Original Article

Multi-stream and multi-scale fusion rib fracture segmentation network based on UXNet

Yusi Liu, Liyuan Zhang*, Zhengang Jiang*

School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China

Contributions: (I) Conception and design: Y Liu; (II) Administrative support: L Zhang; (III) Provision of study materials or patients: Y Liu, L Zhang; (IV) Collection and assembly of data: Y Liu; (V) Data analysis and interpretation: Y Liu, L Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*These authors contributed equally to this work.

Correspondence to: Liyuan Zhang, PhD; Zhengang Jiang, PhD. School of Computer Science and Technology, Changchun University of Science and Technology, 7186 Satellite Road, Changchun 130022, China. Email: zhangliyuanzly@cust.edu.cn; jiangzhengang@cust.edu.cn.

Background: Accurate segmentation of rib fractures represents a pivotal procedure within surgical interventions. This meticulous process not only mitigates the likelihood of postoperative complications but also facilitates expedited patient recuperation. However, rib fractures in computed tomography (CT) images exhibit an uneven morphology and are not fixed in position, posing difficulties in segmenting fractures. This study aims to enhance the accuracy of elongated rib fracture segmentation, ultimately improving the efficiency of clinical diagnosis.

Methods: In this study, we propose multi-stream and multi-scale fusion network based on efficient attention UXNet (M2SUXNet). It aims to enhance the segmentation accuracy of elongated rib fractures through multi-scale fusion attention enhancement. Firstly, we propose the multi-stream and multi-scale fusion (M2SF) module in the feature extraction stage. The module is designed with two parallel paths. Each path analyzes the image content using a different feature level. Then, the module effectively distinguishes the more critical feature information in the channel according to the feature weight ratio. The M2SF module integrates information from different scales to obtain comprehensive information on global and local features, achieving a more diverse feature representation. Secondly, the efficient attention (EA) module combines different channel information of input features to integrate channel and spatial features of different channels. The module better combines the context information, establishes the dependency between the space and the channel, enhances the focusing ability of the network on the fractures of different shapes, and improves the segmentation accuracy. Thirdly, the joint loss function of BCE with Logits Loss and Dice Loss is used to solve the sample imbalance problem.

Results: We verified the effectiveness of the proposed model on the public RibFrac dataset. The experimental results demonstrated that the model achieved a Dice coefficient of 75.34%, a joint intersection over union (IoU) of 60.44%, and a precision of 93.79%.

Conclusions: The proposed model for rib fracture segmentation has higher accuracy and feasibility than other existing models. Besides, the M2SUXNet can effectively improve the segmentation performance of elongated rib fractures.

Keywords: Rib fracture segmentation; multi-scale feature fusion; attention mechanisms; UXNet

Submitted Jul 03, 2024. Accepted for publication Oct 28, 2024. Published online Dec 30, 2024.
doi: 10.21037/qims-24-1356
View this article at: https://dx.doi.org/10.21037/qims-24-1356

Introduction

Rib fracture accounts for 40–80% of all chest injuries and increases annually (1). Such fractures can lead to a range of complications (2). Computed tomography (CT) images can display the fracture area more thoroughly and carefully, which meets the requirements of clinical rib fracture diagnosis (3). Therefore, CT has become one of the most used methods for diagnosing fracture injury. A previous study showed that a certain proportion of fractures are often missed in imaging diagnostic evaluation, resulting in poor prognosis (4). Segmenting rib targets can be challenging due to their different sizes, unfixed positions, and irregular shapes. Consequently, rib fracture segmentation in CT images is still complex. Therefore, realizing accurate computer-aided diagnosis technology for rib fractures is of paramount significance for clinical treatment.

Convolutional neural network is one of the traditional image segmentation algorithms. Fully convolutional network (FCN) is a breakthrough application of deep learning (DL) in image semantic segmentation (5). It extracts image features by entire convolution operation and generates a semantically segmented image of the original size by up sampling. Similar to FCN, U-Net (6) ultimately adopts a convolutional neural network for semantic segmentation. Its structure consists of a symmetric encoder and decoder. U-Net effectively connects the encoder and decoder features by simultaneously using skip connection paths between the encoding and decoding segments, reducing the loss of feature information. Due to its performance, U-Net has become the dominant medical image segmentation method. Many researchers have developed and extended the U-Net structure. U-Net++ (7) effectively optimizes the feature map and improves the precision and accuracy of segmentation tasks by introducing dense skip connection paths based on U-Net. By adding a soft attention mechanism to the skip connection, AttUNet (8) can stop the network from reusing features from irrelevant regions and make it pay more attention to the essential features of specific local areas. ResU-Net (9) and DenseU-Net (10) replace the convolutional block of U-Net with ResNet and DenseNet, respectively, which strengthens the ability of the network to extract features and improves the convergence speed of the model. nnUnet (11) is a base robust and adaptable network framework. It obtains good segmentation results through simple preprocessing methods and reduces the complexity of the data preparation stage. UXNet (12) employs three-dimensional (3D) depth convolutions for volume segmentation tasks, which utilizes LK-sized depth convolutions as a generic feature extraction backbone and introduces pointwise depth convolutions to scale the extracted representations efficiently with fewer parameters.

Meanwhile, with the progress of artificial intelligence, many architectures inspired by U-Net have been proposed in medical image processing. Rehman *et al.* (13) and Lin *et al.* (14) respectively improved the model based on U-Net to achieve brain tumor segmentation. Zhou *et al.* (15) developed a novel M-DDC architecture based on a joint U-Net segmentation network and a deep convolutional network to implement MRI-based classification of demyelinating diseases. Ryu *et al.* (16) built Seg-Net based on U-Net and improved the accuracy of model segmentation by allowing accurate retinal vessel segmentation to be fused with dense multi-scale features.

Applying the DL image segmentation algorithm to the auxiliary diagnosis system for rib fractures can improve the accuracy and efficiency of rib fracture diagnosis. Currently, fracture diagnosis systems are classified into two categories (17). The first category of traditional fracture diagnosis models (18) usually relies on manual feature extraction. It uses rule-driven image analysis methods to diagnose suspected fracture regions. This approach combines a doctor's experience with manual intervention. In contrast, the second type of model based on DL can automatically learn features from data. It performs fracture diagnosis more efficiently and with less or even no manual intervention. This approach significantly improves the automation level and accuracy of diagnosis. Most available studies have focused on rib fracture detection in CT images (19-22), and only a few have explored rib fracture segmentation. Cao *et al.* (23) proposed a shape perception method based on DL. By utilizing contrastive learning, numerous unlabeled CT images were utilized for training the model, resulting in the accurate detection and segmentation of rib fractures. Jin *et al.* (24) proposed the FracNet model for rib fracture diagnosis and formulated the detection task as a 3D segmentation task for the first time. This method introduced sliding window sampling to generate region samples from CT images, reducing the model's computational complexity in non-rib regions. The above methods have different contributions to rib fracture segmentation. Jin *et al.* and Gao *et al.* (25) showed that accurate segmentation of slender rib fractures is more difficult than that of circular fractures. The shape of slender rib fractures is often complex, and feature information needs to be obtained from different scales. However, current research on rib fracture segmentation uses the same scale to normalize all fractures, ignoring local details. The multi-scale feature fusion strategy considers the context information of the image while better handling the local details. Therefore, we propose a novel method based on a multi-scale feature fusion strategy to accurately segment rib fractures while considering the fine details and characteristics of the fracture.

This paper proposes a new solution called M^{2}SUXNet, which addresses the difficulty of segmenting rib fractures in CT images through a multi-scale segmentation framework. First, we propose a multi-stream and multi-scale fusion (M^{2}SF) module. The M^{2}SF module integrates feature information from different scales after feature extraction. At the same time, the module can effectively integrate the information of different scales to obtain comprehensive information on global and local features. Inspired by the attention mechanism (26-28), this paper proposes an efficient attention (EA) module. The EA module combines different parts of input features to integrate and weigh features in other channels and spatial locations. It improves the ability of the model to distinguish between forms of rib fractures and suppress noise by adaptively adjusting the channel relationship in the input feature map. The dual attention mechanism allows the model to fully obtain feature information from different scales and enhance attention to subtle details. Unlike other UNet variants, M^{2}SUXNet uses the ConvNet module to tune the hierarchical transformer for robust voxel segmentation. In addition, M^{2}SUXNet can fuse multi-scale image information to achieve more diverse feature representation and fuse channel features and spatial features of different channels.

In summary, our main contributions are as follows:

(I) This paper provides an innovative multi-stream and multi-scale fusion network architecture for rib fracture segmentation in CT images.

(II) We propose a multi-scale attention structure in the encoder stage to fuse multi-scale features and obtain delicate global and local detail information.

(III) In the decoder stage, we propose an EA module to acquire the features of different receptive fields and communicate the complete context information.

(IV) The proposed M^{2}SUXNet can effectively improve the segmentation performance of elongated rib fractures.

The remainder of the paper is organized as follows: In Section 2, the framework structure of the proposed model is comprehensively explained. Section 3 describes the experimental setup and evaluation methodology, and analyzes the results of the experiment. Section 4 discusses some related issues. Finally, Section 5 summarizes the conclusions.

