Neural Computing and Applications (2022) 34:18773–18785
https://doi.org/10.1007/s00521-022-07412-0

ORIGINAL ARTICLE

Attention mechanism-based deep learning method for hairline fracture detection in hand X-rays

Wenkong Wang[1] · Weijie Huang[1] · Quanli Lu[2] · Jiyang Chen[2,3] · Menghua Zhang[1] · Jia Qiao[1] · Yong Zhang[1]

Received: 13 January 2022 / Accepted: 9 May 2022 / Published online: 24 June 2022

Abstract

Wrist and finger fractures detection is always the weak point of associate study, because there are small targets in X-rays, such as hairline fractures. In this paper, a dataset, consisting of 4346 anteroposterior, lateral and oblique hand X-rays, is built from many orthopedic cases. Specifically, it contains a lot of hairline fractures. An automatic preprocessing based on generative adversative network (GAN) and a detection network, called WrisNet, are designed to improve the detection performance of wrist and finger fractures. In the preprocessing, an attention mechanism-based GAN is proposed for obtaining the approximation of manual windowing enhancement. A multiscale attention-module-based generator of the GAN is proposed to increase continuity between pixels. The discriminator and the generator can achieve 93% structural similarity (SSIM) as manual windowing enhancement without manual parameter adjustment. The designed WrisNet is composed of two components: a feature extraction module and a detection module. A group convolution and a lightweight but efficient triplet attention mechanism are elaborately embedded into the feature extraction module, resulting in richer representations of hairline fractures. To obtain more accurate locating information in this condition, the soft non-maximum suppression algorithm is employed as the post-processing method of the detection module. As shown in experimental results, the designed method can have obvious average precision (AP) improvement up to 7% or more than other mainstream frameworks. The automatic preprocessing and the detection net can greatly reduce the degree of artificial intervention, so it is easy to be implemented in real clinical environment.

Keywords Attention mechanism · Generative adversative network · Soft non-maximum suppression · Hairline fractures

1 Introduction

Deep learning [1] is one branch of artificial intelligence that creates computer models to tackle some vision tasks including object detection [2], image classification, etc. The great success of deep learning in the field of computer vision inspires many scholars to apply it to medical image analysis. A classic example is fracture detection in X-rays by using deep learning-based method. Current deep learning-based object detection algorithms are mainly divided into two branches: one-stage and two-stage [3]. The one-stage detector shows great real-time performance, but the detection precision is lower than that of the two-stage detector. In practice, the fracture detection has no need for real time operating, so two-stage detector is widely used in the diagnosis of X-rays [4].

The gray scale of X-rays is compressed in a small range that is not conducive to differentiate crack features. So as to make the skeletal structure more conspicuous, manual windowing enhancement is adopted as preprocessing to deal with this problem, but each image needs to manually select the window level and window width [5]. Furthermore, during wrist and finger fracture detection, hairline

✉ Weijie Huang
cse_huangwj@ujn.edu.cn

✉ Yong Zhang
cse_zhangy@ujn.edu.cn

[1] School of Electrical Engineering, University of Jinan, No. 336, West Road of Nanxinzhuang, Jinan 250022, Shandong, China

[2] Shandong Zhengzhong Information Technology Co., LTD, Jinan 250014, Shandong, China

[3] Shandong University, Jinan 250061, Shandong, China

fractures in X-rays are difficult to detect with state-of-the-art methods in that small object detection [6, 7] is a challenging task for deep learning-based ways. To solve the two problems mentioned above, an automatic preprocessing based on GAN [8] and a detection network, called WrisNet, are designed in this paper. The main contribution can be concluded as follows.

(1) A dataset, consisting of 4346 anteroposterior, lateral and oblique hand X-rays, is built from many orthopedic cases. It should be pointed that hairline fractures account for more than 50 percent of the total targets in the dataset, way more compared to the published datasets.
(2) An attention mechanism-based GAN is proposed as the preprocessing to expand the gray scale range. The goal of the proposed GAN is obtaining the approximation of manual windowing enhancement. We design a novel generator, consisted of multiscale attention-module-based net to process the input image, respectively. The GAN can achieve 93% SSIM of manual windowing enhancement without manual parameter adjustment and greatly reduce the degree of artificial intervention.
(3) In order to deal with hairline fractures of the dataset, a novel network, called WrisNet, is proposed to improve the detection performance. A feature extraction module and a detection module are formed WristNet. In the feature extraction module, the ResNeXt with the triplet attention (TA) is designed to extract the features while in the detection module, the soft non-maximum suppression (Soft-NMS) algorithm is used as the post-processing mechanism to improve the omission of hairline fractures. The results show that the AP can achieve 7% or more improvement than the state-of-the-art frameworks.

This paper is organized as follows. In Sect. 2, medical image preprocessing methods and deep learning-based fracture detection methods are reviewed. The proposed preprocessing and WrisNet are detailed in Sect. 3. In Sect. 4, several experimental results are illustrated to validate the improved detection performance. Finally, the conclusion is given in Sect. 5.

2 Previous work

2.1 GAN in medical image processing

GANs have great application potential in the field of medical image processing. The main tasks they can solve can be divided into image generation and image translation. In the aspect of image generation, the structure information existing in the train dataset is used to generate new medical images. GANs are often used to increase the number of train datasets to foster the accuracy of classification tasks. A new generation method called generating adversarial Unet was developed by Chen et al. [9], which can realize the generation of various medical images to alleviate the over fitting phenomenon in training. A method that using cycle-consistent adversarial networks to generate COVID-19 samples was suggested by Morís et al. [10] to improve the accuracy of classification. The applicability of generating images by GAN in oncology was demonstrated by Han et al. [11]. The image translation of medical images mainly includes super-resolution reconstruction, image denoising and so on. The conditional generation adversarial network (CGAN) [12] was used as a denoising algorithm in [13] for low dose chest images and the proposed method was proved that was superior to the traditional method. A new super-resolution generation countermeasure network was proposed by Zhu et al. [14], which combines CGAN and super-resolution generation adversarial network (SRGAN) to generate super-resolution images. By extracting useful information from different channels and paying more attention to meaningful pixels, a new convolutional neural network was proposed by Gu et al. [15] for super-resolution in medical imaging. Jiang et al. [16] proposed an improved loss function obtained by combining four loss functions, and this loss function achieved good results in the field of super-resolution CT image reconstruction. In this paper, GAN is firstly used as medical image preprocessing to expand the gray scale range. Meanwhile, a multiscale attention-module-based generator is proposed to process the image. The result of GAN achieves 93% SSIM as manual windowing enhancement without manual parameter adjustment.

2.2 Fracture detection by deep learning-based method

Considering the accuracy of fracture classification and fracture location, Guan et al. [17] proposed an improved object detection algorithm for the detection of arm fractures and obtained a model with a high AP. Qi et al. [18] trained an object detection model to locate femoral fractures by using a framework based on Faster-RCNN [19] and achieved a good result. In [20], a dilated convolutional feature pyramid network was designed, which was applied to thigh fracture detection. In [21], the deep learning method was employed to process the CT images of spine as well as to locate spinal fracture. In [22], the top layer of the original model was retrained by using inception v2 network [23] for leg bone fracture detection. Nonetheless, the above methods could not be applied to the proposed dataset due to

poor hairline break detection performance. To better solve the problem of detecting small targets, a feature extraction module, called ResNeXt-TA, is proposed to make fracture features more prominent. In addition, Soft-NMS is designed as the specialized post-processing of a detection module to improve the omission of hairline fractures.

