Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

A Dual-Selective Channel Attention Network for Osteoporosis Prediction in Computed Tomography Images of Lumbar Spine

Linyan Xue[1,2,3], Ya Hou[1], Shiwei Wang[1], Cheng Luo[4], Zhiyin Xia[4], Geng Qin[1], Shuang Liu[1,2,3], Zhongliang Wang[1], Wenshan Gao[4*], Kun Yang[1,2,3*]

[1] College of Quality and Technical Supervision, Hebei University, 071002 Baoding, China
[2] Hebei Technology Innovation Center for Lightweight of New Energy Vehicle Power System, 071002 Baoding, China
[3] National & Local Joint Engineering Research Center of Metrology Instrument and System, Hebei University, 071002 Baoding, China
[4] Department of Orthopedics, Affiliated Hospital of Hebei University, 071002 Baoding, China

[*] Correspondence: Wenshan Gao (wsg6813@163.com); Kun Yang (yangkun@hbu.edu.cn)

Received: 07-09-2022 **Revised:** 08-10-2022 **Accepted:** 09-12-2022

Citation: L. Y. Xue, Y. Hou, S. W. Wang, C. Luo, Z. Y. Xia, G. Qin, S. Liu, Z. L. Wang, W. S. Gao, and K. Yang, "A dual-selective channel attention network for osteoporosis prediction in computed tomography images of lumbar spine," *Acadlore Trans. Mach. Learn.*, vol. 1, no. 1, pp. 30-39, 2022. https://doi.org/10.56578/ataiml010105.

Abstract: Osteoporosis is a common systemic bone disease with insidious onset and low treatment efficiency. Once it occurs, it will increase bone fragility and lead to fractures. Computed tomography (CT) is a non-invasive medical examination method that can identify the bone condition of patients. In this paper, we propose a novel channel attention module, which is subsequently integrated into the supervised deep convolutional neural network (DCNN) termed DSNet, which can perform feature fusion from two different scales, and use the method of quadratic weight calculation to enhance the interconnection among feature map channels and improve the detection and classification performance for the bone condition in lumbar spine CT images. To train and test the proposed framework, we retrospectively collect 4805 CT images of 133 patients, using DXA as the gold standard. According to the T-value diagnostic criteria defined by WHO, the vertebral bodies of L1 - L4 in CT images are labeled and classified into osteoporosis, osteopenia and normal bone mineral density. Meanwhile, the training set and test set are constructed in the ratio of 4:1. As a result, the DSNet achieves a prediction accuracy of 83.4% and a recall rate of 90.0% on the test set, indicating that the proposed model has the potential to assist clinicians in diagnosing individuals with abnormal BMD and may alert patients at high risk of osteoporosis for timely treatment.

Keywords: Deep convolutional neural network (DCNN); DSNet; Dual-selective channel attention; CT image; Osteoporosis

1. Introduction

Osteoporosis is a common and systemic bone disease, and its early symptoms are not obvious. Most patients with osteoporosis undergo relevant examinations when complications arise, which are usually in the late stage and bring a lot of inconveniences and even death [1]. Therefore, early screening is crucial for the timely prevention and treatment of osteoporosis fractures [2]. Meanwhile, all kinds of orthopedic surgery need to refer to a bone status evaluation to formulate a better surgical plan. As the gold standard for bone mass measurement, DXA testing is expensive. Even in many developed countries, the opportunity to use DXA is still insufficient. CT image examination has a large number and clear images, which is of great significance in early screening for the prevention and treatment of osteoporotic fractures [3].

In order to reduce this preventable injury and subsequent complications, more and more researchers are focusing on methods that combine computer-aided detection and machine learning with radionics to assist clinicians with osteoporosis prediction. For example, Aouache et al. [4] designed a fuzzy decision tree (FDT) model to identify osteoporosis by identifying patients' cervical spine images. Devikanniga and Raj [5] proposed an artificial neural

network optimized for monarch butterflies, which identified osteoporosis and normal subjects by recognizing hip X-ray images of patients and combining them with demographic attributes. However, their algorithms were based on small data sets and required complex data processing before feature extraction, causing inaccuracy in medical image processing.

In recent years, convolutional neural networks (CNNs) have achieved high performance in visual recognition tasks. Instead, this technique can automatically locate the region of interest (ROI) and extract features, avoiding empirical errors in manual feature extraction [6]. Therefore, the application of deep learning in medical imaging diagnosis has received extensive attention for osteoporosis classification [7-10] and bone mineral density (BMD) prediction [11-13]. For instance, Pan et al. [14] developed a deep learning-based system for bone mineral density (BMD) classification in chest CT images. To extract the density information of trabecular bone more accurately, they firstly segmented and labeled all vertebral bodies, then their system can automatically extract the rectangular area in the center of the trabecular bone as the ROI, thereby realizing automatic measure BMD and make two classifications predictions of osteoporosis and osteopenia. Yasaka et al. [15] used an improved CNN model to extract the features of the manually labeled central circular region of trabecular bone from unenhanced abdominal CT images, achieving a correlation of 0.840 between the CNN network and the corresponding DXA results. Lee et al. [16] proposed a method based on the combination of the VGG network and random forest to classify of normal and abnormal BMD in spinal X-ray images. To extract features more accurately, they selected the central region of the fourth lumbar spine as ROI and achieved a 71% accuracy for the two-category classification. Gonzalez et al. [17] selected chest CT images as input data, proving that deep neural network has an excellent performance in osteoporosis identification.

