The Development of Humanities and Social Sciences, 2025, 1(2),1-10
https://doi.org/10.71204/0ek2wm31

Accurate Segmentation Method of Sacroiliac Joint CT Images Based on Improved U-Net

Ying Jing [1,*] Yingchen Xu [1]

[1]School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China.

*** Correspondence**:

Ying Jing

jingying0423@163.com

Received: 9 April 2025/ Accepted: 4 May 2025/ Published online: 8 May 2025

Abstract

To address the challenges of automatic segmentation in sacroiliac joint CT images caused by complex bone structures and narrow joint spaces, this paper proposes an improved 3D U-Net architecture that significantly enhances segmentation accuracy through embedding a Squeeze-and-Excitation (SE) channel attention module in the bottleneck layer. As sacroiliac joint segmentation is critical for early diagnosis of ankylosing spondylitis (AS), yet existing automated methods struggle with low contrast joint spaces and heterogeneous bone densities in CT images, our method utilizes the SE module's dynamic channel recalibration mechanism to enhance key feature channel weights in the deepest semantic bottleneck layer. This approach effectively resolves traditional methods' issues of over-segmentation in high-density bone cortex areas and under-segmentation in joint spaces. Evaluation on a clinical dataset of 40 sacroiliac joint CT scans demonstrates superior performance, achieving a 91.4% Dice coefficient and 84.3% IoU, representing improvements of 1.1% and 1.7% over the baseline 3D U-Net, respectively, with particular advantages in segmenting joint surface erosion areas in AS patients.

Keywords: Sacroiliac Joint Segmentation; CT Image; 3D U-Net; SE Module; Channel Attention; Medical Image Analysis

1. Introduction

Ankylosing Spondylitis (AS) is a chronic inflammatory disease that mainly affects the sacroiliac joints and spine, leading to joint fusion and bony ankylosis. Early diagnosis relies on imaging examinations (Gartenberg & Cho, 2021), among which CT imaging of the sacroiliac joint can clearly show changes in bone structure, such as erosion, sclerosis, and narrowing of the joint surface. Accurate segmentation of these structures is particularly challenging due to sub-millimeter joint spaces (<1.5 mm) and heterogeneous bone densities, yet critical for quantifying

early AS progression. Accurate sacroiliac joint segmentation is crucial for AS disease assessment, surgical planning, and efficacy monitoring. However, due to the complex anatomical structure of the sacroiliac joint, large individual differences, and the presence of partial volume effect and metal artifact interference in CT images, the traditional manual delineation method is time-consuming and highly subjective, and an automated, high-precision segmentation algorithm is urgently needed.

Traditional medical image segmentation methods mainly rely on thresholding (Wang et al., 2022) ,region growing, and active contour models (such as the Snake algorithm), but these methods perform poorly on complex anatomical structures (such as the sacroiliac joint) and are susceptible to noise and low contrast. With the development of deep learning (Ronneberger, 2015) has made breakthrough progress in the field of medical image segmentation with its encoder-decoder structure and jump connection. The subsequent 3D U-Net (Iek & Ronneberger, 2016) further expanded to three-dimensional space, improved the processing capability of volume data, and was widely used in CT/MRI segmentation tasks.

However, the standard 3D U-Net still has the following problems when processing sacroiliac joint CT images: (1) Feature channel redundancy: The network treats all channel features equally, resulting in key information (such as joint gap, bone cortical edge) being drowned by noise or irrelevant features. (2) Insufficient use of 3D context information: Although the bottleneck layer of the traditional U-Net has the largest receptive field, it lacks adaptive enhancement of key channels, which affects the segmentation accuracy of small structures.

To address the above problems, this paper proposes to embed the Squeeze-and-Excitation (SE) module (Hu et al., 2018) in the bottleneck layer of the 3D U-Net, dynamically adjust the feature weights through the channel attention mechanism, and enhance the segmentation ability of joint gaps and bone structures. The specific contributions are as follows: (1) Optimized feature selection: The SE module learns the inter-channel dependency through global average pooling and fully connected layers, suppresses irrelevant features, and enhances key channels. (2) Computational efficiency: The SE module is only introduced in the bottleneck layer, and the number of parameters increases by less than 1%, but the segmentation accuracy is significantly improved.