3 Methodology

An automatic preprocessing based on GAN and WrisNet is proposed for X-ray diagnosis of wrist and finger fractures, which are detailed in Sects. 3.1 and 3.2, respectively. The original image is input into the GAN for gray stretch. The output is operated into WrisNet for detecting fractures of X-rays.

3.1 GAN-based preprocessing

The X-ray gray value is compressed in a small range, which is not conducive to the identification of crack features. A very efficient way of gray stretch is manual windowing enhancement but the window level and window width of each image are need to be manually set. In this paper, a GAN is firstly proposed to expand the gray scale. Inspired by pix2pix [24], a multiscale attention-module-based generator and a discriminator are designed to form the GAN. The structure of the generator is shown in Fig. 1. The architecture is modeled with encoding process and decoding process, which are corresponding to 8 down-samplings and 8 up-samplings, respectively. A CBAM module [25] is embedded at each scale. 16 CBAM modules and the encoding-decoding architecture are formed the generator. The discriminator of pix2pix is directly transplanted to the proposed GAN. The designed generator can greatly increase continuity between pixels of the generated image, compared with pix2pix, and the comparing results can be seen in Sect. 4. The gray scale of the output can be controlled in a reasonable range, which can help the following WrisNet to detect hairline fractures better.

3.2 WrisNet-based fracture detection

The network diagram of WrisNet mainly consists of two components and is shown in Fig. 2. The first component is the feature extraction module to extract the feature maps of the X-rays, which is detailed in Sect. 3.2.1. The second component is the detection module, which can output the exact location of the fractures by analyzing the feature maps obtained in the first component and is detailed in Sect. 3.2.2.

3.2.1 Feature extraction module

The proposed feature extraction module is inspired by Faster-RCNN, mainly composed of ResNeXt-TA and FPN [26].

(1) *ResNeXt-TA*

ResNeXt-TA is a proposed backbone, composed of C1, C2, C3, C4 and C5. A convolution layer, a batch normalization layer [27], a ReLU activation function [28] and a maxpool layer are formed as C1.

C2, C3, C4 and C5 are designed with different number (3, 4, 23, and 3) of blocks. The structure of each block is inspired by ResNet-block [29], and the chart of one block is described as Fig. 2. Each block is formed by a residual connection and a ReLU layer. The residual connection contains the following components in order:

(a) a convolution layer,
(b) a batch normalization layer,
(c) a ReLU layer,
(d) a group convolution [30],
(e) a batch normalization layer,
(f) a ReLU layer,
(g) a convolution layer,
(h) a batch normalization layer,
(i) a TA module,
(j) a shortcut connection.

The group convolution: The input tensor is firstly divided into 64 groups in the channel dimension, then they are convolved with 64 different convolution layers, respectively. Finally, the results of the convolution are concatenated on the channel dimension as the output of group convolution. When the depth and width of the network are increased to a certain extent, increasing the number of groups can improve the performance of feature extraction module effectively.

TA module: The TA module is used from [31] and the detailed structure of TA module is shown in Fig. 3, which is composed of three different sub-branches. The TA module can be expressed as Eq. (1):

$$M(F) = AVG\left[M_{0,1,2}(F) + M_{1,0,2}(F) + M_{1,2,0}(F)\right] \tag{1}$$

The formulas of the three branches are expressed as Eqs. (2), (3) and (4):

$$M_{0,1,2}(F) = \sigma\left(f^{7\times7}(Z - \mathrm{Pool}(F))\right) \tag{2}$$

$$M_{1,0,2}(F) = P_{0,1,2}\left(\sigma\left(f^{7\times7}\left(Z - \mathrm{Pool}\left(P_{1,0,2}(F)\right)\right)\right)\right) \tag{3}$$

$$M_{1,2,0}(F) = P_{0,1,2}\left(\sigma\left(f^{7\times7}\left(Z - \mathrm{Pool}\left(\left(P_{1,2,0}(F)\right)\right)\right)\right)\right) \tag{4}$$

where F is the input tensor with size $C \times H \times W$. $P_{0,1,2}(\cdot)$, $P_{1,0,2}(\cdot)$, $P_{1,2,0}(\cdot)$ refer to the dimensional transformation operations that convert the size of F to $C \times H \times W$, $H \times$