Despite the above studies yielded exciting results in identifying osteoporosis, they still left significant limitations on the diagnostic ability of clinical osteoporosis recognition. Firstly, those designed network models were limited to two classifications: Osteoporosis and osteopenia, which could not meet the clinical requirements of various conditions of bone condition. Moreover, these studies mostly selected the trabecular structure as ROI, while ignoring the cortical bone thickness which plays an important role in the real discrimination scenarios of bone by clinical experts. Finally, the methods proposed in these studies were not effective in bone image texture recognition due to the insufficient ability of feature extraction, which affected the prediction accuracy.

In view of the limitations of CNN-based methods for accurate osteoporosis prediction, we proposed an improved network based on Faster R-CNN, in which a new channel attention module was integrated to enhance the texture, shape and other features of vertebral trabecular bone in CT images. The specific analysis is as follows: 1) We have achieved a three-category classification of osteoporosis, osteopenia and normal bone mineral density, which brings greater applicability to professional doctors in clinical diagnosis; 2) According to the actual clinical needs, the whole vertebral body including cortical bone and cancellous bone was tested for the region of interest; and 3) We adopt an improved channel attention mechanism, named DS module, to improve the interrelation among feature map channels and deepen the learning of extracted feature information.

2. Methodology

This paper aims to improve the osteoporosis detection performance for the lumbar spine CT images based on an improved Faster R-CNN network.

2.1 Faster R-CNN

The general framework of the proposed DSNet is based on the mainstream detection network of Faster R-CNN due to the high performance in many visual detection tasks [18]. In Faster R-CNN, the backbone network extracts feature from input images using ResNet [19], LeNet5 [20], AlexNet [21] or GoogLeNet [22]. The backbone network loads the officially trained model parameters to extract features, which can reduce the amount of model training data and speed up the training speed. Afterward, the target proposal boxes generated by the region proposal network (RPN) are projected to the feature map through the region of interest pooling layer (ROI pooling). Finally, the features are calculated by classification and bounding regression to achieve end-end detection. The framework of Faster R-CNN is shown in Figure 1.

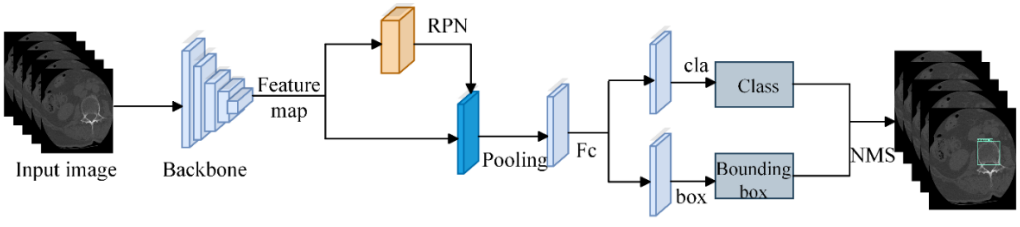

Figure 1. Faster R-CNN

In clinical practice, radiologists commonly distinguish normal, osteoporosis, and osteopenia BMD from the thickness of cortical bone and the sparsity of cancellous bone. For the automatic osteoporosis detection task, the network needs to pay more attention to the texture details of the whole vertebral body. With respect to these aspects, we pronounce a novel attention module and subsequently integrate it into the backbone of ResNet50 to improve the interconnection among feature map channels and deepen the learning of the extracted texture feature information. As a result, the proposed DSNet extracts more texture features of the entire vertebral body than the traditional ResNet50.

2.2 Channel Attention Module

Channel attention is one of the most widely used attention mechanisms in various fields such as natural language processing (NLP), computer vision (CV) and speech recognition. The representative ones are SENet [23], SKNet [24], ECANet [25], etc. As channel attention modules, they can be easily embedded in the deep learning network to achieve improved performance. As shown in subgraph (a) of Figure 2, the input feature map with the size of H*W*C is mapped to the feature map with the size of H1*W1*C1 through a transformation of a 3*3 convolution. Then the SE attention mechanism establishes the inter-dependence between feature channels through a global pooling and a full convolution. Its mechanism is that each channel's feature map is assigned a weight, which represents the correlation between the channel and the key information. The larger the weight, the higher the correlation. In this way, it simulates the brain signal processing mechanism of human vision and filters useful channel information through the weight. The SK attention mechanism introduces a 5*5 spatial dimension on the basis of SE to fuse feature channels, as shown in subgraph (b) of Figure 2. The fusion feature map from two different scales captures attention through a global pooling and a full convolution. Then it assigns weights by introducing a softmax calculation to select adaptively different spatial scales of information without sacrificing the amount of computation. Although SKNet captures target features at different scales, it fuses feature maps before weight assignment and allocates the same attention weight to the feature map after the global pooling and the full convolution, thereby losing the interactive feature information from different scales and not adequately condensing the model's attention ability.

