Research on rib fracture auxiliary diagnosis based on convolutional neural network

He Li[1], RuiJing Xu[1], Jinwei Li[2*]

[1] Huazhong University of Science and Technology Tongji Medical College Affiliated Union Hospital

[2]Dongfeng Motor Group Co., Ltd

ABSTRACT. Rib fractures are a common condition in chest injuries, which may be caused by traffic accidents, falls from heights, and daily exercise. Traditional detection methods such as CT scanning are prone to high rates of missed and false detections due to the complexity of rib fractures. However, detection methods based on deep learning technology have better detection performance and speed advantages. This article uses a network model called YOLOv12+Mamba for rib fracture detection, and compares it with other common detection algorithms through experiments. The test results show that the YOLOv12+Mamba network significantly improves the average accuracy (mAP_0.5), and its accuracy and recall also exceed other models, proving the effectiveness of this network. On the basis of the YOLOv12 model, this article has improved the problems and developed a new Mamba model. The outstanding performance of this model in image recognition comes from its efficient object detection and classification capabilities, as well as its support for faster training and inference processes, resulting in outstanding performance when processing large-scale image data. The improvement plan increases mAP_0.5 from 0.8612 to 0.9354, mIoU by 0.41% -1.2%, and inference speed by only 3%. Compared with the initial model, the improved prediction results are significantly better. The YOLOv12+Mamba network proposed in this article significantly improves the accuracy of rib fracture detection and effectively reduces the workload of doctors. This model provides a satisfactory solution to the problems of missed and false detections in traditional detection methods, improving diagnostic efficiency and accuracy.

Keywords: rib fracture; Deep learning; YOLOv12 ; Mamba model

* Corresponding Author: Jinwei Li (kjb-lijw@dfmc.com.cn)

1 Introduction

Rib fracture is a common chest injury [1]. Although rib fractures can usually heal on their own, severe fractures can affect lung function and lead to complications such as pneumothorax, lung contusion, or infection [2]. Therefore, timely and appropriate handling of such injuries is very important. Understanding and managing rib fractures is crucial, especially in trauma care and emergency medicine [3]. Research and development of new technologies to assist in the diagnosis and treatment of rib fractures can improve treatment outcomes and patients' quality of life. Traditional medical imaging analysis relies heavily on computed tomography (CT), however, due to the large number of image slices contained in CT data, it poses difficulties for physicians to manually review the images [4]. In addition, the complexity of rib fractures and their similarity to other chest injuries result in high rates of missed and false detections in traditional detection methods. The rib fracture detection method based on deep learning can overcome these problems and has good detection performance and fast speed [5]. This article will investigate a rib fracture detection algorithm that combines YOLOv12 with Mamba model architecture optimization [6]. Deep learning technology has made significant progress in multiple fields, especially in the medical field [7], with enormous potential for application. With the powerful computing power of computers, deep learning can assist doctors in accurately detecting lesions [8], greatly reducing the visual burden on doctors in medical image analysis. Through automation and efficient processing mechanisms [9], deep learning not only improves diagnostic efficiency and reduces operating costs, but also significantly improves the supply-demand imbalance in the healthcare industry. Therefore, applying deep learning technology to the medical field [10] can have significant implications for alleviating the shortage of medical resources.

2. Methodology

2.1 ADD MSAA（Multi-Scale Attention Aggregation) module

In the field of deep learning, the size of the dataset is often an important challenge for training models. In the YOLOv12+Mamba network model, the MSAA (Multi Scale Attention Aggregation) module efficiently aggregates feature maps of different scales through Mamba to improve segmentation accuracy. The MSAA module plays an important

role in enhancing detail segmentation capabilities in complex scenes, providing more global and rich information. However, due to the large size of the dataset, the computational load of the MSAA module also increases accordingly, resulting in a significant increase in training time. In response to this issue, this section focuses on optimizing the MSAA module in the model. First, fuse the feature maps from different stages of the encoder to form a new feature map ˆ F. Reduce the number of channels and dimensionality of the feature map using 1x1 convolution. Then sum up the convolution results of different kernel sizes, such as 3x3, 5x5, 7x7 convolution, to fuse features of different scales. Finally, average pooling and max pooling are used to aggregate spatial features, and non-linear transformation is performed through 7x7 convolution and sigmoid activation function. Finally, the spatial dimension of the feature map is compressed to 1x1 and subjected to global average pooling to extract global information. Generate channel attention maps using 1x1 convolution and ReLU activation function, and expand their size to match the dimensions of the input feature map. Can add the results of spatial and channel paths at the element level to obtain the final output feature map.