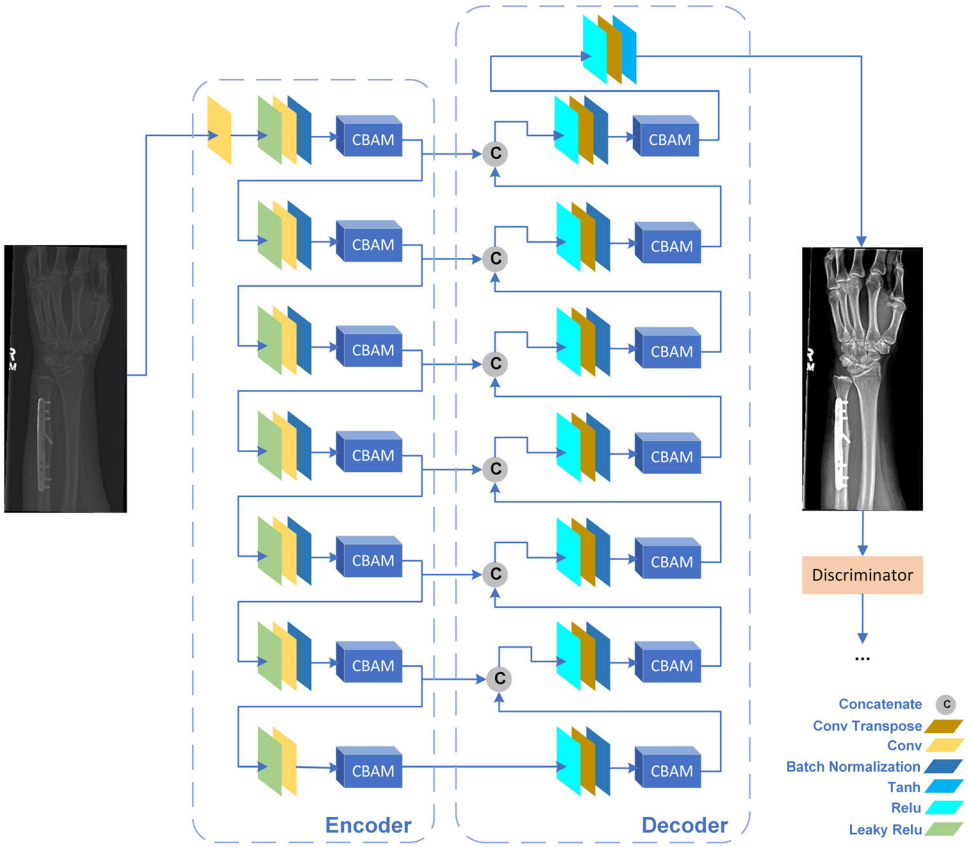

Fig. 1 Network diagram of proposed GAN

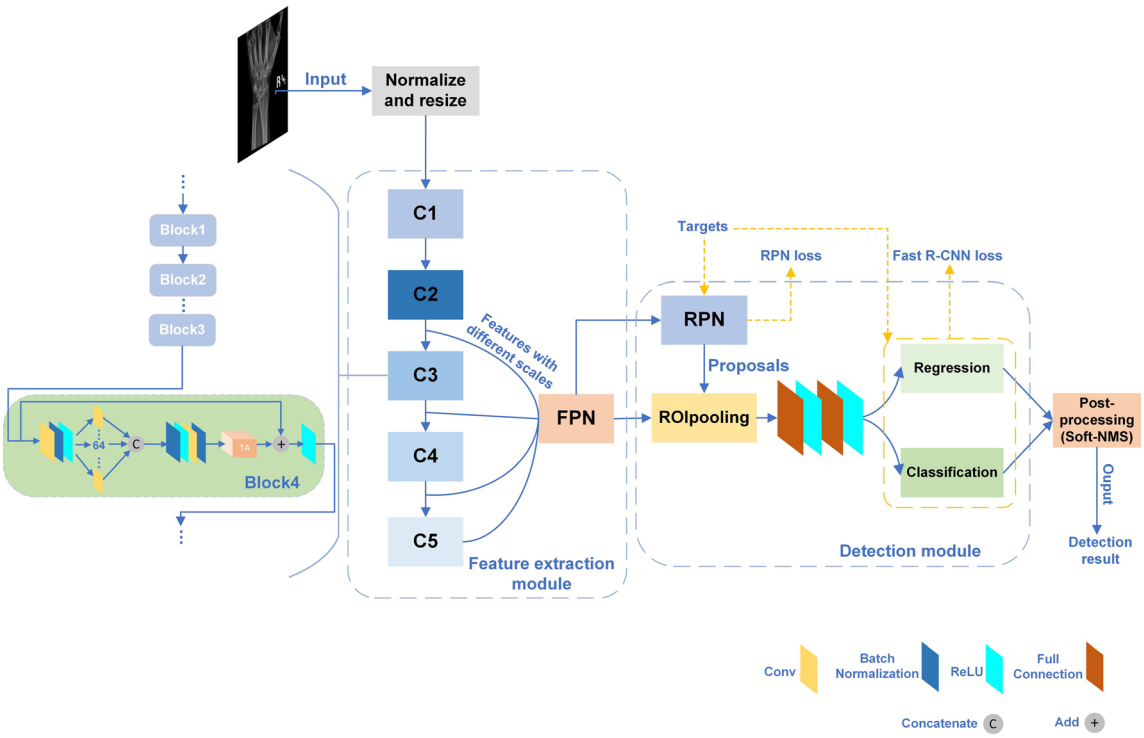

Fig. 2 Network diagram of WrisNet

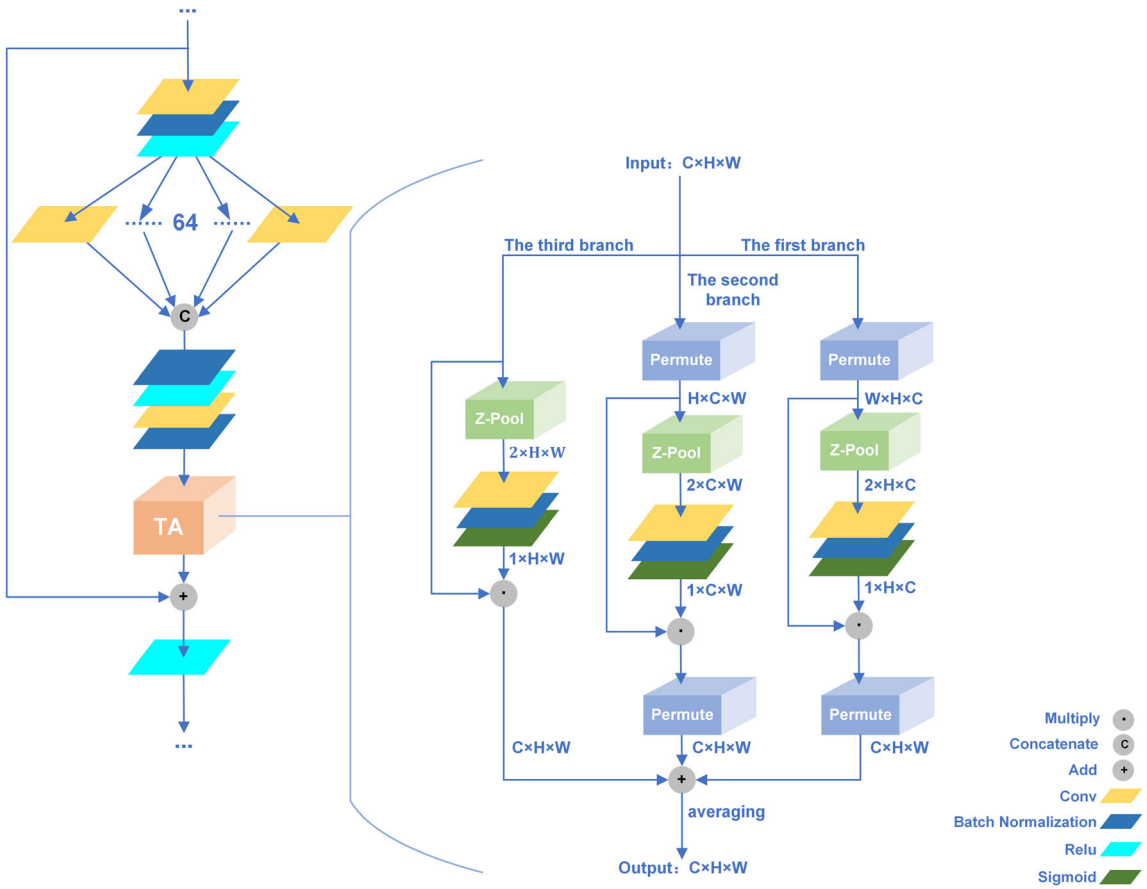

Fig. 3 Block diagram of ResNeXt-TA and the three-branch structure of TA module

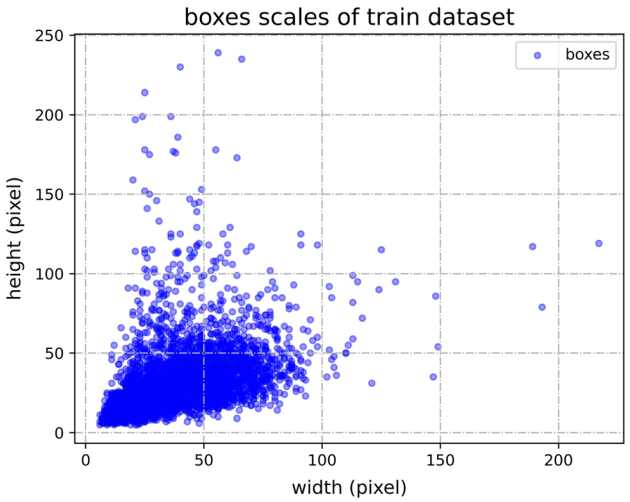

Fig. 4 Size distribution of ground truth boxes in train dataset

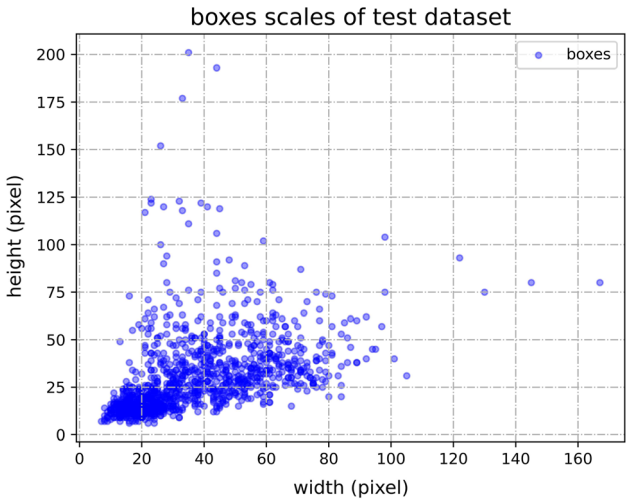

Fig. 5 Size distribution of ground truth boxes in test dataset

$C \times W$ and $H \times W \times C$, respectively. $f^{7\times7}(\cdot)$ refers to the convolution operation with 7×7 kernel size. And $\sigma(\cdot)$ is the sigmoid operation. Z-Pool(.) in the above formula can be expressed Eq. (5):

$$Z - \text{pool}(F) = [\text{MaxPool}(F), \text{AvgPool}(F)] \tag{5}$$

where MaxPool$(\cdot)$ and AvgPool$(\cdot)$ refer to the global maximum pooling and the global average pooling operations, respectively.

The lightweight TA module is located after the third BN layer of each block without adding too many parameters.

Although few parameters does it contain, it could still help each block effectively to understand what information should be laid more emphasis on in the X-ray. In addition, spatial attention is combined with channel attention [32] so that the module could learn the interdependencies between different dimensions and generate more meaningful representations of wrist and finger fractures.

(2) *Multi-scale feature extraction for small targets*

According to the analysis of the statistical data (as shown in Figs. 4 and 5), we find that the size distribution of ground truth boxes is scattered and there are a large number of hairline fractures. FPN is used in feature extraction module to prevent the features of small fractures from being lost during feature extraction. As shown in Fig. 2, feature maps of different scales are extracted from the output of $C2$, $C3$, $C4$ and $C5$ in ResNeXt-TA. These feature maps are fused from top to bottom to obtain more meaningful feature maps.

3.2.2 Detection module

In the detection module, a large number of regular anchors are artificially preset in the RPN, and then the proposal coordinates representing the foreground area are obtained through selection and regression. Then they are projected onto the multi-scale feature maps generated in Sect. 3.2.1. The feature matrixes are segmented on the feature maps according to the corresponding proposals and flattened with the ROI pooling layer. Next, the predicted location and the label information are obtained through the regression layer and Softmax layer, respectively. Finally, a post-processing method Soft-NMS [33] is used to filter the redundant output of the network. The execution process is defined in the Algorithm 1. Set $\mathcal{B}$ contains $\mathcal{N}$ detected boxes and each detection box has its own confidence. Soft-NMS reduces the confidence of the possibly redundant detection boxes instead of removing them directly. First, the confidences in set $\mathcal{S}$ are sorted from high to low. The detected box b_m with the highest confidence is added to the set $\mathcal{M}$, which is merged into $\mathcal{D}$, and b_m is removed from $\mathcal{B}$. Then, the remaining boxes in $\mathcal{B}$ are checked one by one, and their confidence scores are reduced by the function $f(i\sigma u(\mathcal{M}, b_i))$ which are shown in Eq. (6). The progressive loops until all the boxes in $\mathcal{B}$ are put into $\mathcal{D}$. Finally, the boxes with confidence lower than the threshold in $\mathcal{D}$ are considered as repeated fracture localization. Soft-NMS can greatly improve the detection effect in the above-mentioned special case by this kind of scoring reduction mechanism.

Algorithm 1 Soft-NMS

```
Input: B = {b_1, b_2, ..., b_N}, S = {s_1, s_2, ..., s_N}
Begin
    D ⇐ {}
    While B ≠ empty do
        m ⇐ argmax S
        M ⇐ b_m
        D ⇐ D ⋃ M; B ⇐ B − M
        for b_i in B do
            s_i ⇐ s_i f (iσu (M, b_i))
        end
    end while
end
```

$$s_i = \begin{cases} s_i, & \text{iou}(\mathcal{M}, b_i) < N_t \\ s_i(1 - \text{iou}(\mathcal{M}, b_i)), & \text{iou}(\mathcal{M}, b_i) \geq N_t \end{cases} \tag{6}$$

where N_t is the NMS threshold.

4 Experiment

4.1 Dataset

4346 X-rays of wrist and finger fractures including distal radius fractures, scaphoid fractures, phalanx fractures, and other types are utilized in the experiment, which are collected from real medical environment in regular hospitals. The labels of ground truth boxes are completed by the experienced radiologists using LabelImg over one month. The annotations are stored as XML files with PSACAL VOC format. This dataset brings more challenges because there are many X-rays with steel nails, plates, and plaster on the hand. The 4346 X-rays are randomly divided into a train dataset and a test dataset with a ratio of 8 : 2, which is

Table 1 Data statistics

Maximum height	Maximum width	Number of targets	Number of small targets	Number of medium targets
512 (pixel)	512 (pixel)	1116	600	511

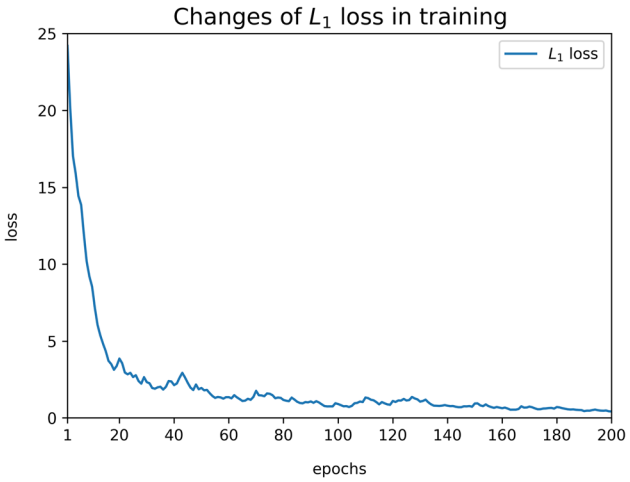

Fig. 6 Changes of L_1 loss in training GAN

always guaranteed during the experiment. We make statistics on the size of the X-rays and the number of targets of different sizes in test dataset(see Table 1). In this paper, a target with a size less than 32×32 is defined as a small target [34], which accounted for more than 53.7% in the test dataset. And targets with a size between 32×32 and 96×96 are considered as medium targets. The distribution of targets in the train dataset and test dataset is shown in Figs. 4 and 5.