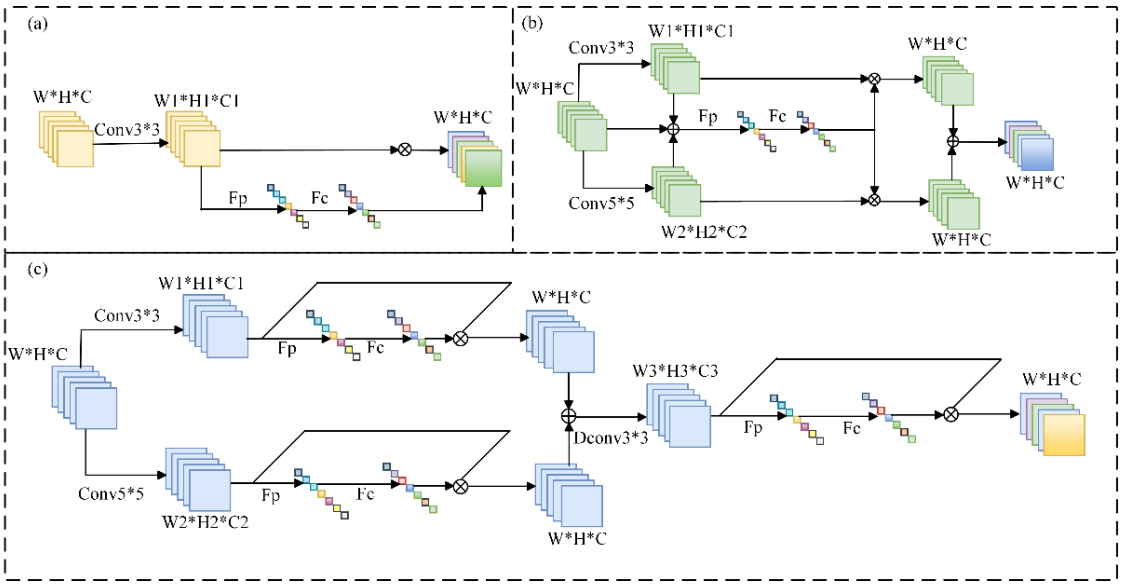

Figure 2. Schema of different attention modules

2.3 The Improved Dual-Selective Attention Module

This paper proposes a new attention mechanism module named dual-selective, as shown in subgraph (c) of Figure 2. The feature map extracted by Backbone is taken as the input. Our attention mechanism module performs feature fusion from two different scales of kernel sizes 3 and 5. The global pooling layer and the full convolution layer are used to fuse the weights of each channel. Then the two fused branches are superimposed. In order to adaptively adjust the multi-scale feature information and capture more accurate information from the target objects with different scales, we adopt a dilated convolution with a 3×3 kernel and a dilation size of 2 to fuse the output features. Therefore, our DS module can fuse more resolution information in the convolutional feature map from different attention weights, and deepen the texture information of the feature map from different scales by the

secondary weight calculation. It is conducive to paying more accurate attention to the feature information of the image texture.

In subgraph (c) of Figure 2, the computational steps can be mathematically expressed as follows. The input of the channel attention block is a feature map $x \in R^{W*H*C}$, in which W, H and C are the width, the height and the channel, respectively [26]. The feature map x is conducted with two transformations with kernel sizes 3 and 5, whose results are represented as x_c. The x_c is embed the global information by a global pooling. The global pooling $F_P(x_c)$ is calculated as:

$$F_P(x_c) = \frac{1}{H \times W} \Sigma_{i=1}^{W} \Sigma_{j=1}^{H} x_{ij} \tag{1}$$

Further, the output feature map $x_o \epsilon R^{1*1*C}$ after the global pooling is the one-dimensional vector. The feature map x_0 achieves precise and adaptive selections by a full convolution. The full convolution $F_c(x_o)$ is calculated is:

$$F_c(x_o) = x_o W_{ij} + B_i \tag{2}$$

where, $W_{ij} \epsilon R^{i*j}$ is the weight matrix and $B \in R^{i*1}$ is the one-dimensional bias [27].

2.4 Dual-Selective Channel Attention Network

Dual-selective channel attention network (DSNet) serves as the backbone of the improved Faster R-CNN, which is a fusion network of ResNet50 and DS module. As shown in Figure 3, the first stage has a relatively simple structure, consisting of a 7*7 convolutional kernel and a 3*3 max pooling layer. The second stage consists of a DS module and a residual layer. Stages 3 and 4 consist of one residual layer, respectively. The last stage consists of a residual layer and a subsequent DS module. Each residual layer is composed of 1*1, 3*1, and 1*1 convolution kernels in sequence. In this way, DSNet improves the network's ability to extract texture features and can use transfer learning to improve the stability and training speed of the model.