Methods

In this section, we first introduce the M^{2}SUXNet architecture for rib fracture segmentation. Then, the basic structure and function of M^{2}SF module and EA module are introduced in detail. Finally, the data set used in the experiment, implementation details, and evaluation metrics are introduced. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Overview of the model

Compared with two-dimensional (2D) images, 3D images are more complex and require higher computing resources. At the same time, rib fractures in CT images have the characteristics of unfixed positions and different shapes. Compared with traditional convolution, deep convolution extracts more complex feature representations through multi-level convolution and pooling, reduces the number of parameters, and improves efficiency and generalization ability. Point convolution reduces the number of parameters and calculations by reducing the channel dimension, enhances the expression ability of the model, and improves the quality of feature representation and model performance. Based on the above factors, we designed the M^{2}SUXNet network architecture based on 3DUXNet. M^{2}SUXNet aims to obtain feature information of different scales from 3D chest CT images and enhances the model's attention to tiny details of images through the dual attention mechanism. The specific structure of M^{2}SUXNet is shown in *Figure 1*.

UXNet Block

In order to perform feature extraction efficiently with fewer parameters, 3DUXNet is used as the baseline network and the original UXNet Block is maintained in the proposed model M^{2}SUXNet. *Figure 2* illustrates the UXNet Block structure. The UXNet Block comprises four stages, each with two large kernel (LK) convolutional blocks (i.e., L=8 total layers). Each block has a depthwise convolutional scaling (DCS) layer followed by a depthwise weighted convolution

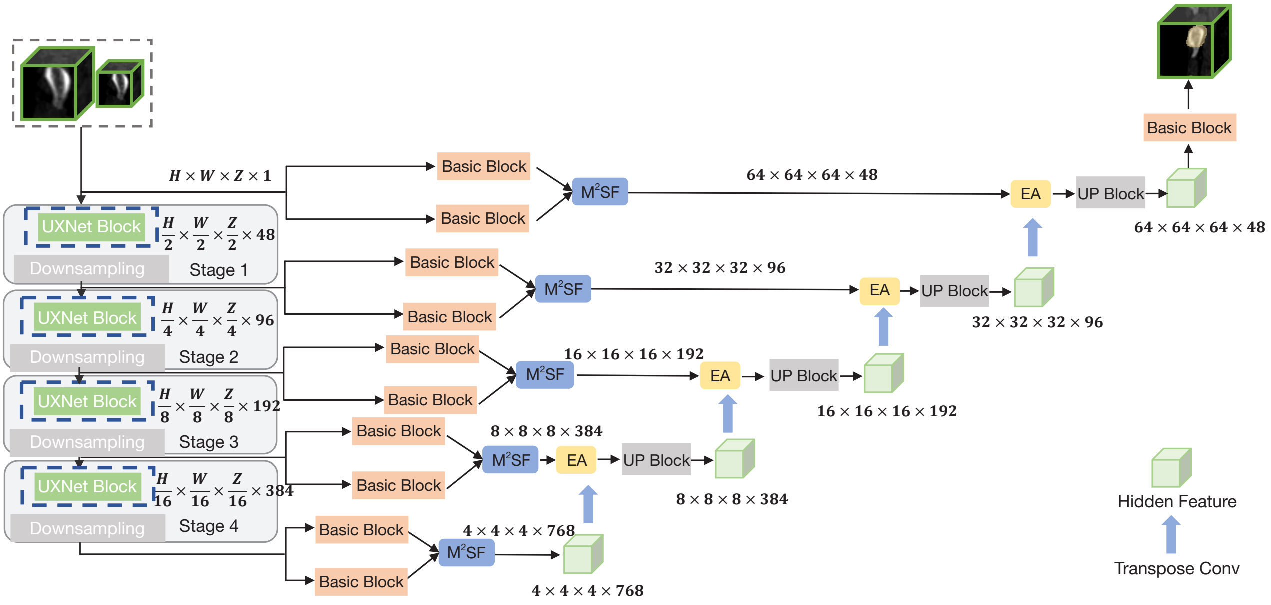

Figure 1 The overall framework of M^{2}SUXNet. M^{2}SF, multi-stream and multi-scale fusion; EA, efficient attention.

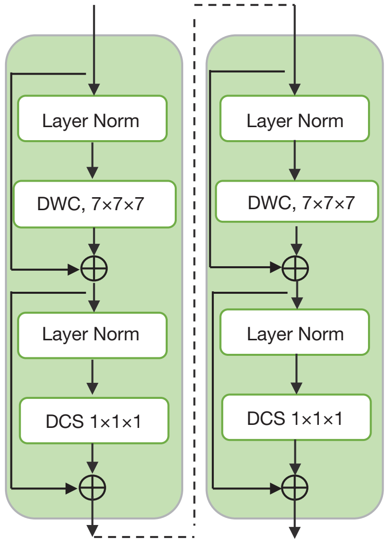

Figure 2 The basic structure of UXNet Block. DWC, depthwise weighted convolution; DCS, depthwise convolutional scaling.

(DWC) layer. The DCS layer helps to scale the feature map dimension (up to 4 times the input channel size) without increasing the model parameters. It minimizes redundancy in the volumetric context learned across channels. In order to exchange information between channels, we use a standard convolutional block with a kernel size of 2×2×2 and a stride of 2. The same process continues in phases 2, 3, and 4. Such hierarchical representations are extracted in each stage in a multiscale setting and are further used to learn dense volumetric segmentation. The outputs of layers *l* and *l+1* are defined as follows.

$$\hat{z}^{l} = DWC\left(LN\left(\hat{z}^{l-1}\right)\right) + \hat{z}^{l-1} \qquad [1]$$

$$z^{l} = DCS\left(LN\left(\hat{z}^{l}\right)\right) + \hat{z}^{l} \qquad [2]$$

$$\hat{z}^{l+1} = DWC\left(LN\left(\hat{z}^{l}\right)\right) + \hat{z}^{l} \qquad [3]$$

$$z^{l+1} = DCS\left(LN\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1} \qquad [4]$$

Where, $\hat{z}^{l}$ and $\hat{z}^{l+1}$ are the outputs of the deep convolutional layers at layer (*l*) and layer (*l*+1), respectively. z^{l} and z^{l+1} are the outputs of the depth convolutional scaling layers of layer (*l*) and layer (*l*+1), respectively. *DWC*() represents the deep convolution operation, which is used to apply a convolution kernel on each channel independently. *DCS*() stands for deep convolutional scaling and is used to scale the feature representation in the channel dimension. *LN*() stands for layer normalization, which is used to stabilize the training process and accelerate convergence.

Feature extraction stage

In the encoder module, a novel M^{2}SF module is proposed, which aims to strengthen the focus of the encoder on tiny image features while fusing feature information at different scales. The encoder module includes Basic Block, M^{2}SF, and a downsampling module. In the encoder path, the feature information from different scales first passes through the feature extraction stage, such as Basic Block. Then, fusion and refinement of different scale features are realized by the M^{2}SF module. Finally, the fused features were used as the encoder module output.

Basic Block

The Basic Block consists of two convolutional layers: batch normalization (BN) and ReLU activation functions. The convolutional layer extracts input features, BN accelerates training and improves network stability, and the ReLU activation function introduces nonlinearity to enhance the representation ability of the model. The specific structure is shown in *Figure 3*.

Multi-stream and multi-scale fusion module

The normalization-based attention module (29) uses a sparse weight penalty to improve computational efficiency and maintain the same performance. Based on this concept, this paper suggests the M^{2}SF module with a particular structure, as *Figure 4* illustrates. The M^{2}SF module uses the design of two parallel paths to integrate the feature information of different scales. In addition, the module can effectively distinguish the more significant feature information in the channel. The scale factor in BN reflects the size of the change in each channel, indicating its significance where the scaling factor is simply the variance of BN. The more prosperous the information in the channel, the more critical the feature information. Hence, the more significant the variance, the more dramatic the change in the channel. Conversely, channels with limited information provide only one type of information and are of limited value. Eqs. [5,6] represent the output features obtained at two different scales.

$$F_{out\_32} = sigmoid\left(W_{\gamma}\left(BN\left(F_{32}\right)\right)\right) \quad [5]$$

$$F_{out\_64} = sigmoid\left(W_{\gamma}\left(BN\left(F_{64}\right)\right)\right) \quad [6]$$

Among them, r reflects the scaling factor of each channel, W_{γ} represents the weight penalty, *sigmoid*() stands for sigmoid activation function, F_{32} represents the input features of scale 32, F_{64} represents the input features of scale 64, F_{out_32} represents the output features of scale 32, and F_{out_64} represents the output features of scale 64.

The final output feature of the M^{2}SF module is represented in Eq. [7].

$$F_{out} = Trans\left(F_{out\_32}\right) + F_{out\_64} \quad [7]$$

Among them, *Trans*() represents deconvolution, F_{out_32} represents a feature map of scale 32, F_{out_64} represents a feature map of scale 64, and F_{out} represents output features.

By embedding the M^{2}SF module into the neural network, feature information from different scales can be integrated. Using different feature levels to analyze image content and improve the model's attention to relevant

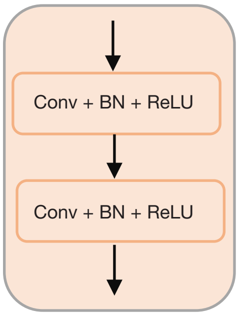

Figure 3 The basic structure of Basic Block. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit.