4.2 Training details of GAN

4.2.1 Manual image preprocessing

The manual window technique is generally used to preprocess the X-ray image. First, a certain range is selected, where the maximum and the minimum are set as the thresholds. The pixel value greater than max is set to 255, and the pixel value less than min is set to 0. Then, the pixel values in the range are mapped to 0–255 using a linear conversion. The formula for pixel value mapping is shown in Eq. (7):

$$P = \frac{255 \times (P_o - \min)}{\max - \min} \tag{7}$$

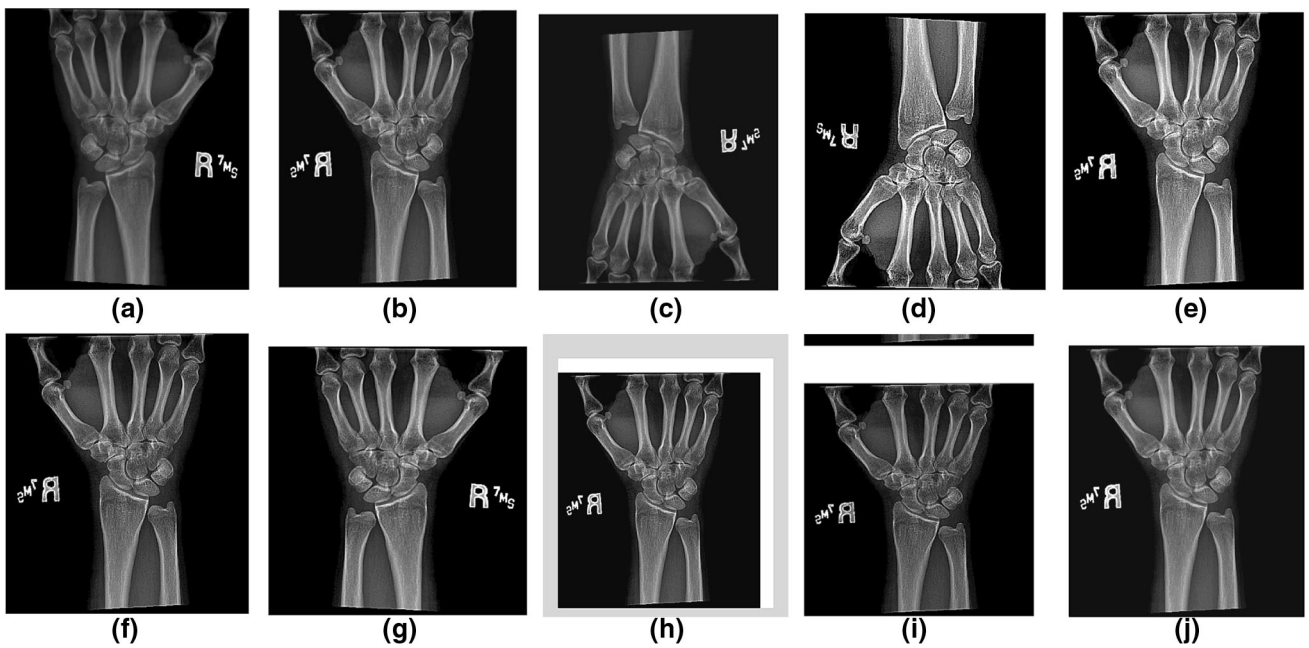

Fig. 7 The data augmentation process includes random flips, brightness transformations, affine transformations, and image sharpening, designed to enhance the X-rays of the train dataset. The input X-rays are randomly subjected to the above four transformations, and the relevant parameters are randomly selected within a certain range. **a** Is the original X-ray, while the data-augmented results are shown in (**b**–**j**)

Table 2 Hyperparameters for training WrisNet

Learning rate	Batch size	Momentum (SGD)	Weight decay	Total epochs
0.02	16	0.9	0.0001	23

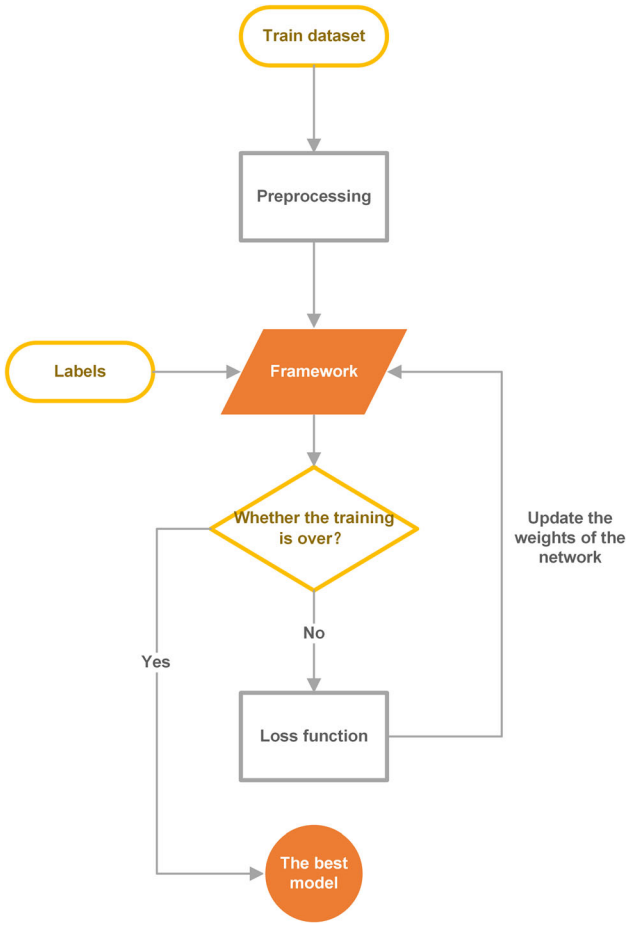

Fig. 8 Process of training. WrisNet loads the images and annotations of the train dataset, updating the weight of the network in repeated iterations

where P_o is the original pixel value and P is the pixel value after linear conversion.

The X-rays with the manual window adjustment are used as the ground truths of GAN.

4.2.2 Training process of GAN

The GAN model is trained on a GPU NVIDIA GeForce RTX 3090. The settings of training are as follows. Adam gradient descent algorithm [35] is adopted. The batch size is set to 1 and a total of 200 epochs are trained. The initial learning rate is set to 0.0002, and a linear learning rate decay strategy is adopted at the 100th epoch.

Table 3 Similarity distribution of test dataset using pix2pix

Threshold of SSIM	Number of images in test dataset
SSIM < 90%	212
SSIM < 80%	45
SSIM < 50%	3

In the training process, so as to ensure that the automatically preprocessed images are similar to the manually preprocessed images, L_1 loss is used in the loss function of the generator to guide the generation of images. The change of L_1 loss during training indicates the process of gradually approaching the pixel values of the automatically preprocessed image and the manually preprocessed image, as shown in Fig. 6.

4.3 Training details of WrisNet

4.3.1 Data augmentation

In the experiment, two data augmentation [36] strategies are set to improve the performance. One strategy is that the data are tripled by flipping the image in random directions. And the other strategy is that the data are increased ten times by using random flips, brightness transformations, affine transformations, and image sharpening. Some transformed images are shown in Fig. 7.

4.3.2 Training process of WrisNet

The pretrained weights on ImageNet [37] are using to initialize the backbone. The model is trained end-to-end on four GPU NVIDIA GeForce RTX 3090. The hyperparameters are shown in Table 2. The warm-up strategy is used in the first 500 iterations. The SGD gradient descent is adopted. Furthermore, the training process is shown in Fig. 8.