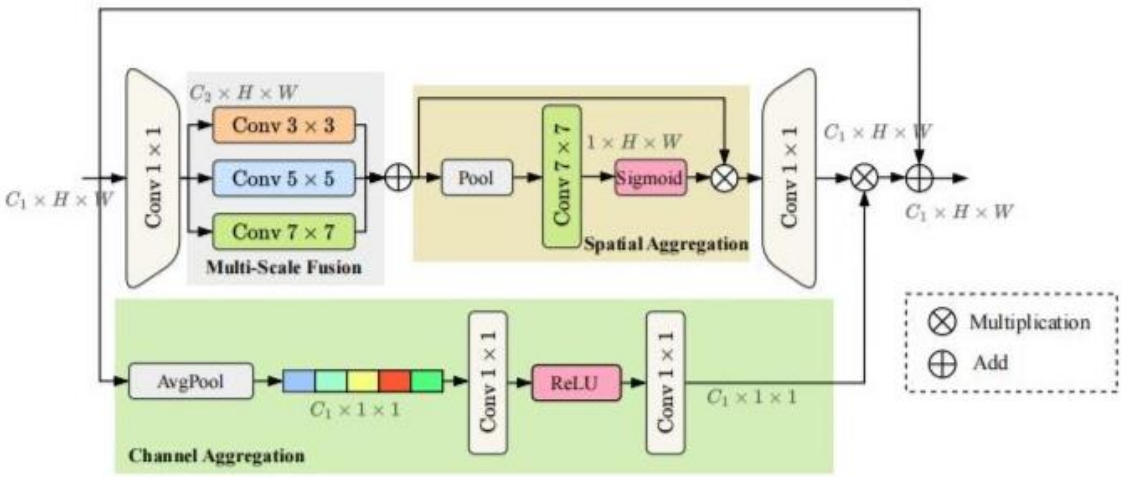

Figure 1. MSAA (Multi-Scale Attention Aggregation) module

The final pixel color is calculated by multiplying the triangle color by the coverage weight

$$Final\ Color = \frac{Triangle\ Color \times Number\ of\ covered\ sampling\ points}{total\ number\ of\ sampling\ points}$$

Parameter Description：

Triangle Color：The shading color of the current triangle at that pixel. The number of covered sampling points and the number of sub sampling points covered by the current triangle. The total number of sampling points, the preset total number of sub sampling points. The final depth value of the pixel is the average depth of the sampling points covering the lid

$$Depthpixel = \frac{1}{k}\sum_{i=1}^{k} Depthi \qquad \text{k} = \text{ Number of covered sampling points}$$

The core of MSAA lies in coverage driven color mixing and subsampling point depth averaging. The actual code implementation requires strict separation of sampling point coverage judgment and depth storage to avoid edge anomalies.

2.2 Add attention mechanism

The human visual system uses a neural mechanism called "visual attention" to quickly focus on key information in the field of view while filtering out irrelevant interference. This biological mechanism has inspired the design of attention models in deep learning, whose core idea is to automatically identify and reinforce key regions in input data by calculating feature importance weights. In the field of computer vision, this biomimetic mechanism has been successfully applied to various neural network architectures, which enhances the model's perception ability of important features by dynamically allocating computing resources, thereby significantly improving the performance of tasks such as object detection and image classification. This article will add different attention modules to the feature fusion network part of the YOLOv12+Mamba network model, and conduct experiments to compare which attention mechanism brings better promotion effect to the model. The GAM Attention in the following figure represents the attention mechanism module to be added.