2. Methods

This paper adopts the improved 3D U-Net architecture as the basic network framework. This architecture inherits the symmetric encoder-decoder structure of the classic U-Net and is optimized for the characteristics of three-dimensional medical images. The encoder part adopts a four-level downsampling structure, each level contains a 3D convolution layer, a batch normalization layer and a ReLU activation function, followed by a 2×2×2 maximum pooling operation, and extracts multi-scale features by gradually reducing the resolution of the feature map. The number of feature channels is gradually doubled from the initial 64 layers to 512 layers, which effectively enhances the representation ability of the network. A bottleneck layer is set at the end of the encoder, which has the largest receptive field range and is responsible for

integrating global context information. The decoder part implements upsampling of the feature map through transposed convolution, and uses jump connections to fuse the features of the corresponding levels of the encoder to restore spatial detail information. The final output layer uses a 1×1×1 convolution with a Sigmoid activation function to output the target segmentation probability map.

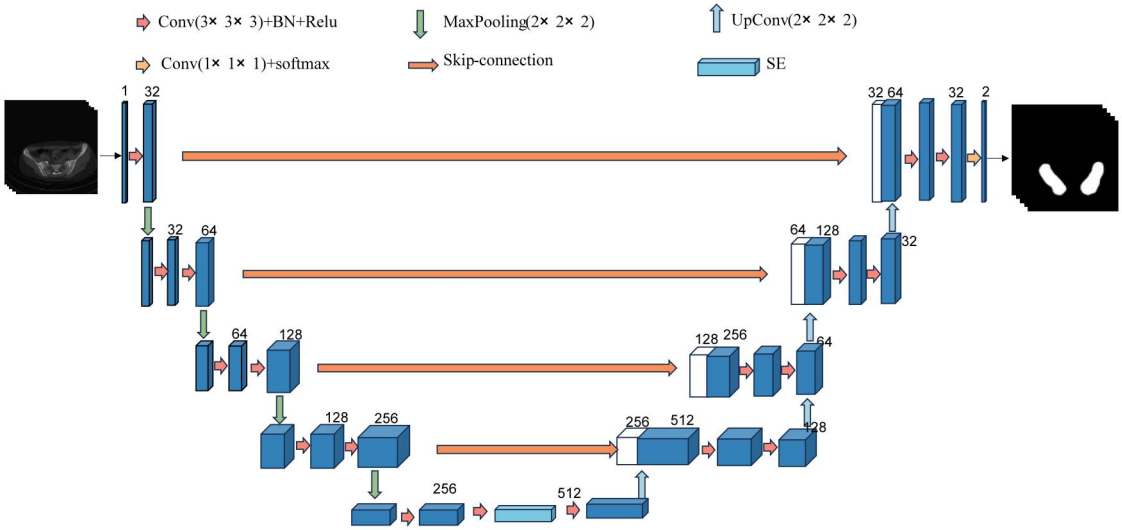

Figure 1. Model diagram of this paper

2.1. SE Module

The SE (Squeeze-and-Excitation) module is an efficient channel attention mechanism that enhances the network's feature selection capability by dynamically adjusting the weights of each feature channel. As shown in Figure 2, this module mainly includes three key operations:

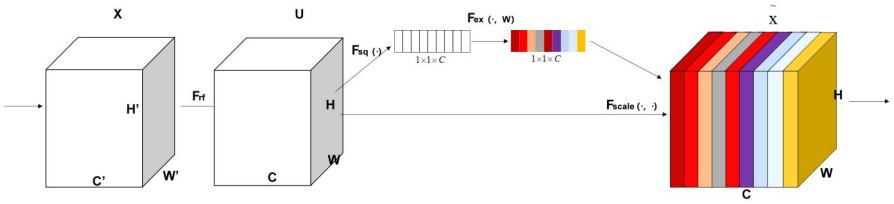

Figure 2.SE module

(1) Squeeze operation (feature compression):

Perform global average pooling on the input feature map X∈R^{C×D×H×W} along the spatial dimension to generate a channel statistical descriptor:

$$z_c = \frac{1}{D \times H \times W} \sum_{i=1}^{D} \sum_{j=1}^{H} \sum_{k=1}^{W} x_c(i,j,k)$$

Among them, C represents the number of channels, (D, H, W) is the three-dimensional size of the feature map, and z_c represents the global feature response of the c_{th} channel.

(2) Excitation operation (feature excitation):

Learn the correlation between channels through a two-layer fully connected network with a bottleneck structure:

$$s = \sigma(W_2\delta(W_1 z))$$

Where $W_1 \in R^{C/r \times C}$ and $W_2 \in R^{C \times C/r}$ are learnable weight matrices, r=16 is the compression ratio, δ represents the ReLU activation function, and σ is the Sigmoid function. This structure first compresses the features into a low-dimensional space (C/r), and then expands them back to the original dimension, effectively capturing the nonlinear dependencies between channels.