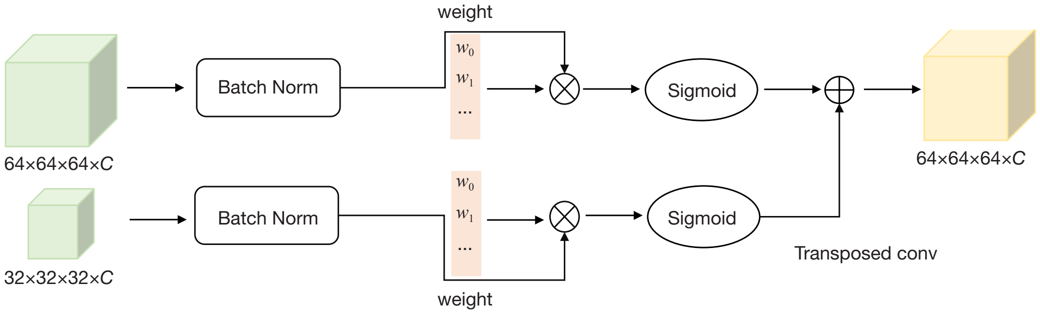

Figure 4 The basic structure of the multi-scale attention fusion module.

feature information can enhance the model's ability to pay attention to different fracture shapes in rib CT images and improve segmentation accuracy.

Feature integration truncation

A new EA module is proposed in the decoder module, which aims to enhance the attention of the network to the tiny details of the fracture. The EA module effectively captures channel interaction information from low- and high-level features and enables the attention mechanism to adjust the two sets of features before fusing them. The decoder module includes UP Block, EA module, and deconvolution. In the decoder path, the low-level features from the encoder module and the high-level features from the decoder module are first spatially refined by the EA module and then concatenated and fused by the convolutional block. Finally, UP Block learns features and adapts to attention mechanisms.

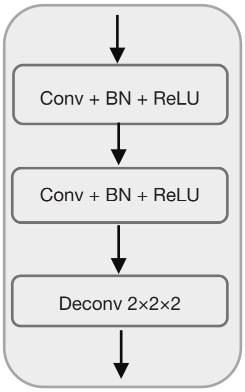

Figure 5 The basic structure of UP Block. Conv, convolutional layer; BN, batch normalization; ReLU, rectified linear unit; Deconv, deconvolutional layer.

UP Block

The UP Block consists of two convolutional layers, a BN layer, a ReLU activation function, and a deconvolution layer. The convolutional layers extract the input data features, and after BN and ReLU activation, the deconvolution layer performs upsampling to recover the spatial dimension of the input data. *Figure 5* shows the specific structure of UP Block.

EA module

This paper proposes an attention enhancement module named EA to improve the ability of the network model to obtain minute details of rib fractures. The specific structure is shown in *Figure 6*. The module combines different parts of input features to integrate and weigh features in other channels and spatial locations. It improves the ability of the model to distinguish different shapes of rib fractures and suppress noise by adapting the channel relationship in the input feature map. The EA module consists of two identical branches. There are two parts involved in the processing of features. One processes the features passed through the encoder, and the other processes the features passed through the decoder. Taking the forward propagation process of a single branch as an example, it is as follows.

The input feature map is split into two parts X_1 and X_2, performing 3×3 convolution operations on each part. Then, the feature map X is obtained by integrating the information from the different branches by concatenating the output feature maps of the two branches together. Next, channel shuffling is performed, and the average and maximum values of the feature maps are extracted using adaptive

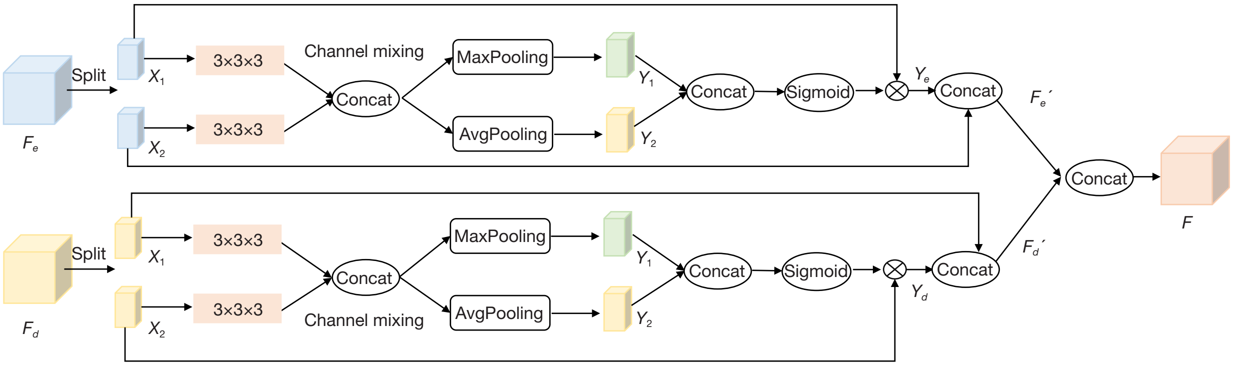

Figure 6 The basic structure of the efficient attention module.

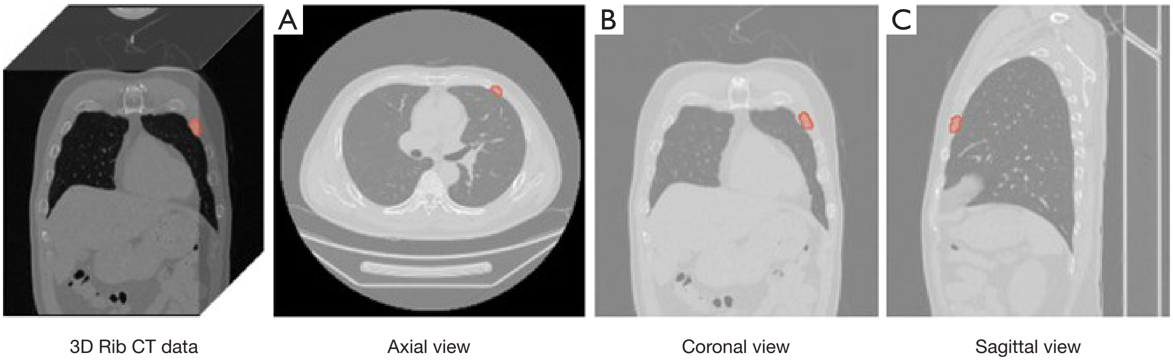

Figure 7 Example image of 3D rib CT data. The red marked area represents the presence of rib fractures in this area. (A) Axial view, (B) coronal view, and (C) sagittal view.

average pooling and adaptive max pooling to obtain Y_1 and Y_2 respectively. This process is shown in Eqs. [8,9]. Where *maxp*() represents the max pooling operation and *avgp*() represents the average pooling operation.

$$Y_1 = maxp(X) \quad [8]$$

$$Y_2 = avgp(X) \quad [9]$$

In order to introduce a non-linear relationship and enhance the discriminative ability to distinguish between different elements, we utilize the Sigmoid function as an activation function. Each element of the output feature map is mapped to a range from 0 to 1. Finally, the output of the Sigmoid function is multiplied at the element level with the original feature map X_1 to obtain the final enhanced feature map Y.

This process is shown in Eq. [10]. Where, *mul*() represents multiplication, *sigmoid*() represents the sigmoid activation function, and *concat*() represents feature concatenation.

$$Y = mul\left(sigmoid\left(concat\left(Y_1, Y_2\right)\right), X_1\right) \quad [10]$$

The final output feature map $F_e^{'}$ of a single branch is formed by concatenating the enhanced feature submap Y and the original feature map X_2. This process is shown in Eq. [11], where *concat*() represents feature concatenation.

$$F_e^{'} = concat(Y, X_2) \quad [11]$$

The EA module performs the task of integrating and weighing input features from various channel locations. It enhances the model's ability to segment different shapes of rib fractures and suppress noise by adaptively adjusting the channel relationship in the input feature map.

Loss function

The loss function used for model training consists of classification loss BCE With Logits Loss and segmentation loss Dice Loss. The formula is as follows:

$$L = \lambda_1 L_{BCE}(P, G) + \lambda_2 L_{Dice}(P, G) \quad [12]$$

Where, L_{BCE} denotes BCE With Logits Loss and L_{Dice} represents Dice Loss. P is the predicted image and G is the ground truth. λ_1 and λ_2 are weight coefficients, λ_1 is set as 0.5, and λ_2 is set as 1.

The joint loss function of BCE With Logits Loss and Dice Loss is applied to the rib fracture segmentation task, which improves the robustness and generalization ability of the model by considering pixel-level and region-level information in CT images.

Dataset

The experimental data were obtained from the publicly available RibFrac Dataset of the MICCAI 2020 RibFrac Challenge. The dataset contains 420 samples from the training set, 80 from the validation set, and 160 from the test set. The data are 3D rib CT images annotated by several radiologists with varying experience in chest CT interpretation. An example image of 3D rib CT data is shown in *Figure 7*.

Three views show the results of slicing the data in the axial, coronal, and coronal planes. In addition, rib fractures of different shapes are shown in *Figure 8*. Considering that unannotated test set samples cannot be used to calculate prediction accuracy, 160 samples were removed from this experiment. *Table 1* shows the specific data set allocation.