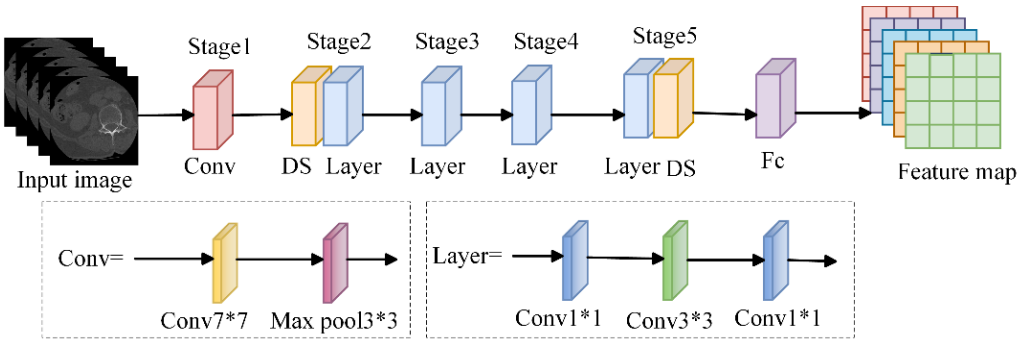

Figure 3. Pipeline of the proposed DSNet

3. Experiments

We used axial unenhanced lumbar spine CT images (L1 - L4) as input data and the DXA report of each lumbar spine as reference data. All data were manually annotated for the entire lumbar vertebrae and labeled with the bone condition category of osteoporosis, osteopenia and normal. The training set and test set were divided at a ratio of 4:1. The training set is used to train the network model. The performance of the network model is evaluated by the test set.

3.1 Subjects and Dataset Description

This study was approved by the ethical board of the affiliated hospital of Hebei university. We retrospectively collected personal data of lumbar spine CT scan (full bone window) and DXA examination between May 2016 and April 2020 from the Department of Orthopedics, Affiliated Hospital of Hebei University, with a total of 50,296 lumbar spine CT images and corresponding DXA detection results of 1,132 patients. Moreover, we conducted screening and disorientation processing for patient privacy. All patients underwent single-photon emission computed tomography (CT) scans of the lumbar spine (from L1 to L4). Their gender, height, and weight were recorded, and the original images of the cases were in DICOM format. For all data, we performed rigorous data

cleaning by learning from medical professionals: (1) Exclude images of poor quality, such as artifacts produced by spinal implants (instruments) implanted during surgery; (2) Exclude non-lumbar regions or images showing lumbar L1-L4 insufficiency, leaving only the lumbar cones Complete images; (3) Image data that has undergone lumbar spine surgery, such as bone nails and bone cement filling; (4) Individuals with a history of fractures, history of spinal surgery, primary or metastatic tumors, bone hyperplasia, or vertebral bodies are excluded. Finally, we achieved 4805 images (including any of L1, L2, L3 or L4 for the training dataset; including all L1, L2, L3 and L4 for the test dataset) of 133 patients (men, n=29, women, n=104). The whole vertebral body (including cortical bone and cancellous bone) was manually annotated by experienced physicians.

According to the criteria for diagnosis of osteoporosis defined by WHO [28]. We divided the subjects into three groups, namely the normal group with T value ≥ -1.0 SD, the osteopenia group with -2.5 SD < T value < -1.0 SD, and the osteoporosis group with T value ≤ -2.5 SD. Finally, we divide all the data at a ratio of 4:1, one part is used for the training of network parameters, namely the training set, which includes 3,844 CT images, and the other part is used to evaluate the generalization performance of the network, namely the test set, which includes 961 CT images.

3.2 Experimental Details

A new convolutional neural network should be firstly initialized the weights. Otherwise, the activation layer function will fail to output during the training of deep neural networks, which will result in an explosion of the loss gradient and a prolonged convergence of the network. Therefore, we use the weight parameters learned from the COCO dataset as the initial network parameters, which can avoid effectively the problem of overfitting during the training process and improve the stability and training speed of the model.

Deep learning model training was performed on a computer equipped with a core I7-7800x 3.5-GHz central processing unit, 256 GB memory, and a 2080 Ti graphics processing unit. We used the programming language of python 3.7 and the deep learning framework of PyTorch. The loss function of the model was the dice loss, SGD was used as the optimizer of the model, the learning rate was set to 0.01, the batch size was 9, and the training epoch was 100.

3.3 Evaluation Indicators

To evaluate the performance of the proposed model, the metrics of mean average precision (Map), recall, accuracy, and frame per second (Fps) are adopted for object detection. The vertebral bodies of the lumbar CT images (IOU ≥ 0.5) are detected and classified correctly as correct predictions, otherwise, it is considered incorrect predictions.

Mean average precision (Map) is the proportion of the predictions for all categories that successfully predicted the true target, which is defined as:

$$Map = \frac{TP}{N(TP + FP)} \tag{3}$$

Recall rate (Recall) represents the ability of the model to find all relevant targets, which is the number of true targets in the results predicted by the model. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Accuracy (Acc) represents the proportion of all predictions that the model predicts correctly.

$$Acc = \frac{TP}{TP + TN + FN} \tag{5}$$

Frame per second (Fps) is defined as the number of pictures processed by the model per second. The higher the frame per second, the faster the model processing speed, and vice versa.

$$Fps = \frac{Fn}{S} \tag{6}$$

where, *TP* is the number of predicted bounding boxes with correct classification and correct bounding box coordinates. *FP* is the number of predicted bounding box classification errors or bounding box coordinates that do

not meet the standard. *FN* is the number of ground truths falsely detected. Number (*N*) represents the number of categories for classification. *Fn* is the number of picture frames processed by the network. *S* represents the time.

