convolutional neural networks.

Guoshan Zhang is currently a Professor in School of Electrical and Information Engineering, Tianjin University, China. He received his B.S. degree in Mathematics from Northeast Normal University, China, M.S. degree in Applied Mathematics, and Ph.D. degree in Industrial Automation from Northeastern University, China, respectively. His research interests include linear and nonlinear system control, intelligent control and pattern recognition.

Jinkun Yao is currently a deputy chief physician at Department of Radiology, Linyi People's Hospital, China. He received his B.S degree in Clinical Medicine from Binzhou Medical College, China in 1996, M.S. degree in Clinical Medicine from Shandong University, China in 2015. His research focuses on X-ray, CT and MRI imaging diagnosis.

Xinbo Wang is currently working toward Ph.D. degree in Control theory and Control Engineering, School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include deep learning, pattern recognition and image processing.

Mengxuan Wang is currently working toward M.S. degree at School of Electrical and Information Engineering, Tianjin University, China. His research interests include object detection, computer aided detection and deep learning.