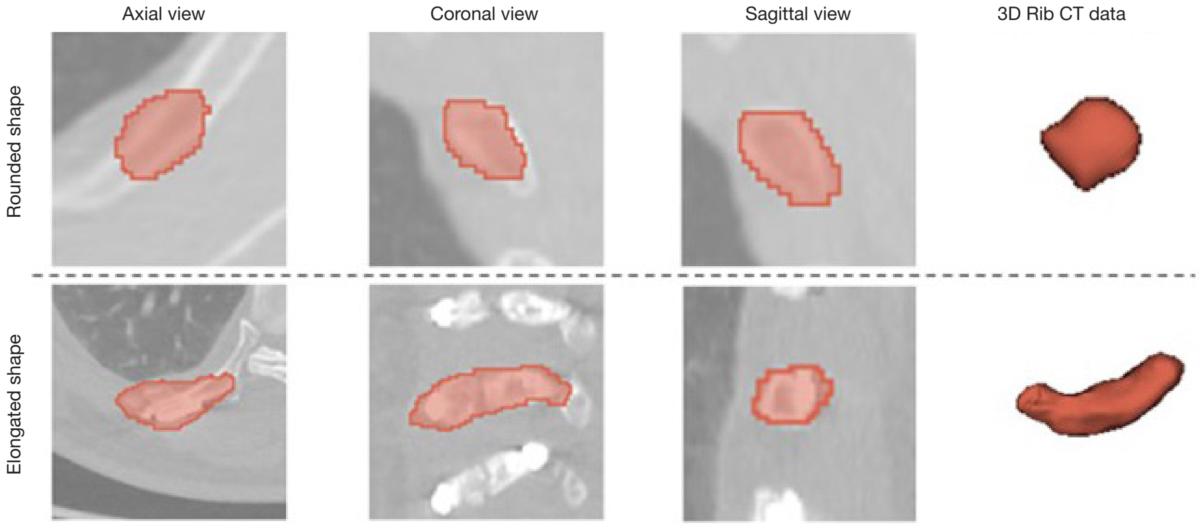

Figure 8 Schematic representation of rib fractures of different shapes. These include axial view, coronal view, sagittal view, and 3D visualization results of rib fracture.

Table 1 Detailed dataset distribution, number of samples, and number of fractures included

Dataset	Sample size	Number of fractures
Training set	395	3,096
Validation set	50	442
Testing set	55	885

Table 2 Experiment environment and configuration parameters

Parameters	Details
Optimizer	SGD
Data augmentation	Horizontal flipping
CUDA	11.7
Python	3.9
Deep learning framework	Fastai
Maximum learning rate	1e–1
Epoch	650
Batch size	6

Implementation details

The images at two scales, 32×32×32 and 64×64×64, are input into the network encoder to obtain the feature information of images at different scales fully. Considering that in the rib fracture segmentation task, the appearance of the case may vary horizontally; for example, the lesion may appear on the left or right side. More training samples in different directions can be generated by horizontal flipping, which improves the generalization ability of the model and the ability to identify lesions in different directions and enhances the robustness of the model. Therefore, we employ horizontal flipping for data augmentation during training. For the programming and development environment, CUDA-Toolkit 11.7 is utilized. Python 3.9 is the programming language used, while the fastai is employed for the DL framework. In the experiment, we use the SGD optimizer to train the network with the batch size of 6, the epoch of 650. The learning rate increases linearly within the first epoch, from 0.00001 to 0.1, and gradually decreases and stabilizes in subsequent training. The specific details are shown in *Table 2*.

Evaluation metrics

This study applied the Dice coefficient (Dice), the joint intersection over union (IoU), and Precision to evaluate segmentation performance for rib fractures. The specific formula is as follows:

$$Dice = \frac{2*|A \cap B|}{|A| + |B|} \qquad [13]$$

$$IoU = \frac{|A \cap B|}{|A \cup B|} \qquad [14]$$

Here, A represents the set of pixels in the binarized image of the prediction result. B represents the set of pixels in the binarized image of the true label. $|A|$ denotes the number of elements in set A, and $|B|$ denotes the number of elements in set B.

In computer-aided diagnosis technology, high Dice/IoU accuracy ensures precise segmentation of fracture regions. This enhances the accuracy of fracture diagnosis and aids doctors in formulating more effective surgery and rehabilitation plans. Low Dice/IoU can lead to false or missed fracture detection. This affects diagnostic decisions and treatment outcomes, especially in cases of multiple or complex fractures. It can directly impact patients' clinical prognosis and the quality of care.

$$Precision = \frac{TP}{TP + FP} \quad [15]$$

Here, TP represents the number of samples correctly predicted as positive class and FP represents the number of samples incorrectly predicted as positive class. Precision measures the proportion of regions predicted by the model as existing fractures. High Precision indicates that the model effectively reduces false positives in cases where normal areas are misclassified as fractures.

Results

Main result

Studies have shown that it is harder to segment elongated rib fractures accurately. To verify the effectiveness of the model M^{2}SUXNet, we first report the predicted segmentation illustration of the proposed M^{2}SUXNet for circular and elongated shapes in the test set (*Figure 9*). Our proposed model can effectively solve the problem of low segmentation accuracy in elongated rib fractures. By adopting a multi-scale fusion approach and adding the dual attention mechanisms, our model significantly improves the segmentation of elongated shape rib fractures.

Ablation experiment

To investigate the contribution of different components of M^{2}SUXNet to improve rib fracture segmentation accuracy, we selected different network architectures for experiments on public datasets. Firstly, the performance of the joint loss function of Dice Loss and BCE Loss was verified. The joint Loss function of Dice Loss and BCE Loss was also compared with the joint loss function of Dice Loss, Dice Loss and Focal Loss. M^{2}SUXNet is used as a benchmark, as shown in *Figure 1*. *Table 3* compares loss functions.

Compared with the joint loss function of Dice Loss, Dice Loss and Focal Loss, the joint loss function of Dice Loss and BCE Loss achieved better training results, mainly because of their complementarity in segmentation tasks. Dice Loss performed well when coping with small targets and sample imbalance. In contrast, BCE Loss provided stable gradient signals early in training and accelerated model convergence through pixel-by-pixel loss calculation. Combining these two loss functions can improve segmentation accuracy and enhance training stability. In contrast, the joint loss of Dice and Focal Loss does not work well because their functions partially overlap. Focal Loss deals with class imbalance and focuses on hard-to-classify samples. However, Dice Loss solved the imbalance problem in segmentation tasks well, resulting in the limited contribution of Focal Loss. In addition, the Focal Loss may cause large gradient fluctuations early in training. This affects the model's convergence stability and leads to poorer segmentation results than a joint scheme of Dice and BCE Loss.

Secondly, the effectiveness of the multi-stream and multi-scale feature fusion strategy was verified. To fully demonstrate the contribution of the proposed module, we fused three classical attention mechanisms in the feature extraction stage of the model: convolutional block attention module (CBAM) module, SE module (30), and efficient channel attention (ECA) module (31), respectively. U represents the UXNet network with a joint loss function. We used U as a benchmark to assess network architectures and compare fusion versions of M^{2}SF and EA. Here, "U+32" means that only 32×32×32 scale images are input into the network, "U+64" implies that only 64×64×64 scale images are input into the network, "U+32+64" indicates that the two different scales are combined by deconvolution, "U+32+64+CA" means adding CA module to U while fusing two different scales by deconvolution, "U+32+64+SE" means adding SE module to U while fusing two different scales by deconvolution, "U+32+64+ECA" means adding ECA module to U while fusing two different scales by deconvolution, and "U+M^{2}SF" signifies that the M^{2}SF structure is added to U.

The experimental results in *Table 4* show that the multi-scale fusion method significantly outperforms the single-scale method in the rib fracture segmentation task. In particular, the "U+M^{2}SF" structure achieves the best segmentation performance in the feature extraction stage