4. Results

In this study, we use 1,605 CT images of osteoporosis, 1,600 CT images of osteopenia, and 1,600 CT images of normal bone density for analysis. To estimate the detection performance of the proposed attention mechanism and the novel DSNet, we compared the effects of ResNet50, VGG16, and MobileNet as the backbone under different attention mechanisms.

4.1 Comparison of Different Attention Modules Based on ResNet50

We adopted Faster R-CNN as the main framework and used the backbone with different attentional mechanisms to extract features. We performed an ablation experiment using the modules of SK, SE, and our DS based on the traditional ResNet50 without the integration of the channel attention algorithm. The metrics of Map, Recall, Acc with the IOU of 0.5, and Fps were used to evaluate and compare the detection performances among different methods. As shown in Table 1, ResNet50 integrated with the attention module achieved increased Map and Recall, indicating that the attention module can improve the network's performance. The integration of DS attention into ResNet50 achieved the best performance and the detection speed is the closest to the ResNet50, reaching a Map of 83.4%, a Recall of 90.0%, an Acc of 75.9%, and a Fps of 0.1846, indicating that our DS module has better performance on network feature extraction. Meanwhile, our DS module improved the shortcomings of other attention mechanisms on Acc, with a 1.8% improvement over Resnet50. Therefore, our model was effective at fusing multiscale attention weights.

Table 1. Detection results of the resnet50 with different attention modules

Method	Backbone	Map (%)	Recall (%)	Acc (%)	Fps (f/s)
Faster R-CNN	ResNet50	78.2	88.1	74.1	**0.1849**
	ResNet50+SE	80.4	88.5	73.7	0.1834
	ResNet50+SK	81.4	89.2	73.8	0.1829
	ResNet50+DS (ours)	**83.4**	**90.0**	**75.9**	0.1846