4.4 Results and analyses

4.4.1 GAN-based preprocessing

The proposed GAN is compared with Unet-based [38] pix2pix in two different ways, which are described as follows:

Table 4 SSIM comparison between pix2pix and the proposed GAN

	pix2pix (%)	The proposed GAN (%)
Average of test dataset (SSIM < 80%)	69.86	74.14
Maximum of test dataset (SSIM < 80%)	79.99	95.53
Average of test dataset (SSIM < 90%)	82.84	86.17
Maximum of test dataset (SSIM < 90%)	89.99	99.12
Average of test dataset (SSIM < 100%)	92.62	92.90
Average of test dataset (SSIM < 100%)	99.53	99.59
Maximum improvement of single image	77.75	95.53

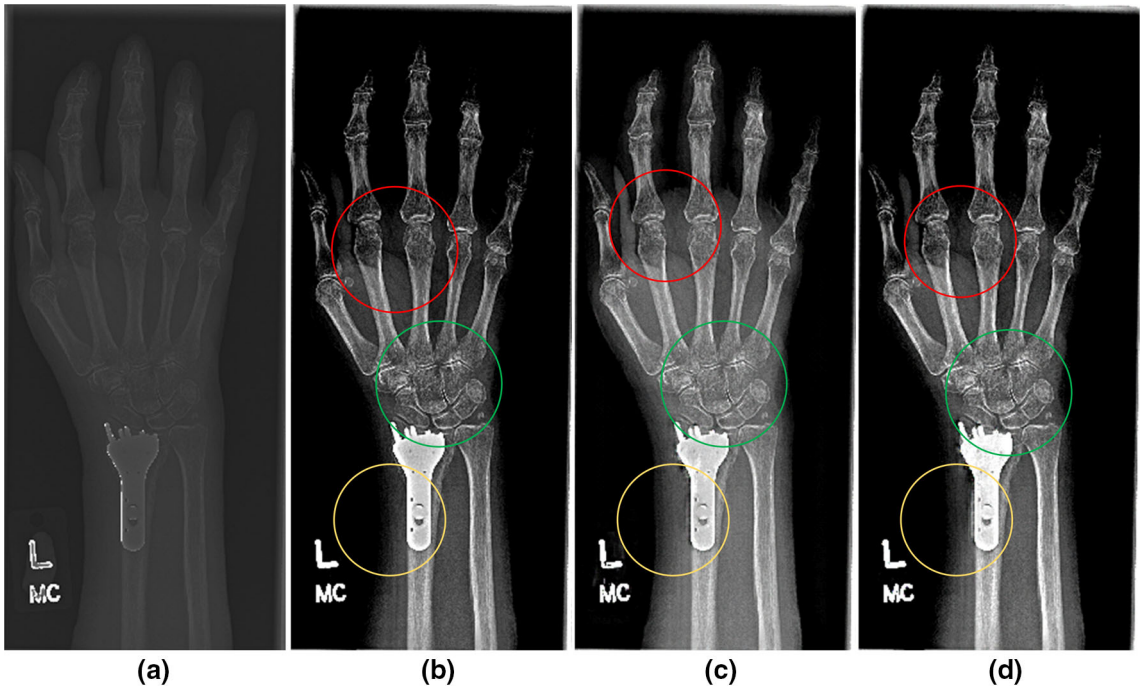

Fig. 9 Comparison of the generated images. **a** Is the original X-ray image. **b** is the ground truth. **c**, **d** Are generated by pix2pix and the proposed GAN, respectively

Table 5 Effect of manual and generated images on object detection

Algorithm	Manual image (AP%)	Generated by Unet (AP%)	Generated by proposed generator (AP%)
Faster R-CNN (ResNet50)	47.4	47.1	47.4
Faster R-CNN (ResNeXt101)	49.2	48.9	49.4

(1) SSIM value

The distribution of SSIM value in test dataset using pix2pix are shown in Table 3. The SSIM value which is less than 90% can be great improved by using the proposed GAN. The SSIM comparison between pix2pix and the proposed GAN is shown in Table 4, where the SSIM of single image can be increased from 77.75 to 95.53%, which proves that the proposed GAN can obtain the approximation of manual windowing enhancement. The comparison of the generated images is shown in Fig. 9. The result shows that the image generated by the proposed GAN is more similar than pix2pix with ground truth (see circles in Fig. 9) and proves that the attention-module-based generator can greatly improves the correlations between pixels.

(2) AP value

As shown in Table 5, the proposed generator ensures the consistency of the detection results between the generated images and the manual images, compared with Unet. The

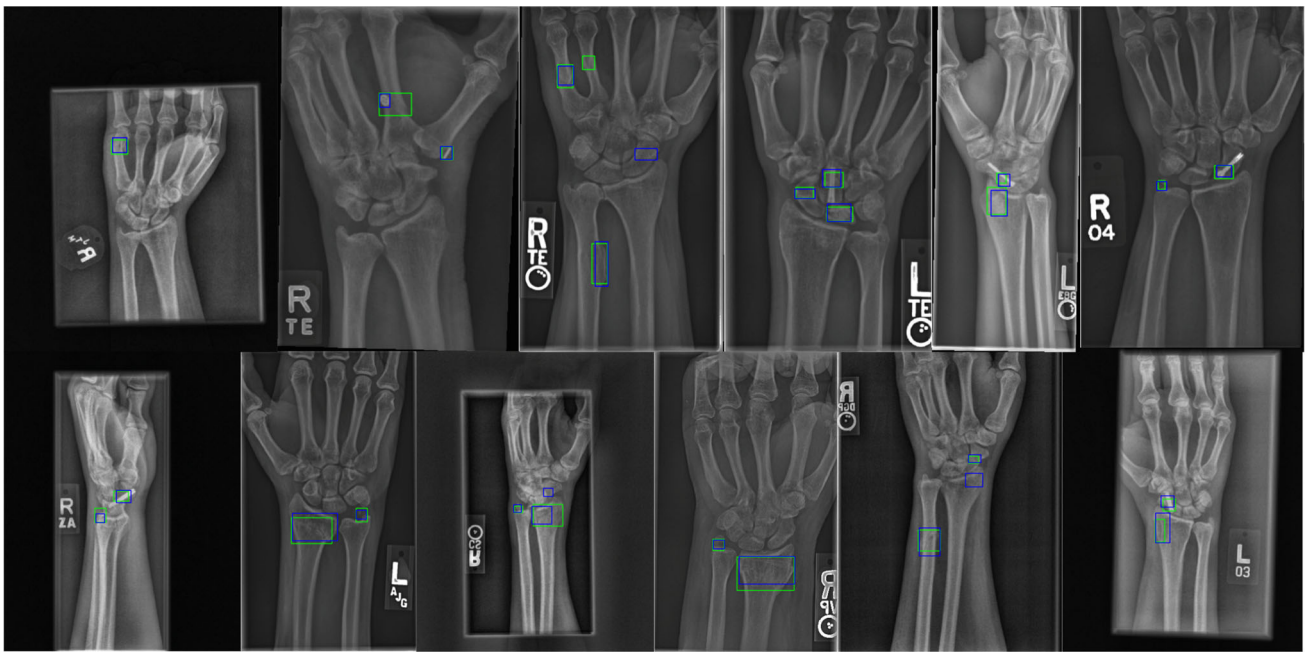

Fig. 10 Our model has a good effect on the detection of phalanx, hand scaphoid and distal radius. The first to third results from the upper left include the detection of phalangeal fractures. Others include the scaphoid and distal radius fractures