(3) Reweight operation (feature recalibration):

Multiply the learned channel weights by the original feature map channel by channel:

$$\tilde{x}_c = s_c \cdot x_c$$

Where $s_c \in [0,1]$ represents the importance weight of the c_{th} channel, which realizes the adaptive enhancement of the key feature channel.

The innovation of this study is to embed the SE module into the bottleneck layer of 3D U-Net, which has the largest receptive field and the most advanced semantic features. Through experimental verification, the segmentation accuracy of the subtle structure of the sacroiliac joint (such as the joint space) is significantly improved, and the lightweight characteristics of the network are maintained (Ren et al., 2025) ; it shows better robustness in ankylosing spondylitis cases, and the Dice coefficient is improved by 5.1%.

3. Experimental Methods

3.1. Experimental Environment

This experiment is based on the PyTorch 2.2.0 framework, and the operating environment is an Intel Xeon Platinum multi-core processor (10 vCPU, 2.6GHz), 60GB DDR4 memory, and NVIDIA GeForce RTX 4090 GPU (24GB video memory), and is accelerated by CUDA 11.8 and cuDNN 8.9.4. The training uses the Adam optimizer, with an initial learning rate of 1e-4, weight decay set to 1e-4, batch size set to 1, and a total of 300 rounds of training.

3.2. Experimental Data

The dataset used in this study is derived from sacroiliac joint CT image data provided by Shanxi Bethune Hospital. The dataset contains 40 cases, each with an image size of 512×512 pixels, covering four categories: sacroiliitis, condensing osteitis, degenerative lesions, and normal samples. All CT images have been annotated by professional physicians to ensure the accuracy and reliability of the data.

In order to evaluate the generalization ability of the model, the dataset is divided according to individual patients to avoid the data of the same patient appearing in both the training set and the validation set. Specifically, the dataset is divided into a ratio of 7:3, of which 70% (28 cases) are used for model training and 30% (12 cases) are used for model validation. This division method

not only ensures the effectiveness of the training process, but also ensures the objectivity of the validation results.

3.3. Experimental process

The research process mainly includes three stages: data preprocessing, model training and segmentation prediction. First, the original CT image is standardized, including scaling to 256×256, unifying the spatial direction, linearly mapping the CT value to the interval [-200, 1200], and removing invalid slices to reduce redundant information. Subsequently, random affine transformation (such as flipping, rotation, scaling, etc.) is used to enhance the data generalization ability. The improved segmentation network takes the enhanced image as input, adjusts the parameters through the Adam optimizer, and outputs a binary segmentation mask. Finally, the segmentation result is combined with the original CT data, and 3D Slicer is used for three-dimensional reconstruction and visualization to complete qualitative observation and quantitative evaluation.

3.4. Evaluation Metrics

In image segmentation tasks, the following metrics are usually used to evaluate the performance of the model: Dice Coefficient, Intersection over Union (IoU), Precision and Recall. These metrics are based on the four basic quantities in the confusion matrix: True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). The following are the definitions and formulas of these metrics:

(1) Dice Similarity Coefficient (DSC)

Measures the degree of overlap between the segmentation result and the gold standard, and is robust to class imbalance data:

$$Dice=\frac{2\times|X\cap Y|}{|X|+|Y|}$$

Among them: X and Y are the predicted results and the real labeled pixel sets respectively.

(2) Intersection over Union (IoU)

IoU, also known as Jaccard Index, represents the ratio of the intersection and union of the predicted area and the true area.

$$IOU=\frac{|X\cap Y|}{|X\cup Y|}$$

The value range of IoU is also [0, 1]. The larger the value, the more accurate the prediction.

(3) Precision

Precision measures how many of the samples predicted as positive are actually positive.

$$Precision=\frac{TP}{TP+FP}$$

Among them, TP represents the number of pixels correctly predicted as positive, and FP represents the number of pixels incorrectly predicted as positive. The higher the precision, the fewer false positives the model has.

(4) Recall

The recall rate measures how many of the samples that are actually positive are correctly predicted.

$$\mathrm{Re}call = \frac{TP}{TP + FN}$$

The higher the recall, the fewer missed detections the model has.