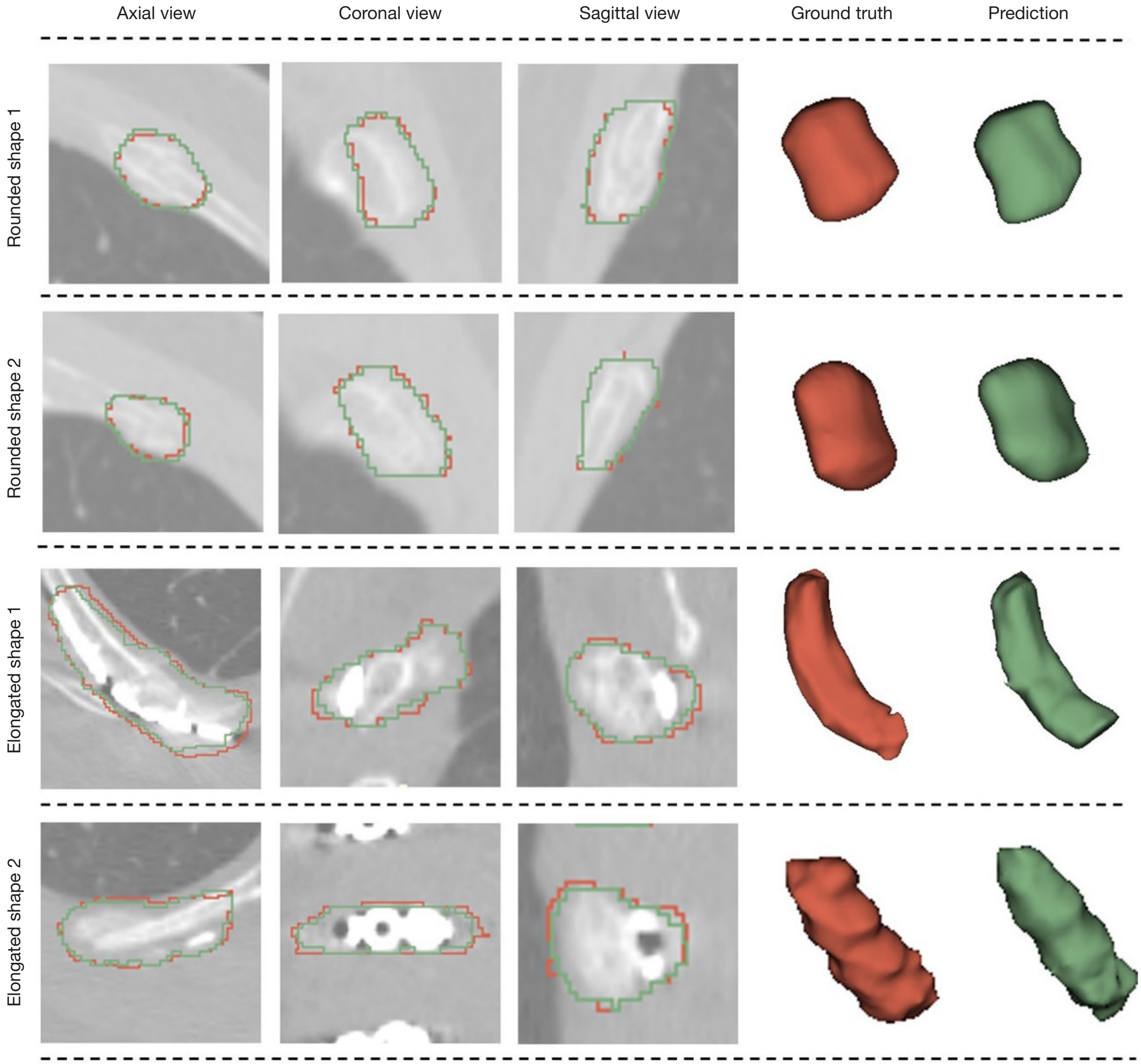

Figure 9 Visualization of M^{2}SUXNet segmentation results for circular and elongated bar rib fractures in the test set. The true label is shown in red, and the prediction of the model is shown in green.

Table 3 Comparing the segmentation performance of different loss functions on the test set

Loss function	Dice	IoU
Dice Loss	74.82%	59.77%
Dice Loss + Focal Loss	73.85%	58.54%
Dice Loss + BCE Loss (Ours)	75.34%	60.44%

Dice, Dice similarity coefficient; IoU, intersection over union; Dice Loss, Dice Similarity Coefficient Loss; Focal Loss, Focal Cross-Entropy Loss; BCE Loss, Binary Cross-Entropy Loss.

of the model by introducing the M^{2}SF module. This result indicates that different fracture morphologies (such as circular and elongated fractures) have different requirements for multi-scale features. The M^{2}SF module better captures the delicate features of these complex morphologies through multi-scale feature fusion, thereby improving the accuracy of segmentation. In contrast, although SE, CBAM, and ECA modules can improve the efficiency of feature extraction through the attention mechanism, they need to be improved in dealing with the complexity of fracture morphology. The SE module is weighted only for

Table 4 Comparing the segmentation performance of different architectures on the test set

Model	Dice	IoU
U+32	69.58%	53.35%
U+64	70.76%	54.75%
U+32+64	71.52%	55.67%
U+32+64+CBAM	69.12%	52.81%
U+32+64+SE	71.94%	56.18%
U+32+64+ECA	69.48%	53.23%
U+32+64+M^{2}SF	73.08%	57.58%

U, the UXNet network with a joint loss function; U+32, 32×32×32 scales image inputs; U+64, 64×64×64 scales image inputs; U+32+64, two different scales are combined by deconvolution; U+32+64+CBAM, adding CBAM module to U while fusing two different scales by deconvolution; U+32+64+SE, adding SE module to U while fusing two different scales by deconvolution; U+32+64+ECA, adding ECA module to U while fusing two different scales by deconvolution; U+M^{2}SF, the M^{2}SF structure is added to U. Dice, Dice similarity coefficient; IoU, intersection over union.

channels. The CBAM module combines channel and spatial attention. Although it enhances basic features, it cannot adequately handle complex fracture details. Meanwhile, the ECA module focuses on channel attention, which improves the efficiency of feature utilization but insufficiently captures subtle spatial features. The M^{2}SF module uses normalization and channel attention weighting to assess the importance of each feature, capturing relationships across different scales. This is especially effective for difficult-to-segment fractures, such as elongated fractures. Therefore, the strategy of multi-scale fusion significantly improves the accuracy of segmentation, especially in the case of complex and diverse morphologies of rib fractures.

To verify the model's segmentation performance on accurate data, we visualized the predicted segmentation map of the model. As shown from the visualization results in *Figure 10*, the segmentation performance of the "U+32" structure in circular and elongated rib fractures is not ideal. The main reason is that the scale of the input image is too small, so the model cannot capture enough feature information. The small scale limits the ability of the model to extract the details of the fracture region, especially when dealing with complex or subtle fractures. It is difficult for the model to distinguish the fracture region from the background accurately. In addition, the single small-scale input gives the model insufficient expression at the feature level, which weakens the overall understanding of the fracture shape, thus affecting the segmentation accuracy. The model with the M^{2}SF module achieved the best results for circular and elongated fractures. This indicates that the multi-scale feature fusion mechanism in the M^{2}SF module offers significant advantages during feature extraction. The M^{2}SF module adjusts feature representation by weighting each channel based on importance. This enables it to fully capture the detailed features of rib fractures, particularly in complex shapes such as elongated fractures.

In contrast, although the model integrated with the SE module performs well on segmenting circular fractures, it performs relatively poorly on elongated fractures. The SE module mainly enhances features by channel weighting but needs to adequately capture the local detail changes in fracture shape. Therefore, the SE module has limitations in handling complex and morphologically variable elongated fractures, and it performs worse than the M^{2}SF module. It can be seen from the figure that although the "U+M^{2}SF" structure performs well in the overall segmentation effect, it still has shortcomings in regards to detail. Therefore, we will make further improvements in the model's feature integration stage to enhance the model's ability to capture subtle features.

Finally, this paper verifies the effectiveness of the proposed EA module. We fused three classical attention mechanisms in the feature integration stage of the model: CBAM module, SE module, and ECA module. Here, "U+M^{2}SF" signifies that the M^{2}SF structure is added to U, "U+M^{2}SF+CA" indicates fusing M^{2}SF structure and CA structure to U, "U+M^{2}SF+SE" indicates fusing M^{2}SF structure and SE structure to U, "U+M^{2}SF+ECA" indicates fusing M^{2}SF structure and ECA structure to U, and "U+M^{2}SF+EA" indicates fusing M^{2}SF structure and EA structure to U.

Table 5 summarizes the performance comparison of several structures. From the segmentation accuracy of each structure in the table, "U+M^{2}SF+EA" has the best result, with its Dice value reaching 75.34% and its IoU value reaching 60.44%. The proposed EA module achieves effective feature enhancement and fusion in the feature integration stage by combining branch convolution and channel attention mechanisms. In contrast, when ECA and CBAM modules are introduced, the segmentation accuracy of the model decreases. This may be because the channel attention mechanism introduced by the ECA module is too global. The design of the ECA module is based on channel