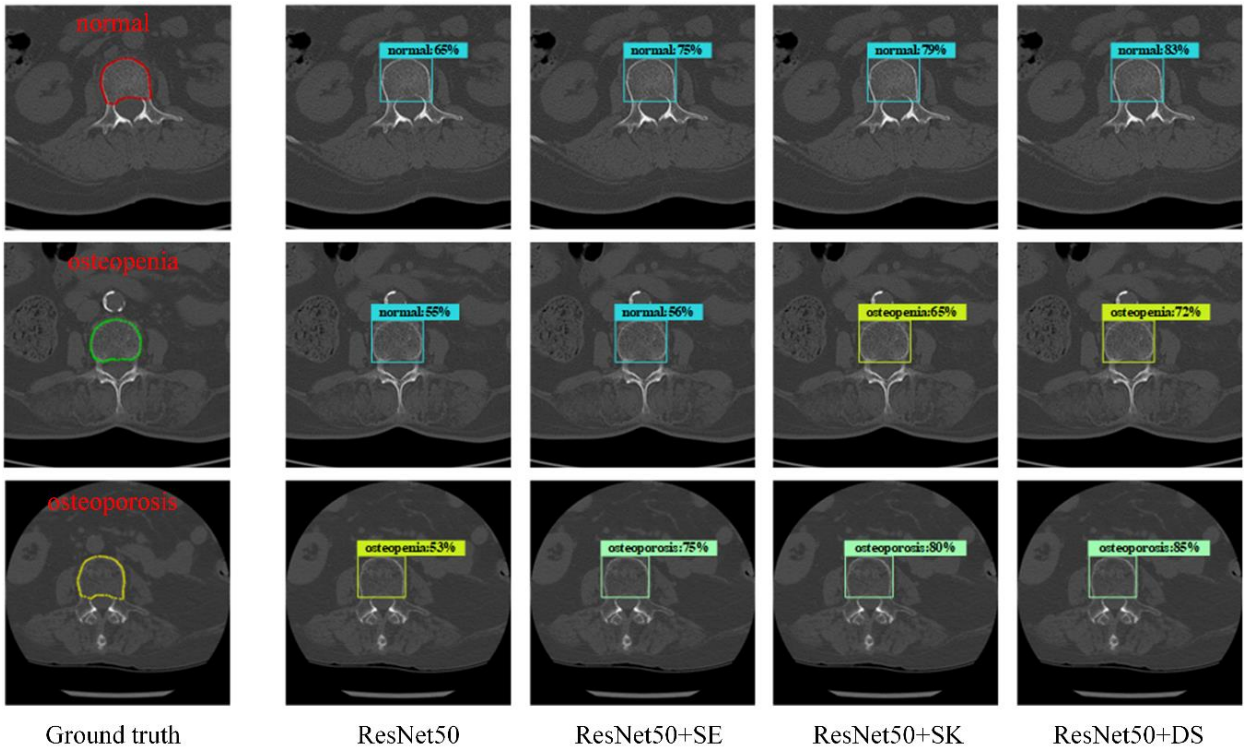

Figure 4. The visual detection results of different backbone networks

To evaluate the detection performance of the network more intuitively, we performed a visual analysis to compare the results of different modules. Figure 4 shows the detection results of different backbone networks with the vertebrae's boundary and a single probability value. The first column is the ground truth of normal bone density, osteopenia, and osteoporosis. The remaining columns are the visual detection results of different backbone networks, followed by Resnet50, Resnet50+SE, Resnet50+SK, and Resnet50+DS. All backbone networks accurately identified the normal class. But because the texture feature of osteopenia is more difficult to distinguish than another two classes, the Resnet50 and Resnet50+SE mistakenly identified osteopenia as normal. Meanwhile, the Resnet50 mistakenly identified osteoporosis as osteopenia, indicating that the attention module can improve the network's detection performance. This can seriously affect the clinician's diagnosis and delay the patient's treatment. For the three categories, our Resnet50+DS module can accurately identify with the highest confidence. It can be seen that the ResNet50+DS has a higher detection efficiency for improving the network's attention ability.

4.2 Comparison of Our Module Network with Other Backbone Networks

To verify the applicability of the proposed DS attention module, we tested the performance of different backbone networks with and without the fusion DS attention module. To be fair, these three sets of comparative experiments were tested under the same training conditions and the same data set (4,805 images of 133 patients). As shown in Table 2, ResNet50 with the DS attention module has improvements of 0.54%, 0.19%, and 0.18% in Map, Recall, and Acc with a 0.0003f/s decrease in Fps. We further compared the performance with or without the DS attention module on VGG [29], and MobileNet [30] and found that the integration of the DS attention module would also reduce their speed. The VGG16 integrated with the DS attention module achieved increased Map, Recall, and Acc. The integration of DS attention into MobileNet achieved better performance on Map and Recall with a 0.3% reduction in Acc. Among ResNet50 with the integration DS module has the best improvement. In general, our proposed DS module is also applicable to other backbones and the fusion method of ResNet50 had the best effect, which can effectively improve the feature extraction ability of the network.

Table 2. Detection results of the faster R-CNN with different backbones

Method	Backbone	DS	Map (%)	Recall (%)	Acc (%)	Fps (f/s)
Faster R-CNN	ResNet50		78.2	88.1	74.1	**0.1849**
		√	**83.4**	**90.0**	**75.9**	0.1846
	VGG16		80.0	88.5	71.1	**0.1873**
		√	**80.2**	**89.9**	**71.8**	0.1870
	MobileNet		80.9	86.4	**73.0**	**0.1856**
		√	**83.0**	**88.1**	72.7	0.1846

4.3 Comparison of Our Method with Other Methods

Table 3 illustrated the comparative results between our improved network method and the other osteoporosis prediction networks. Our method produced remarkable recall and detection categories on the biggest number of images. The traditional osteoporosis prediction methods had the results of the two classifications. Our method added a detection category of osteopenia to alert patients before they reach the severity of osteoporosis. As we all know, as the detection category increases, the detection effect of the network decreases. Therefore, the Acc of our proposed method was reduced than the other two-classification tasks. But our proposed method achieved the highest Recall of 89.2% with a tri-classification identification of osteoporosis, osteopenia, and normal. (We have no the results of AUC due to the tri-classification detection task.) Thus, the overall detection and classification effect of our method achieved better performance and provide value for the prediction of three categories of bone status in the future [31].

Table 3. Comparative result of the proposed method with other methods

Methods	Number of images	Category	Acc (%)	AUC (%)	Recall (%)
Machine learning [32]	120	2	-	83.0	-
LeNet [33]	4000	2	88.4	-	-
3D U-net [14]	200	2	-	**92.7**	85.7
VGGnet16+BCR [16]	334	2	71.0	74.0	-
ENSEMBLE [31]	247	2	**92.0**	-	88.0
Faster R-CNN+ResNet50+DS (Ours)	**4805**	**3**	75.9	-	**90.0**

5. Conclusions

This study proposes a new deep-learning method for the automatic detection of lumbar vertebrae and classification of bone status using 4,805 lumbar CT images of 133 patients. We train the network model with a large amount of clinical data and compare the test results with DXA diagnosis results. Meanwhile, the network model has been improved and optimized. At present, the prediction of osteoporosis is based on the measurement of bone density. However, bone density can only reflect about 70% of the degree of osteoporosis. The geometric characteristics of bone microstructure and the heterogeneity of density structure also have a certain impact on the diagnosis of osteoporosis. Therefore, the thickness of cortical bone and the thinning of cancellous bone are the basis for doctors to judge whether there is osteoporosis. Different from previous studies, our model is highly close to the doctor's diagnostic level. The entire lumbar vertebral body including cortical and cancellous bones is annotated as regions of interest, as the dataset for model training.

In this paper, we propose a new backbone integrating ResNet50 and DS module based on Faster R-CNN. In this network, the lumbar vertebral body is identified and detected first, and the detection results are marked in the form of rectangular boxes with predicted confidence. Then the vertebral body structure is extracted to class the categories of osteoporosis, osteopenia, and normal. The proposed module improves the performance of feature extraction and pays more attention to the texture features of cortical and cancellous bones. Compared with other attention mechanisms, our module achieves an improvement on Map, Recall, and Acc reaching 83.4% and 90.0%, and 75.9% respectively. The feasibility, effectiveness, and compatibility of the model are also verified.