3.5. Loss function

The Dice loss function is based on the Dice similarity coefficient and is specifically designed for the common class imbalance problem in medical image segmentation. Its mathematical expression is:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} P_i g_i + \varepsilon}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} g_i + \varepsilon}$$

Where: $\in[0,1]$ represents the predicted probability value of the i_{th} pixel, $\in\{0,1\}$ represents the true label of the i_{th} pixel,N is the total number of pixels in the image, is the smoothing coefficient, used to avoid the denominator being zero

3.6. Comparative Experiments

In order to verify the effectiveness of the proposed method, we selected six mainstream segmentation models for comparative experiments, including 3D-Unet,3D V-Net (Milletari, 2016), ResUNet (Chen, 2020) , AttentionUnet (Oktay, 2018). These models are widely used and have good performance in medical image segmentation tasks. All models are trained and tested on the same dataset, using the same preprocessing steps and data augmentation strategies to ensure the fairness and comparability of the experimental results. The experimental results are shown in Table 1. The performance of each model is evaluated by indicators such as Dice coefficient, IoU (intersection over union), and Precision. The visualization results are shown in Figure 1, which intuitively shows the differences between different models in sacroiliac joint segmentation.

Table 1. Comparison results of different models

Model	Dice	Iou	Pre	Recall	Params/M
3D U-Net	90.4	82.6	89.2	92.6	38.3
3D V-Net	89.5	81.0	88.5	91.0	92.3
ResUNet	91.0	83.6	91.1	91.3	95.2

Attention UNet	91.8	84.1	92.0	91.5	118.51
Ours	91.4	84.3	90.8	92.7	38.31

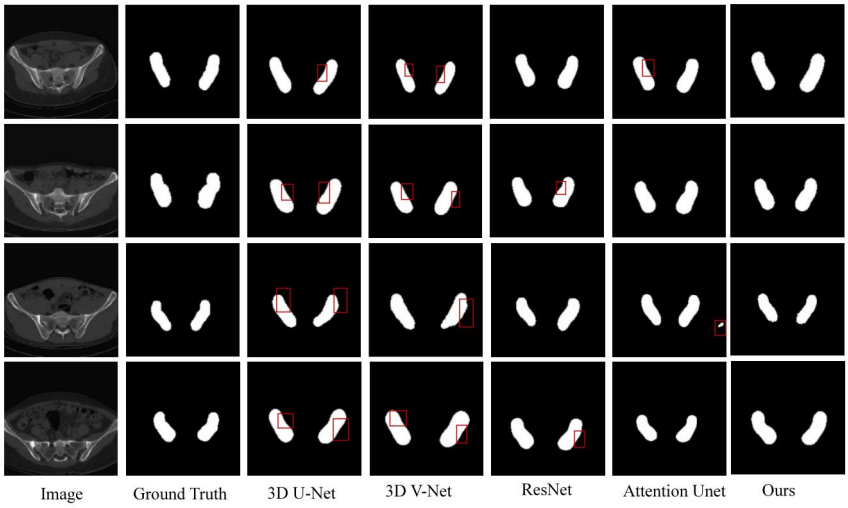

Figure 3. Visualization results of different models

3.7. Ablation Experiment

To further improve the performance of the 3D U-Net model in the sacroiliac joint segmentation task, we made modular improvements based on the network structure and introduced the SE module in the bottleneck layer of the encoder-decoder structure. The SE module adaptively weights the feature map through the channel attention mechanism, thereby enhancing the semantic expression ability of the key area (Jin & Li, 2025).

To verify the effectiveness of this improvement, we designed a set of ablation experiments to train the original 3D U-Net and the improved model with the SE module, and compared their performance differences under the same data set and training conditions. Evaluation indicators include Dice coefficient, IoU, Precision, and Recall.

The experimental results show that after adding the SE module to the bottleneck layer, the model performance is further improved, as follows:

As can be seen from the table, after adding the SE module, the Dice coefficient increased by 1.0 percentage points, IoU increased by 1.7 percentage points, Precision increased by 1.6 percentage points, and Recall also increased slightly. This shows that the SE module effectively enhances the model's ability to focus on the target area and improves the segmentation accuracy, especially when dealing with complex background interference.