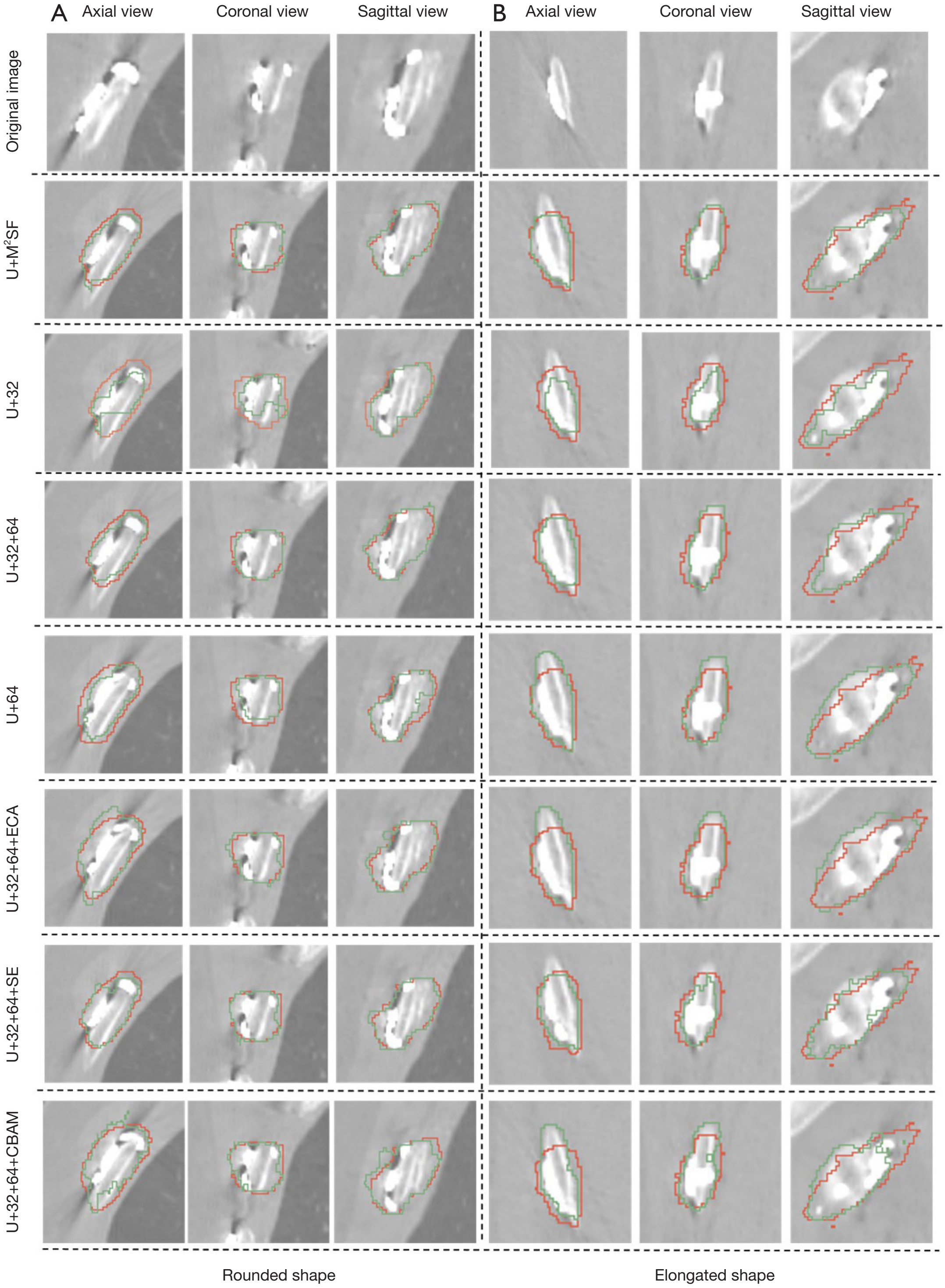

Figure 10 Visualization of the ablation experiment. (A) segmentation renderings of different models for circular rib fractures, (B) segmentation renderings of different models for elongated shape rib fractures. The true label is shown in red, and the prediction of the model is shown in green. U, the UXNet network with a joint loss function; U+M^{2}SF, the M^{2}SF structure is added to U; U+32, 32×32×32 scales image inputs; U+64, 64×64×64 scales image inputs; U+32+64, two different scales are combined by deconvolution; U+32+64+ECA, adding ECA module to U while fusing two different scales by deconvolution; U+32+64+SE, adding SE module to U while fusing two different scales by deconvolution; U+32+64+CBAM, adding CBAM module to U while fusing two different scales by deconvolution.

Table 5 Comparing the segmentation performance of different architectures on the test set

Model	Dice	IoU
U+M^{2}SF	73.08%	57.58%
U+M^{2}SF+CBAM	72.47%	56.83%
U+M^{2}SF+SE	73.84%	58.53%
U+M^{2}SF+ECA	70.68%	54.66%
U+M^{2}SF+EA	75.34%	60.44%

U+M^{2}SF, the M^{2}SF structure is added to U; U+M^{2}SF+CBAM, fusing M^{2}SF structure and CBAM structure to U; U+M^{2}SF+SE, fusing M^{2}SF structure and SE structure to U; U+M^{2}SF+ECA, fusing M^{2}SF structure and ECA structure to U; U+M^{2}SF+EA, fusing M^{2}SF structure and EA structure to U. Dice, Dice similarity coefficient; IoU, intersection over union.

attention to global information, which needs to pay more attention to detailed features. The CBAM module combines channel attention and spatial attention, attempting to adjust the information distribution of the feature map through many aspects. However, this combination method may lead to excessive adjustment of the feature map, especially when dealing with complex fracture shapes. The local information of the feature may be overly smoothed or adjusted, and then the accurate recognition of the target structure is affected. ECA and CBAM modules may not effectively enhance feature expression and detail capture. Their added complexity could negatively impact the model's performance.

In addition, to visually verify the segmentation effect of the model on the rib fracture region, we visualized the segmentation effect of the model, as shown in *Figure 11*. It can be seen that "U+M^{2}SF+ECA", "U+M^{2}SF+SE", and "U+M^{2}SF+CBAM" perform well in dealing with circular fractures but have some shortcomings in segmenting elongated bar fractures, which is caused by the relatively limited expressive power and fusion strategy of these modules in the feature integration stage. The "U+M^{2}SF+EA" structure performs best in the segmentation task for circular and elongated fractures. This indicates that the EA module has better adaptability to fractures of different shapes and sizes while effectively capturing subtle features in the feature integration stage.

Comparative experiment

To demonstrate the effectiveness of the proposed rib fracture segmentation model M2SUXNet, we selected seven state-of-the-art semantic segmentation models in comparison experiments, including 3DUNet, 3D Attention UNet, 3DUXNet, Unetr (32), SwinUnetr (33), nnFormer (34), and FracNet. In addition, we use the same hyperparameter configuration and post-processing strategy as M2SUXNet in the training phase.

Table 6 displays the segmentation results of various models on the public dataset for the fracture lesion region. Traditional 3DUNet and 3D Attention UNet use a symmetric encoder-decoder structure. Although they can extract deep features, they struggle with capturing detailed information in complex objects. 3DUNet relies only on a simple up–down sampling structure and weakly understands the complex shape of fractures, which explains its low Dice (70.71%) and IoU (54.69%). M2SUXNet may enhance the capture of fracture details at different scales by introducing the M^{2}SF module. For the task of fracture segmentation, the model not only needs to capture global context information to identify the overall structure of the rib but also needs to be able to focus on local details to locate the fracture area accurately. This is improved in 3D Attention UNet, which better handles the weight allocation of features between regions by introducing an attention mechanism. However, due to the lack of multi-scale information capture, its Dice and IoU performance still needs improvement to M2SUXNet. In contrast, SwinUnetr and nnFormer introduced the Transformer architecture, which is good at global information modeling and performs well when processing large-scale medical images. SwinUnetr deals with local attention using a sliding window and can understand the global context in fracture segmentation. Dice reaches 74.18%, but its capture of local feature details is not as good as M2SUXNet, resulting in insufficient segmentation of the level of detail. nnFormer further improves the feature modelling ability by the Transformer encoder, reaching Dice 72.87%. However, compared with M2SUXNet, its comprehensive processing of local details and global context still needs to be improved in fracture shape. The advantage of M2SUXNet is that through an effective multi-scale structure, it can capture the details of the fracture area and establish cross-level information association between different scales. This structure makes fracture segmentation more accurate, and the segmentation Dice can reach 75.34%, and the IoU is 60.44%, significantly better than the segmentation performance of other models.

In addition, we evaluated the precision of each model. Among them, 3DUNet, AttUNet, and FracNet have lower