This paper can expand the application of artificial intelligence-assisted diagnosis systems and contribute to the clinical identification of osteoporosis. To a certain extent, it can solve the problem that fracture osteoporosis cannot be detected in time to improve the treatment rate of osteoporosis so that it has important theoretical value and clinical significance.

Author Contributions

L.X. and Y.H. devised the project and drafted the manuscript; G.Q. and S.W. carried out the data collection and analyses; C.L., Z.X. and S.L. participated in the design of the study and Z.W., W.G. and K.Y. contributed to analyzing the results of the experiment. All authors have read and agreed to the published version of the manuscript.

Funding

This paper was funded by Hebei University (Grant No.: DXK201914); the President of Hebei University (Grant No.: XZJJ201914); the Post-graduate's Innovation Fund Project of Hebei University (Grant No.: HBU2022SS003); and the Special Project for Cultivating College Students' Scientific and Technological Innovation Ability in Hebei Province (Grant No.: 22E50041D).

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

[1] Y. Zhao, T. Zhao, S. Chen, X. Zhang, M. S. Sosa, J. Liu, X. Mo, X. Chen, M. Huang, S. Li, X. Zhang, and C. Huang, "Fully automated radiomic screening pipeline for osteoporosis and abnormal bone density with a deep learning-based segmentation using a short lumbar mDixon sequence," *Quant. Imaging. Med. Surg.*, vol. 12, no. 2, pp. 1198-1213, 2022. https://doi.org/10.21037/qims-21-587.

[2] P. Snodgrass, A. Zou, U. Gruntmanis, and I. Gitajn, "Osteoporosis diagnosis, management, and referral practice after fragility fractures," *Curr. Osteoporos. Rep.*, vol. 20, no. 3, pp. 163-169, 2022. https://doi.org/10.1007/s11914-022-00730-1.

[3] J. Liu, J. Wang, W. Ruan, C. Lin, and D. Chen, "Diagnostic and gradation model of osteoporosis based on improved deep u-net network," *J. Med. Syst.*, vol. 44, no. 1, pp. 15-23, 2020. https://doi.org/10.1007/s10916-019-1502-3.

[4] M. Aouache, A. Hussain, M. A. Zulkifley, D. W. W. Zaki, H. Husain, and H. B. A. Hamid, "Anterior osteoporosis classification in cervical vertebrae using fuzzy decision tree," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 4011-4045, 2018. https://doi.org/10.1109/MWSCAS.2003.1562304.

[5] D. Devikanniga and J. S. R. Raj, "Classification of osteoporosis by artificial neural network based on monarch butterfly optimisation algorithm," *Healthcare Technol Lett.*, vol. 5, no. 2, pp. 70-75, 2018. https://doi.org/10.1049/htl.2017.0059.

[6] V. Pedoia, F. Caliva, G. Kazakia, A. Burghardt, and S. Majumdar, "Augmenting osteoporosis imaging with machine learning," *Curr. Osteoporos. Rep.*, vol. 19, no. 6, pp. 699-709, 2021. https://doi.org/10.1007/s11914-021-00701-y.

[7] Y. Geng, T. Liu, Y. Ding, W. Liu, J. Ye, L. Hu, and L. Ruan, "Deep learning-based self-efficacy X-ray images in the evaluation of Rheumatoid Arthritis combined with osteoporosis nursing," *Sci. Programming-Neth.*, vol. 2021, pp. 1-8, 2021. https://doi.org/10.1155/2021/9959617.

[8] K. S. Lee, S. K. Jung, J. J. Ryu, S. W. Shin, and J. Choi, "Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs," *Jpn. J. Clin. Med.*, vol. 9, no. 2, pp. 392-400, 2020. https://doi.org/10.3390/jcm9020392.

[9] K. Poole, D. Chappell, E. Clark, J. Fleming, L. Shepstone, T. Turmezei, A. Wagner, K. Willoughby, and S. Kaptoge, "PHOENIX (Picking up hidden osteoporosis effectively during normal CT imaging without additional X-rays): Protocol for a randomised, multicentre feasibility study," *BMJ Open.*, vol. 12, no. 1, pp. 1-8, 2022. https://doi.org/10.1136/bmjopen-2021-050343.

[10] C. Tang, W. Zhang, H. Li, L. Li, Z. Li, A. Cai, L. Wang, D. Shi, and B. Yan, "CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening," *Osteoporosis Int.*, vol. 32, no. 5, pp. 971-979, 2021. https://doi.org/10.1007/s00198-020-05673-w.

[11] P. Xiao, T. Zhang, N. Dong, Y. Han, Y. Huang, and X. Wang, "Prediction of trabecular bone architectural features by deep learning models using simulated DXA images," *Bone Rep.*, vol. 13, pp. 1-8, 2020. https://doi.org/ 10.1016/j.bonr.2020.100295.

[12] H. K. Lim, H. il Ha, S. Y. Park, and K. Lee, "Comparison of the diagnostic performance of CT Hounsfield unit histogram analysis and dual-energy X-ray absorptiometry in predicting osteoporosis of the femur," *Eur. Radiolo.*, vol. 29, no. 4, pp. 1831-1840, 2019. https://doi.org/10.1007/s00330-018-5728-0.