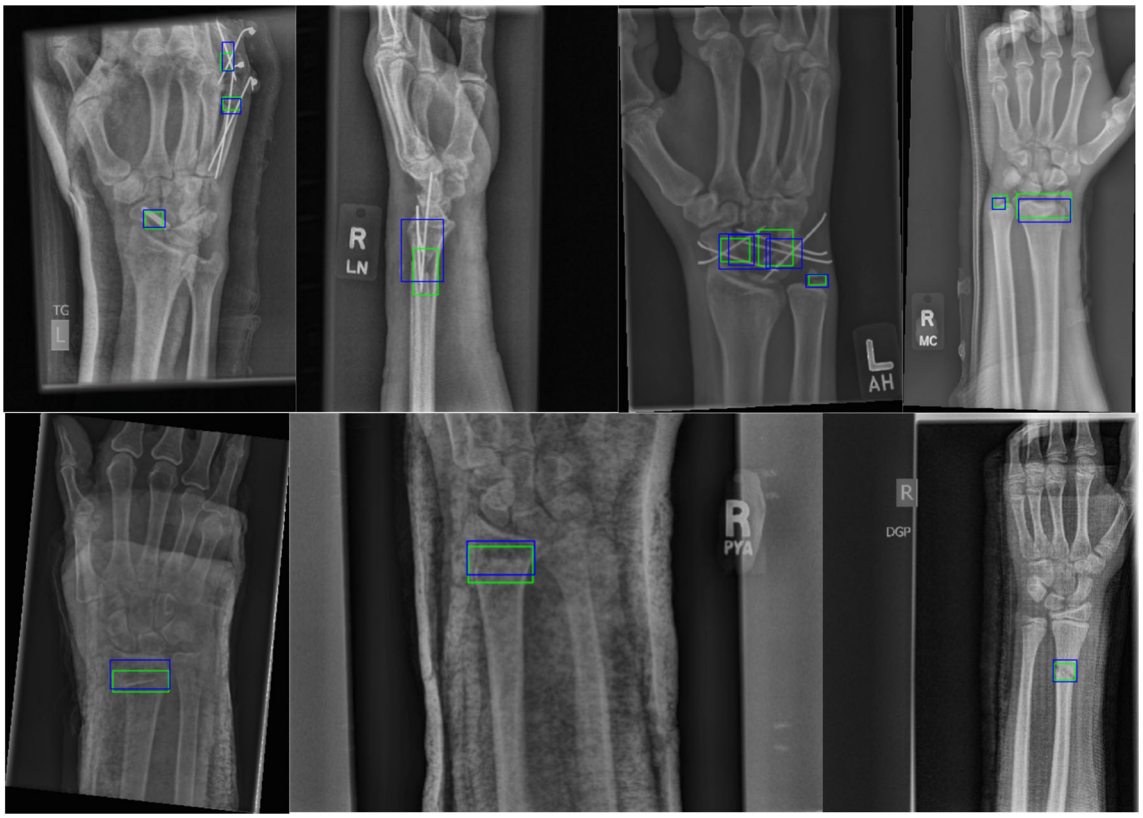

Fig. 11 Under the influence of plaster and steel nails, the model still has considerable detection effect. The first to third results from the upper left include the detection with steel nails. Others include the detection with plasters

Table 6 Comparison of different frameworks

Algorithm	Backbone	AP (%)
Faster R-CNN	ResNet50	47.4
Faster R-CNN	ResNeXt101	49.2
Cascade R-CNN [39]	ResNet50	48.2
Cascade R-CNN	ResNet101	48.4
Cascade R-CNN+DCN [40]	ResNet101	48.3
WrisNet	ResNeXt-TA	**54.7**
WrisNet (best effect)	ResNeXt-TA	**56.6**

detection effect of generated images even over the manual test dataset, due to the elimination of the influence of subjective factors in preprocessing.

4.4.2 Comparison of detection effect

3476 X-rays are using to train the WrisNet, and some detection results of the test dataset are shown in Figs. 10 and 11. The green boxes in figures are the ground truth boxes marked by the doctors, and the blue boxes are detected by WrisNet. As shown in Fig. 10, WrisNet has excellent results in the fracture detection of phalanx, scaphoid, and distal radius, which is reflected in the large overlap area between the detection boxes and the corresponding ground truth boxes. At the same time, as shown in Fig. 11, the model can also perform well in complex environments such as X-rays with nails or plaster. The result can demonstrate that the effectiveness is very close to the diagnosis of radiologists.

The detection effects of the representative object detection frameworks are compared with WrisNet, and the results are shown in Table 6, where the significant improvement of our method is marked in bold. All the frameworks use 3476 X-rays as the train dataset and 870 X-rays are set as the test dataset. The same image preprocessing method and the first data augmentation strategy are used in this part. Furthermore, the pretrained weights on ImageNet are used in all frameworks to initialize the backbone network, and the hyperparameters are adjusted to achieve the best effect, to ensure the validity of the comparative experiment. AP is used as evaluation criteria of detection results, which is the most reliable and commonly used evaluation criteria in current object detection field. And APs of each framework are obtained when IOU is 0.5. As shown in Table 6, our network achieves 54.7% AP, which have an improvement of at least 5.5% in AP over the other frameworks. With the second data augmentation strategy and Soft-NMS, the AP of WrisNet can reach to 56.6%.

4.4.3 Ablation experiment

A simple ablation experiment is performed and the results are shown in Table 7, where the significant improvement of our method is marked in bold. The impact of the proposed data preprocessing, the proposed backbone network, the data augmentation, and the proposed post-processing are gradually tested. In ablation experiments, the results can demonstrate that the proposed WrisNet have obvious AP improvement up to 8.6%. As shown in Table 8, the improvement is mainly due to the enhancement of small target detection and WrisNet have obvious AP improvement of small targets up to 9.4%.

Table 7 Ablation experiment

Data preprocessing	Improved backbone	Data augmentation (10×)	Soft-NMS	AP (%)
×	×	×	×	48.0
✓	×	×	×	49.2
✓	✓	×	×	53.7
✓	×	✓	×	53.3
✓	✓	✓	×	54.0
✓	✓	✓	✓	**56.6**

Table 8 Comparison of AP of different size targets

Algorithm	AP%	AP% (small targets)	AP% (medium targets)
Faster R-CNN(ResNeXt101)	48.0	27.8	67.6
WrisNet	**56.6**	**37.2**	73.4

5 Conclusion

In this paper, an automatic GAN-based preprocessing and WrisNet are proposed for X-ray diagnosis of wrist and finger fractures. The results between the proposed GAN and manual processing show high similarity for X-ray enhancement, as a generator incorporating an attention mechanism is designed. 93% of SSIM indicates that manual window augmentation can be replaced by automatic GAN-based preprocessing. The preprocessed images are fed into WrisNet for wrist and finger fracture detection. To better handle hairline fractures, ResNeXt-TA and Soft-NMS were used to improve the backbone and post-processing. ResNeXt-TA is constructed by using group convolution and attention strategy to extract richer feature maps, while Soft-NMS is used to filter redundant bounding boxes. The AP value of the proposed method improves by 7% compared to that of the current mainstream framework, when the IOU threshold is 0.5. We believe that WrisNet performs better after being trained on a large number of data and has the potential to help doctors in diagnosis.