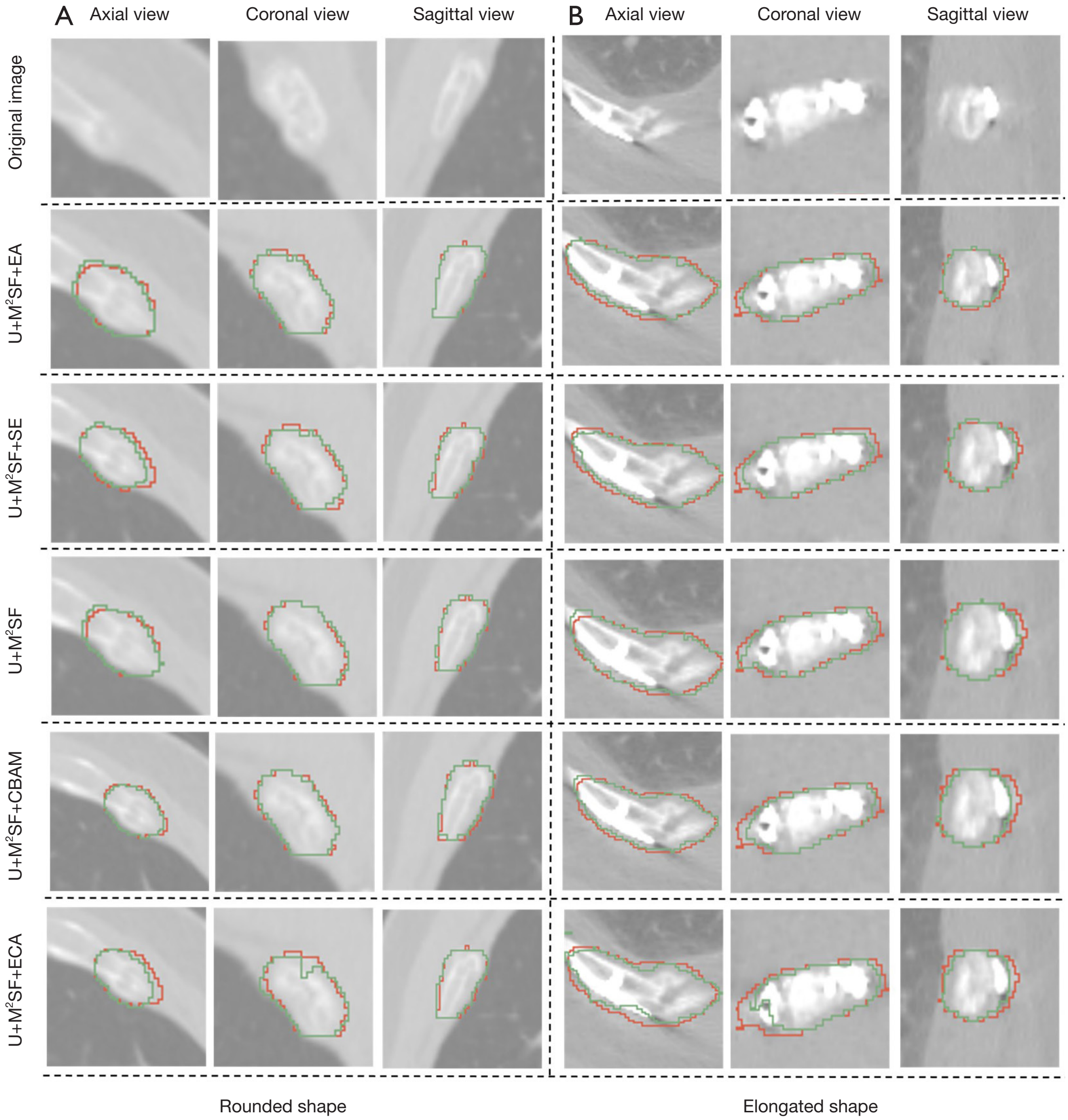

Figure 11 Visualization of the ablation experiment. (A) Segmentation renderings of different models for circular rib fractures, (B) segmentation renderings of different models for elongated shape rib fractures. The true label is shown in red, and the prediction of the model is shown in green. U+M^{2}SF, the M^{2}SF structure is added to U; U+M^{2}SF+EA, fusing M^{2}SF structure and EA structure to U; U+M^{2}SF+SE, fusing M^{2}SF structure and SE structure to U; U+M^{2}SF+CBAM, fusing M^{2}SF structure and CBAM structure to U; U+M^{2}SF+ECA, fusing M^{2}SF structure and ECA structure to U.

precision, possibly related to their simpler model structure and limited feature extraction ability. M^{2}SUXNet has the best performance, with a precision of 93.79%. This result may be attributed to its advanced multi-scale and multi-stream fusion design, which enables it to capture the details of the fracture region better and reduce false positives.

To evaluate the performance of the models in the case of fractures with different morphologies, we visualized the segmentation effects of each model. It can be seen from *Figure 12* that SwinUnetr performs prominently for circular fractures because it introduces a sliding window mechanism, which can effectively capture the global and local information of the image to segment the circular fracture area accurately. However, SwinUnetr is slightly

Table 6 Comparing the segmentation performance of different networks on the test set

Method	Dice	IoU	Precision
3DUNet (2015)	70.71%	54.69%	89.69%
AttUNet (2018)	70.03%	53.88%	89.24%
FracNet (2020)	71.52%	55.67%	90.90%
SwinUnetr (2021)	74.18%	58.96%	92.60%
Unetr (2022)	70.68%	54.66%	91.05%
3DUXNet (2023)	71.45%	55.58%	90.68%
nnFormer (2023)	72.87%	57.32%	91.51%
M^{2}SUXNet	75.34%	60.44%	93.79%

AttUNet, Attention 3DUNet; FracNet, 3D Rib Fracture Diagnosis Network; Unetr, 3DUNet with Transformers, 3DUNet with Swin Transformers; 3DUXNet, U-Net with Large Kernel Convolutions; nnFormer, Neural Network Former; Dice, Dice similarity coefficient; IoU, intersection over union.

inadequate when dealing with elongated fractures. This may be because fractures with elongated shapes depend more on the efficient extraction of local features and information fusion across scales. In contrast, M^{2}SUXNet performs well on the circular and elongated fractures segmentation task. Although nnFormer and 3DUXNet perform well in global information modelling, they are deficient in capturing local features, resulting in poor performance in handling elongated fractures.

Discussion

The uneven shape and unfixed position of rib fractures in CT images complicate the diagnostic process. This can easily lead to misdiagnosis or missed diagnosis. In addition, subtle fractures or multiple fractures further increase the segmentation difficulty. In clinical applications, CT's high-resolution 3D images reveal the fine structure of the rib, including the fracture line and fracture slice. Accurate segmentation enables CT to comprehensively assess chest injuries (35), including rib fractures and other chest structures such as lungs and hearts. This is essential for a comprehensive assessment of patient injuries and treatment planning. At the same time, accurate rib fracture diagnosis and segmentation can help doctors formulate more effective treatment plans (36) and improve treatment effects.

The difficulty of rib fracture segmentation lies in the fracture regions with complex shapes, such as elongated, curved, and even irregular fracture shapes (37). This paper introduces multi-stream and multi-scale fusion into the rib fracture segmentation task for the first time. Moreover, the corresponding information fusion module is designed for this architecture to fully use the feature maps' information and fuse and complement the information between different feature maps. The multi-stream network can better capture the edge information of different scales of rib fractures in low-level feature acquisition. In this way, the low-level feature information of the edge of different scales of rib fractures can be captured. At the same time, the high-level semantic feature information of rib fracture can be supplemented between different scales through the multi-scale information fusion strategy. Studies have shown that elongated rib fractures are more difficult to segment than circular fractures. When segmenting rib fractures, the model effectively discriminates and diagnoses various fracture shapes. We verified the effectiveness of the proposed model on the public RibFrac dataset. Our proposed model can effectively improve the segmentation performance of elongated shape rib fractures.

Despite the limitations of most current rib fracture segmentation methods, this study is of great significance for dealing with the complexity of rib fractures in CT images. The complex morphology and unfixed position of rib fracture in CT images make the segmentation task very challenging. The proposed method achieves a 75.34% Dice coefficient and 60.44% IoU in CT images. It shows higher accuracy and feasibility than existing models, especially in segmenting elongated fractures. Although Dice/IoU metrics can quantify segmentation accuracy, they may not fully capture critical details. Therefore, combining other relevant factors to ensure accurate clinical judgment in practical clinical applications is still necessary. The proposed method can provide valuable segmentation performance in practice and shows potential in dealing with complex fracture morphologies.

This study still has some limitations. The M^{2}SUXNet model has many parameters and high video memory requirements. Future work will consider designing lightweight segmentation networks to reduce parameters and computation while maintaining good segmentation performance. Secondly, the lack of fixed device fracture cases in public datasets may limit the model's generalization ability in complex clinical scenarios. Future studies should incorporate more such data further to improve the applicability and robustness of the model.