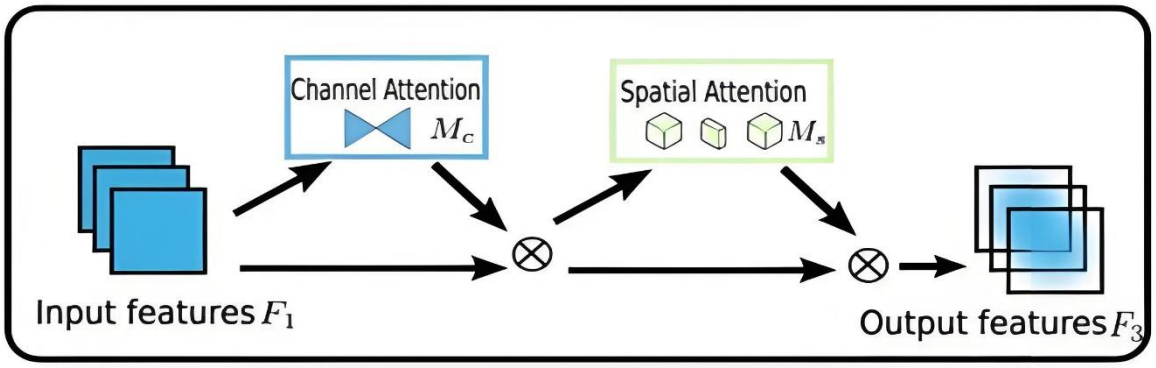

Figure 2. The GAM Attention

3. Results

3.1 Dataset collection

For this rib fracture detection task, we used the publicly available RibFrac dataset, which includes 400 patients with a total of 3321 rib fractures, including 319 male patients and 81 female patients, including 280 training CT scans (all with fractures), 40 validation CT scans (10 without fractures), and 80 test CT scans between the ages of 4 and 79. The number of slices in CT images of rib fractures for each patient ranges from 350 to 630. These data were annotated with accurate fracture sites and categories by the attending radiologist. Both the training and validation sets provided images and annotations, while the testing set only provided images. Each annotation includes a rib fracture area and a four class classification. Can meet the data requirements for rib fracture detection models. Exclude images with respiratory artifacts to prevent affecting the accuracy of the experiment, and ensure that the CT image thickness is between 1mm and 1.25mm. The dataset has consistency and stability, meeting the basic requirements for the detection algorithm experiment in this article.

3.2 experimental setup

In this chapter, two improvement schemes, MSAA module improvement and attention module addition, will be compared with the unimproved YOLOv12+Mamba network to verify their improvement effects. The attention module with the best final usage effect is used as the final improvement plan. Improvements are beneficial to the model to verify that the algorithm model proposed in this chapter has better detection performance.

The evaluation indicators used in this experiment are recall rate, accuracy, and the average accuracy mean (mAP_0.5) with an IoU threshold of 0.5. Among the three evaluation indicators, mAP_0.5 is the most important, followed by recall rate, and finally accuracy. The experimental dataset used in this chapter is a publicly available dataset, which is divided into training set, validation set, and testing set. Then input the dataset into the network model for training, and obtain the optimal weights for the model. Use this optimal weight for the experiment in the testing section. The software and hardware environment of the experiment set the batch size to 16, the number of iterations to 200, and the learning rate to 0.001.

3.3 experimental result

（1） MSAA improvement Moule

The main purpose of improving the MSAA module is to preserve small target details through spatial refinement, enhance key features through channel aggregation, and dynamically fuse across scales. This triple mechanism significantly improves the model's discriminative robustness in dense small target detection while maintaining efficient inference efficiency. It can be seen that after the improvement, mAP_0.5 and recall rate have slightly increased, while accuracy has decreased slightly and is almost unaffected. The main purpose of improving this module is for lightweighting, and it is not required to improve the detection accuracy of the model, ensuring that the detection accuracy does not decreaseModules do indeed contribute to the lightweighting of models.

Table 1. Table of indicators before and after improving MSAA module.

Evaluation	**YOLOv12+Mamba**	**After Improving The MSAA Module**
mAP_0.5	0.8924	0.8953
Precision	0.8501	0.8497
Recall	0.8943	0.9027

（2） Add attention mechanism

Adding attention mechanism, in computer vision tasks, attention mechanism is mainly applied in tasks such as image description generation, image classification, and object detection. Through attention mechanism, the model can dynamically select and focus on important areas in the image, thereby better understanding and processing the content of the

image. Thereby improving the accuracy and robustness of object detection. In this experiment, different attention mechanisms will be added to the original YOLOv12+Mamba network for comparative experiments, and the attention module with the best model promotion effect will be selected as the final improvement plan.