Acknowledgements This work was supported in part by the Key R&D Project of Shandong Province under Grant No. 2022CXGC010503, the Youth Foundation of Shandong Province under Grant No. ZR202102230323, the National Natural Science Foundation for Young Scientists of China under Grant No. 61903155, and the Doctoral Scientific Fund Project under Grant No. xbs1910.

Funding Information Not applicable.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Abdou MA (2022) Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput & Applic 34:5791–5812. https://doi.org/10.1007/s00521-022-06960-9
2. Zhao ZQ, Zheng P, Xu ST, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30(11):3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865
3. Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, Lan X (2020) A review of object detection based on deep learning. Multimed Tools Appl 79(33):23729–23791. https://doi.org/10.1007/s11042-020-08976-6
4. Ren M, Paul HY (2022) Deep learning detection of subtle fractures using staged algorithms to mimic radiologist search pattern. Skeletal Radiol 51:345–353. https://doi.org/10.1007/s00256-021-03739-2
5. Mourya GK, Gogoi M, Talbar SN, Dutande PV, Baid U (2021) Cascaded dilated deep residual network for volumetric liver segmentation from CT image. Int J E-Health Med Commun 12(1):34–45. https://doi.org/10.4018/IJEHMC.2021010103
6. Liu Y, Sun P, Wergeles N, Shang Y (2021) A survey and performance evaluation of deep learning methods for small object detection. Expert Syst Appl 172:114602. https://doi.org/10.1016/j.eswa.2021.114602
7. Lim JS, Astrid M, Yoon H J, Lee SI (2021) Small object detection using context and attention. In: 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, pp 181–186. https://doi.org/10.1109/ICAIIC51459.2021.9415217
8. Singh NK, Raza K (2021) Medical image generation using generative adversarial networks: a review. Health Inform Comput Perspect Healthc 932:77–96. https://doi.org/10.1007/978-981-15-9735-0_5
9. Chen X, Li Y, Yao L, Adeli E, Zhang Y (2021) Generative adversarial U-Net for domain-free medical image augmentation. arXiv preprint arXiv:2101.04793
10. Morís DI, de Moura RJJ, Buján JN, Hortas MO (2021) Data augmentation approaches using cycle-consistent adversarial networks for improving COVID-19 screening in portable chest X-ray images. Expert Syst Appl 185:115681. https://doi.org/10.1016/j.eswa.2021.115681
11. Han C (2021) Pathology-aware generative adversarial networks for medical image augmentation. arXiv preprint arXiv:2106.01915
12. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784
13. Kim HJ, Lee D (2020) Image denoising with conditional generative adversarial networks (CGAN) in low dose chest images. Nucl Instrum Methods Phys Res Sect A 954:161914. https://doi.org/10.1016/j.nima.2019.02.041
14. Zhu Y, Zhou Z, Liao G, Yuan K (2020) Csrgan: medical image super-resolution using a generative adversarial network. In: 2020 IEEE 17th international symposium on biomedical imaging workshops (ISBI workshops). IEEE, pp 1–4. https://doi.org/10.1109/ISBIWorkshops50223.2020.9153436
15. Gu Y, Zeng Z, Chen H, Wei J, Zhang Y, Chen B et al (2020) MedSRGAN: medical images super-resolution using generative adversarial networks. Multimed Tools Appl 79:21815–21840. https://doi.org/10.1007/s11042-020-08980-w
16. Jiang X, Liu M, Zhao F, Liu X, Zhou H (2020) A novel super-resolution CT image reconstruction via semi-supervised generative adversarial network. Neural Comput Appl 32:14563–14578. https://doi.org/10.1007/s00521-020-04905-8
17. Guan B, Zhang G, Yao J, Wang X, Wang M (2020) Arm fracture detection in X-rays based on improved deep convolutional neural network. Comput Electr Eng 81:106530. https://doi.org/10.1016/j.compeleceng.2019.106530
18. Qi Y, Zhao J, Shi Y, Zuo G, Zhang H et al (2020) Ground truth annotated femoral X-ray image dataset and object detection based method for fracture types classification. IEEE Access 8:189436–189444. https://doi.org/10.1109/ACCESS.2020.3029039
19. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
20. Guan B, Yao J, Zhang G, Wang X (2019) Thigh fracture detection using deep learning method based on new dilated convolutional feature pyramid network. Pattern Recognit Lett 125:521–526. https://doi.org/10.1016/j.patrec.2019.06.015
21. Sha G, Wu J, Yu B (2020) Detection of spinal fracture lesions based on improved Yolov2. In: 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, pp 235–238. https://doi.org/10.1109/ICAICA50127.2020.9182582

22. Abbas W, Adnan SM, Javid MA, Majeed F, Ahsan T, Hassan SS (2020) Lower leg bone fracture detection and classification using faster RCNN for X-rays images. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, pp 1–6. https://doi.org/10.1109/INMIC50486.2020.9318052
23. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 2818–2826. https://doi.org/10.1109/cvpr.2016.308
24. Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 1125–1134. https://doi.org/10.1109/cvpr.2017.632
25. Woo S, Park J, Lee JY, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19. https://doi.org/10.1007/978-3-030-01234-2_1
26. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 2117–2125. https://doi.org/10.1109/cvpr.2017.106
27. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning (ICML). PMLR, pp 448–456
28. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: International conference on machine learning (ICML), pp 807–814
29. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/cvpr.2016.90
30. Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 1492–1500. https://doi.org/10.1109/cvpr.2017.634
31. Misra D, Nalamada T, Arasanipalai AU, Hou Q (2021) Rotate to attend: convolutional triplet attention module. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 3139–3148. https://doi.org/10.1109/WACV48630.2021.00318
32. Niu Z, Zhong G, Yu H (2021) A review on the attention mechanism of deep learning. Neurocomputing 452:48–62. https://doi.org/10.1016/j.neucom.2021.03.091
33. Bodla N, Singh B, Chellappa R, Davis LS (2017) Soft-NMS—improving object detection with one line of code. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 5561–5569. https://doi.org/10.1109/iccv.2017.593
34. Lin TY, Maire M, Belongie S, Hays J, Perona P et al (2014) Microsoft coco: common objects in context. In: European conference on computer vision. Springer, Cham, pp 740–755. https://doi.org/10.1007/978-3-319-10602-1_48
35. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
36. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A (2021) A review of medical image data augmentation techniques for deep learning applications. J Med Imaging Radiat Oncol 65:545–563. https://doi.org/10.1111/1754-9485.13261
37. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 248–255. https://doi.org/10.1109/CVPR.2009.5206848
38. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, Cham, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
39. Cai Z, Vasconcelos N (2018) Cascade r-cnn: delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 6154–6162. https://doi.org/10.1109/cvpr.2018.00644
40. Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y (2017) Deformable convolutional networks. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 764–773. https://doi.org/10.1109/ICCV.2017.89