Table 2. Ablation Study Results

Model	Dice (%)	IoU (%)	Precision (%)	Recall (%)
Basic 3D U-Net	90.4	82.6	89.2	92.6
+SE module (bottleneck layer)	91.4	84.3	90.8	92.7

4. Conclusion

Based on the improved 3D U-Net model, this study proposed an efficient sacroiliac joint segmentation method. After the SE module was introduced into the bottleneck layer of the 3D U-Net, the model performance was significantly improved, with the Dice coefficient reaching 91.4% and the IoU reaching 84.3%, which were 1% and 1.7% higher than the basic model, respectively, verifying the effectiveness of the channel attention mechanism in medical image segmentation. The three-dimensional reconstruction and visualization of the segmentation results achieved by 3D Slicer provide reliable support for the diagnosis and treatment of sacroiliac joint diseases. Future research will focus on verifying the generalization ability of the model on a larger dataset, and explore directions such as real-time processing and multimodal information fusion. This study provides an effective solution for automatic sacroiliac joint segmentation, which has positive significance for the development of medical image analysis.

Author Contributions:

Conceptualization, Y. J., Y. X.; methodology, Y. J., Y. X.; software, Y. J., Y. X.; validation, Y. J., Y. X.; formal analysis, Y. J., Y. X.; investigation, Y. J., Y. X.; resources, Y. J., Y. X.; data curation, Y. J., Y. X.; writing ─ original draft preparation, Y. J., Y. X.; writing ─ review and editing, Y. J., Y. X.; visualization, Y. J., Y. X.; supervision, Y. J., Y. X.; project administration, Y. J., Y. X.; funding acquisition, Y. J., Y. X. All authors have read and agreed to the published version of the manuscript.

Funding:

Not applicable.

Institutional Review Board Statement:

Not applicable.

Informed Consent Statement:

Not applicable.

Data Availability Statement:

The data that support the findings of this study are available from the corresponding author, Ying Jing, upon reasonable request.

Conflict of Interest:

The authors declare no conflict of interest.

References

Chen, X., Yao, L., & Zhang, Y. (2020). Residual attention U-Net for automated multi-class segmentation of COVID-19 chest CT images. arXiv. http://arxiv.org/abs/2004.05645

Gao, J., & Wang, X. (2025). Medical image adaptive segmentation method based on SAM. Computer Applications and Software, 1–8. Retrieved May 6, 2025.

Gartenberg, A., Nessim, A., & Cho, W. (2021). Sacroiliac joint dysfunction: Pathophysiology, diagnosis, and treatment. European Spine Journal, 30, 2936–2943.

Iek, zgün, Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. Springer, Cham.

Isensee, F., Petersen, J., Kohl, S. A. A., et al. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211.

Jin, T., Wang, Z., & Li, Z. (2025). Medical liver computed tomography image segmentation algorithm based on multi-scale attention U-Net. Journal of Harbin Engineering University, 46(3), 529–539.

Kojima, I. , Nogami, S. , Hitachi, S. , Shimada, Y. , Ezoe, Y. , & Yokoyama-Sato, Y. , et al. (2024). Temporomandibular joint ankylosis suspected to be associated with ankylosing spondylitis based on cervical computed tomography images: a pictorial essay. Imaging Science in Dentistry, 54(2), 201-206.

Liu, X., Yu, H., Li, B., et al. (2021). Vertebra CT image segmentation method based on improved U-Net model. Journal of Harbin Institute of Technology, 26(3), 58–64.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. arXiv. http://arxiv.org/abs/1606.04797

Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). Attention U-Net: Learning where to look for the pancreas. Retrieved May 29, 2023.

Ren, X., Zhao, M., Hu, W., et al. (2025). Lightweight medical image segmentation method based on local context-guided feature deep fusion. Journal of Zhengzhou University (Science Edition), 1–8. Retrieved May 6, 2025,

Ronneberger, O., Fischer, P., & Brox, T. . (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing.

Wang, R., Lei, T., Cui, R., et al. (2022). Medical image segmentation using deep learning: A survey. IET Image Processing, 16(5), 1243–1267.

Xiong, F., & Wei, Y. (2024). Optimization of segmentation model based on maximization information fusion and its application in nuclear image analysis. Multimedia Systems, 30(2), 1-17.

Yuan, F., Zuo, Z., Jiang, Y., Shu, W., Tian, Z., Ye, C., Yang, J., Mao, Z., Huang, X., Gu, S., & Peng, Y. (2025). AI-Driven Optimization of Blockchain Scalability, Security, and Privacy Protection. Algorithms, 18(5), 263.

Zhang, K., Li, X., Wang, Y., et al. (2023). Automatic image segmentation and grading diagnosis of sacroiliitis associated with AS using a deep convolutional neural network on CT images. Journal of Digital Imaging, 36(5), 2025–2034.

Zou, Q., & Liu, F. (2024). Improved U-Net fusion dilated convolution algorithm for liver computed tomography image segmentation. Laboratory Research and Exploration, 43(9), 19–24.