Table 2. Table of changes in various indicators before and after adding each attention module.

Moule	MAP_0.5	Precision	Recall
YOLOv12+Mamba	0.8931	0.8584	0.8962
YOLOv12+Mamba+GAM	0.9014	0.8643	0.8975
YOLOv12+Mamba+SA	0.8943	0.8594	0.8912

Experimental data shows that both improvement strategies proposed in this paper exhibit significant performance improvements. It is worth noting that when multiple strategy combinations are used for optimization, the model performance exhibits a synergistic enhancement effect, and the improvement is significantly better than the effect of a single improvement strategy. This fully validates the effectiveness and complementarity of each improved module, and the resulting improved YOLOv12+Mamba fusion model demonstrates excellent accuracy and clinical practical value in rib fracture detection tasks. From this, it can be seen that improving the network does indeed have a good effect on the model's prediction of rib fractures. Verified the correctness and practicality of the improvement measures proposed in this article.

4. Discussion

(1) The limitations of 3D spatial feature modeling are that the current method of training using 2D slice data suffers from spatial information loss, which may result in the model not fully exploiting the stereoscopic features of 3D CT images. It is recommended that future research adopt 3D convolutional neural networks or Transformer architectures to directly process volumetric data, in order to enhance spatial feature extraction capabilities.

(2) The improvement of clinical deployment scheme needs to establish a dual track deployment system, develop a Web visualization platform based on B/S architecture that can support DICOM standard protocols, design edge computing schemes for resource

constrained scenarios, such as carrying NVIDIA Jetson series equipment, supporting the development of lightweight clients based on QT framework, and integrate DICOM viewer and AI reasoning module.

(3) It is recommended to develop an automated 3D reconstruction pipeline for 3D visualization diagnosis support. The 2D detection results can be reconstructed into a 3D model through algorithms such as Marching Cubes, and support: 3D annotation of fracture sites, linkage of multi planar reconstruction (MPR) views, and quantitative measurement of functional fracture line length or displacement distance.

(4) The deep optimization of data imbalance problems is limited by current data augmentation strategies. Conditional Generative Adversarial Networks (cGAN) can be used for lesion region synthesis, and gradient penalty Wasserstein GAN (WGAN-GP) can be introduced to improve generation quality. Attention mechanism can be used to achieve sample generation with anatomical structure constraints.

References

[1] J J J O, Dias A, Anthony G, et al. Delayed hemothorax readmissions after rib fracture in blunt trauma patients. [J]. Journal of clinical orthopaedics and trauma, 2023, 45 102259-102259.

[2] Zhiyi L, Yejie Z. Development paradigm of artificial intelligence in China from the perspective of digital economics [J]. Journal of Chinese Economic and Business Studies, 2022, 20 (2): 207-217.

[3] Yudong Z, Jin H, Shuwen C. Medical Big Data and Artificial Intelligence for Healthcare [J]. Applied Sciences, 2023, 13 (6): 3745-3745.

[4] Ahmed M R, Zhang Y, Liu Y, et al. Single Volume Image Generator and Deep Learning-based ASD Classification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(11): 3044-3054.

[5] Lin T H, Jhang J Y, Huang C R, et al. Deep Ensemble Feature Network for Gastric Section Classification[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(1): 77-87.

[6] Zhao Y, Liu Y, Kan Y, et al. Spatial-Frequency Non-local Convolutional LSTM Network for pRCC Classification[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 22-30.

[7] Li X, Shen L, Xie X, et al. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection[J]. Artificial intelligence in medicine, 2020, 103: 101744.

[8] Li Z, Zhang S, Zhang J, et al. MVP-Net: Multi-view FPN with position-aware

attention for deep universal lesion detection[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 13-21.

[9] Tao Q, Ge Z, Cai J, et al. Improving deep lesion detection using 3d contextual and spatial attention[C]//International Conference on Medical Image Computing and Computer-AssistedIntervention. Springer, Cham, 2019: 185-193.

[10] Fan D P, Zhou T, Ji G P, et al. Inf-net: Automatic covid-19 lung infection segmentation from ct images[J]. IEEE Transactions on Medical Imaging, 2020, 39(8): 2626-2637.