[13] T. Ho-Le, J. Eisman, T. Nguyen, and H. Nguyen, "Prediction of hip fracture in post-menopausal women using artificial neural network approach," In Annual International Conference of the IEEE Engineering in Medicine and Biology Society, (EMBC 2017), Jeju, South Korea, July 11-15, 2017, IEEE, pp. 4207-4210. https://doi.org/10.1109/EMBC.2017.8037784.

[14] Y. Pan, D. Shi, H. Wang, T. Chen, D. Cui, X. Cheng, and Y. Lu, "Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening," *Eur. Radiolo.*, vol. 30, no. 7, pp. 4107-4116, 2020. https://doi.org/10.1007/s00330-020-06679-y.

[15] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, and O. Abe, "Prediction of bone mineral density from computed tomography: Application of deep learning with a convolutional neural network," *Eur. Radiolo.*, vol. 30, no. 6, pp. 3549-3557, 2020. https://doi.org/10.1007/s00330-020-06677-0.

[16] S. Lee, E. Choe, H. Kang, J. Yoon, and H. Kim, "The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population," *Skeletal Radiolo.*, vol. 49, no. 4, pp. 613-618, 2020. https://doi.org/10.1007/s00256-019-03342-6.

[17] G. Gonzalez, G. Washko, and R. Estepar, "Deep learning for biomarker regression: Application to osteoporosis and emphysema on chest CT scans," In Proceedings of SPIE-the International Society for Optical Engineering, (ISOE 2018), Houston, Texas, United States, March 2, 2018, SPIE, pp. 52-60. https://doi.org/10.1117/12.2293455.

[18] X. Li, Z. Xu, X. Shen, Y. Zhou, B. Xiao, and T. Li, "Detection of cervical cancer cells in whole slide images using deformable and global context aware faster RCNN-FPN," *Curr. Oncolo.*, vol. 28, no. 5, pp. 3585-3601, 2021. https://doi.org/10.3390/curroncol28050307.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In 2016 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR 2016), Las Vegas, NV, USA, June 27-30, 2016, IEEE, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90.

[20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *P. IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. https://doi.org/10.1109/5.726791.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM.*, vol. 60, no. 6, pp. 84-90, 2017. https://doi.org/10.1145/3065386.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, "Going deeper with convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (CVPR 2015), Boston, MA, June 07-12, 2015, IEEE, pp. 1-9. https://doi.org/10.1109/CVPR.2015.7298594.

[23] J. Hu, L. Shen, G. Sun, and E. H. Wu, "Squeeze-and-excitation networks," *IEEE T. Pattern Anal.*, vol. 42, no. 8, pp. 2011-2023, 2019. https://doi.org/10.1109/TPAMI.2019.2913372.

[24] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2019), Beach, CA, USA, June 15-20, 2019, IEEE, pp. 510-519. https://doi.org/10.1109/CVPR.2019.00060.

[25] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2020), Seattle, WA, USA, June 13-19, 2020, IEEE, pp. 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155.

[26] M. Lin, Q. Chen, and S. Yan, "Network in network," *ArXiv*., vol. 1, 2014. https://doi.org/10.48550/arXiv.1312.4400.

[27] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (CVPR 2018), Salt Lake City, UT, USA, June 18-23, 2018, IEEE, pp. 7794-7803. https://doi.org/10.1109/CVPR.2018.00813.

[28] L. Liu, M. Si, H. Ma, M. Cong, Q. Xu, Q. Sun, W. Wu, C. Wang, M. Fagan, L. Mur, Q. Yang, and B. Ji, "A hierarchical opportunistic screening model for osteoporosis using machine learning applied to clinical data and CT images," *BMC Bioinformatics*, vol. 23, no. 1, pp. 1-10, 2022. https://doi.org/10.1186/s12859-022-04596-z.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ArXiv.*, vol. 1, 2014. https://doi.org/10.48550/arXiv.1409.1556.

[30] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *ArXiv*., vol. 2017, 2017. https://doi.org/10.48550/arXiv.1704.04861.

[31] M. Islam, M. A. Aowal, and A. T. Minhaz, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," *ArXiv*., vol. 1, 2017. https://doi.org/10.48550/arXiv.1705.09850.

[32] U. Muehlematter, M. Mannil, A. Becker, K. Vokinger, T. Finkenstädt, G. Osterhoff, M. Fischer, and R. Guggenberger, "Vertebral body insufficiency fractures: Detection of vertebrae at risk on standard CT images using texture analysis and machine learning," *Eur. Radiol*., vol. 29, no. 5, pp. 2207-2217, 2019. https://doi.org/10.1007/s00330-018-5846-8.

[33] N. Tecle, J. Teitel, M. R. Morris, N. Sani, D. Mitten, and W. C. Hammert, "Convolutional neural network for second metacarpal radiographic osteoporosis screening," *J. Hand Surg.*, vol. 45, no. 3, pp. 175-181, 2020. https://doi.org/ 10.1016/j.jhsa.2019.11.019.