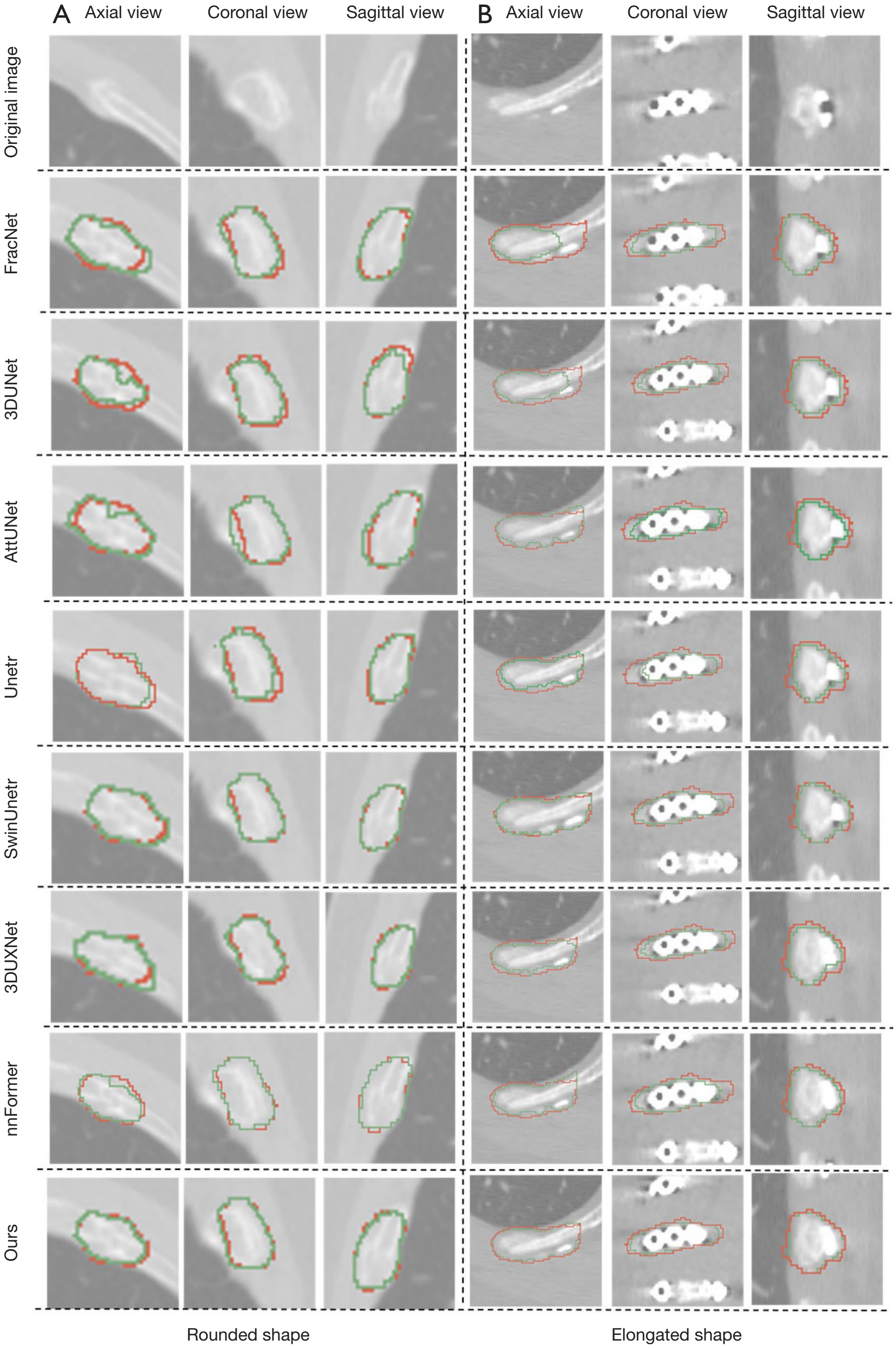

Figure 12 Visualization of the comparison experiment. (A) Segmentation renderings of different models for circular rib fractures, (B) segmentation renderings of different models for elongated shape rib fractures. The true label is shown in red, and the prediction of the model is shown in green. FracNet, 3D Rib Fracture Diagnosis Network; AttUNet, attention 3DUNet; Unetr, 3DUNet with Transformers; SwinUnetr, 3DUNet with Swin Transformers; 3DUXNet, UNet with Large Kernel Convolutions; nnFormer, Neural Network Former.

Conclusions

This paper proposes the M^{2}SUXNet network to enhance rib fracture segmentation accuracy. It particularly targets elongated fractures using multi-stream and multi-scale fusion mechanisms. Experimental results show that the proposed model achieves a Dice value of 75.34%, IoU value of 60.44%, and precision of 93.80% on the RibFrac public dataset, which are better than the values achieved by existing models. M^{2}SUXNet effectively solves the segmentation problem of complex fracture morphology in CT images and provides a reliable tool for clinical application. Despite its promising results, this study still has some limitations. Future research will focus on designing lightweight models and introducing diverse data to improve the applicability and robustness of the model.

Acknowledgments

Funding: This research was funded by The National Natural Science Foundation of China (No. U21A20390) and Education Department Project of Jilin Province (No. JJKH20240945KJ).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-24-1356/coif). All authors report funding from The National Natural Science Foundation of China (No. U21A20390) and Education Department Project of Jilin Province (No. JJKH20240945KJ). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. van Laarhoven JJEM, Hietbrink F, Ferree S, Gunning AC, Houwert RM, Verleisdonk EMM, Leenen LPH. Associated thoracic injury in patients with a clavicle fracture: a retrospective analysis of 1461 polytrauma patients. Eur J Trauma Emerg Surg 2019;45:59-63.
2. Liman ST, Kuzucu A, Tastepe AI, Ulasan GN, Topcu S. Chest injury due to blunt trauma. Eur J Cardiothorac Surg 2003;23:374-8.
3. Traub M, Stevenson M, McEvoy S, Briggs G, Lo SK, Leibman S, Joseph T. The use of chest computed tomography versus chest X-ray in patients with major blunt trauma. Injury 2007;38:43-7.
4. Hu J, Zheng ZF, Wang SH, Si DL, Yuan YQ, Gao BL. Missed rib fractures on initial chest CT in trauma patients: time patterns, clinical and forensic significance. Eur Radiol 2021;31:2332-9.
5. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:3431-40.
6. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing; 2015:234-41.
7. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, Tavares JMRS, Bradley A, Papa JP, Belagiannis V, Nascimento JC, Lu Z, Conjeti S, Moradi M, Greenspan H, Madabhushi A, editors. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing; 2018:3-11.
8. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B,

Glocker B, Rueckert D. Attention u-net: Learning where to look for the pancreas. 2018. arXiv: 1804.03999.

9. Xiao X, Lian S, Luo Z, Li S. Weighted res-unet for high-quality retina vessel segmentation. 2018 9th international conference on information technology in medicine and education (ITME), Hangzhou, China. IEEE; 2018:327-31.
10. Cai S, Tian Y, Lui H, Zeng H, Wu Y, Chen G. Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quant Imaging Med Surg 2020;10:1275-85.
11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-11.
12. Lee HH, Bao S, Huo Y, Landman BA. 3D UX-Net: A Large Kernel Volumetric ConvNet Modernizing Hierarchical Transformer for Medical Image Segmentation. 2023. arXiv: 2209.15076.
13. Rehman MU, Cho SB, Kim JH, Chong KT. Bu-net: Brain tumor segmentation using modified u-net architecture. Electronics 2020;9:2203.
14. Lin SY, Lin CL. Brain tumor segmentation using U-Net in conjunction with EfficientNet. PeerJ Comput Sci 2024;10:e1754.
15. Zhou D, Xu L, Wang T, Wei S, Gao F, Lai X, Cao J. M-DDC: MRI based demyelinative diseases classification with U-Net segmentation and convolutional network. Neural Netw 2024;169:108-19.
16. Ryu J, Rehman MU, Nizami IF, Chong KT. SegR-Net: A deep learning framework with multi-scale feature fusion for robust retinal vessel segmentation. Comput Biol Med 2023;163:107132.
17. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25:30-6.
18. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
19. Pomeranz CB, Barrera CA, Servaes SE. Value of chest CT over skeletal surveys in detection of rib fractures in pediatric patients. Clin Imaging 2022;82:103-9.
20. Ibanez V, Gunz S, Erne S, Rawdon EJ, Ampanozi G, Franckenberg S, Sieberth T, Affolter R, Ebert LC, Dobay A. RiFNet: Automated rib fracture detection in postmortem computed tomography. Forensic Sci Med Pathol 2022;18:20-9.
21. Zhou QQ, Wang J, Tang W, Hu ZC, Xia ZY, Li XS, Zhang R, Yin X, Zhang B, Zhang H. Automatic Detection and Classification of Rib Fractures on Thoracic CT Using Convolutional Neural Network: Accuracy and Feasibility. Korean J Radiol 2020;21:869-79.
22. Zhou Z, Fu Z, Jia J, Lv J. Rib Fracture Detection with Dual-Attention Enhanced U-Net. Comput Math Methods Med 2022;2022:8945423.
23. Cao Z, Xu L, Chen D Z, Gao H, Wu J. A Robust Shape-Aware Rib Fracture Detection and Segmentation Framework with Contrastive Learning. IEEE Transactions on Multimedia. IEEE; 2023;25:1584-91.
24. Jin L, Yang J, Kuang K, Ni B, Gao Y, Sun Y, Gao P, Ma W, Tan M, Kang H, Chen J, Li M. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet. EBioMedicine 2020;62:103106.
25. Gao Y, Chen H, Ge R, Wu Z, Tang H, Gao D, Mai X, Zhang L, Yang B, Chen Y, Coatrieux JL. Deep learning-based framework for segmentation of multiclass rib fractures in CT utilizing a multi-angle projection network. Int J Comput Assist Radiol Surg 2022;17:1115-24.
26. Zhang H, Zu K, Lu J, Zou Y, Meng D. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. In: Wang L, Gall J, Chin TJ, Sato I, Chellappa R, editors. Computer Vision – ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part III. Springer; 2023:541-57.
27. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional Block Attention Module. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision – ECCV 2018. Springer; 2018:3-19.
28. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA. IEEE; 2021:13713-22.
29. Liu Y, Shao Z, Teng Y, Hoffmann N. NAM: Normalization-based Attention Module. NeurIPS 2021 Workshop on ImageNet: Past, Present, and Future. 2021. Available online: https://openreview.net/forum?id=AaTK_ESdkjg
30. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:7132-41.
31. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, 13-19 June 2020. 2020:11534-42.
32. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A,

Landman B, Roth HR, Xu D. Unetr: Transformers for 3d medical image segmentation. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022:574-84.

33. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. In: Crimi A, Bakas S, editors. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Cham: Springer International Publishing; 2021:272-84.
34. Zhou HY, Guo J, Zhang Y, Han X, Yu L, Wang L, Yu Y. nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer. IEEE Trans Image Process 2023;32:4036-45.
35. Ke S, Duan H, Cai Y, Kang J, Feng Z. Thoracoscopy-assisted minimally invasive surgical stabilization of the anterolateral flail chest using Nuss bars. Ann Thorac Surg 2014;97:2179-82.
36. Urbaneja A, De Verbizier J, Formery AS, Tobon-Gomez C, Nace L, Blum A, Gondim Teixeira PA. Automatic rib cage unfolding with CT cylindrical projection reformat in polytraumatized patients for rib fracture detection and characterization: Feasibility and clinical application. Eur J Radiol 2019;110:121-7.
37. Sollmann N, Mei K, Hedderich DM, Maegerlein C, Kopp FK, Löffler MT, Zimmer C, Rummeny EJ, Kirschke JS, Baum T, Noël PB. Multi-detector CT imaging: impact of virtual tube current reduction and sparse sampling on detection of vertebral fractures. Eur Radiol 2019;29:3606-16.

Cite this article as: Liu Y, Zhang L, Jiang Z. Multi-stream and multi-scale fusion rib fracture segmentation network based on UXNet. Quant Imaging Med Surg 2025;15(1):230-248. doi: 10.21037/qims-24-1356