

# BSM-SegNet: Boundary-Scale Synergistic Segmentation Network for Lumbar Muscle MRI Segmentation in Sarcopenia Assessment

Junjie Dang<sup>1,4</sup>, Huanqing Cui<sup>2</sup>, Yining Tang<sup>3</sup> and Ruixia Liu<sup>4</sup>

<sup>1</sup> School of Mathematics and Statistics, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

<sup>2</sup> College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

<sup>3</sup> IEIT SYSTEMS Co., Ltd., Jinan, 250101

<sup>4</sup> Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China

10431241345@stu.qlu.edu.cn, cuihq@sdust.edu.cn,  
tangyinan@ieisystem.com, liurx@sdas.org<sup>(✉)</sup>

**Abstract.** Accurate segmentation of lumbar muscle groups is essential for imaging-based screening and quantitative assessment of sarcopenia. However, due to unclear boundaries, adhesion of adjacent muscles, scale differences between categories, and the presence of artifacts, automatic segmentation of lumbar muscles in magnetic resonance imaging (MRI) remains challenging. To tackle these inherent difficulties, this study introduces a boundary-scale collaborative perception architecture designed for multi-class lumbar muscle segmentation, which we refer to as the Boundary-Scale Synergistic Segmentation Network (BSM-SegNet). The proposed network integrates boundary-sensitive modeling, multi-scale feature representation, and cross-stage feature selection to enhance segmentation accuracy and robustness. Specifically, a Boundary-Sensitive Residual Module (BSRM) is designed to strengthen feature responses in muscle boundary regions through local feature differencing. A Multi-Scale Coupling Module (MSCM) is introduced to improve multi-scale structural modeling via multi-scale pooling and channel attention. In addition, a Cross-Stage Gating (CSG) module is employed to suppress redundant features and enhance semantic consistency during encoder-decoder feature fusion. Extensive experiments are conducted on a private lumbar muscle MRI dataset with 562 subjects and an external multi-center public dataset with 290 subjects. Experimental results show that BSM-SegNet achieves an IoU of 86.68% and a Dice score of 92.85% on the private dataset, and 86.16% IoU and 92.50% Dice on the external dataset, outperforming several representative segmentation methods. Ablation studies further demonstrate the effectiveness and synergistic benefits of the proposed modules. These results indicate that BSM-SegNet is well suited for accurate lumbar muscle segmentation in sarcopenia assessment.

**Keywords:** Sarcopenia, Lumbar muscle segmentation, Magnetic resonance imaging, Deep learning, Multi-class segmentation.

## 1 INTRODUCTION

Sarcopenia is a prevalent age-related disorder involving progressive skeletal muscle loss and dysfunction [1]. Quantitative muscle imaging is essential for sarcopenia diagnosis and monitoring. As lumbar muscles govern spinal stability and trunk motion, precise lumbar muscle segmentation is essential for reliable sarcopenia assessment.

Thanks to its exceptional soft-tissue contrast and radiation-free imaging properties, MRI represents an optimal modality for evaluating lumbar muscle morphology [2]. Nevertheless, automatic MRI lumbar muscle segmentation remains challenging with existing CNN methods [3], due to vague muscle boundaries, muscle adhesion, scale variation, MRI artifacts, and semantic inconsistency from conventional skip connections.

Deep learning methods, especially U-Net variants, dominate lumbar muscle segmentation via multi-resolution feature learning [4,5]. Existing strategies adopt attention mechanisms [6], multi-scale and CNN–Transformer structures [7], and lightweight designs [8]. However, they suffer from poor boundary modeling [9], limited multi-scale representation [7], and inefficient feature fusion [10], restricting segmentation performance on complex lumbar muscles.

This work proposes BSM-SegNet for multi-class lumbar muscle segmentation. It enhances boundary perception, multi-scale feature representation and cross-stage feature fusion, enabling accurate and robust segmentation for sarcopenia analysis.

The main contributions of this study are as follows:

- A BSRM is proposed to enhance feature discrimination in muscle boundary regions and reduce boundary ambiguity between adjacent muscles.
- A MSCM is introduced to strengthen multi-scale feature representation via multi-scale pooling and channel attention.
- A CSG module is designed to suppress redundant information and improve semantic consistency during encoder–decoder feature fusion.
- Comprehensive evaluations conducted on both an in-house clinical dataset and an external multi-center dataset validate the efficacy and cross-dataset generalization performance of BSM-SegNet in performing fine-grained segmentation of lumbar muscles.

## 2 RELATED WORK

### 2.1 Lumbar Muscle Segmentation in MRI

Early approaches for lumbar muscle segmentation mainly relied on atlas-based algorithms and conventional image processing techniques, including intensity thresholding, region growing, and deformable models. Though equipped with interpretable anatomical prior knowledge, these methods are susceptible to intensity inhomogeneity and inter-subject anatomical differences, resulting in poor robustness and limited scalability for large-scale clinical use.

Fueled by the rapid advancement of deep learning, CNNs, especially U-Net and its variants, have dominated MRI-based lumbar muscle segmentation [5]. An encoder-decoder architecture with skip connections can effectively integrate low-level spatial details and high-level semantic features, significantly outperforming traditional methods. Existing studies have verified the reliability of CNN-based frameworks in automatic lumbar and trunk muscle segmentation, supporting clinical evaluation and musculoskeletal analysis [3,11]. Furthermore, residual learning mechanisms and dense connectivity patterns are incorporated to enhance feature reuse efficiency and facilitate stable gradient flow. This design choice ultimately elevates segmentation performance, particularly for morphologically intricate muscle structures [12].

## 2.2 Attention and Multi-Scale Feature Modeling

Attention mechanisms and multi-scale feature learning methods have been widely studied to address the challenges of unclear muscle boundaries and significant variations in muscle scale. Channel and spatial attention modules can highlight valid feature regions and suppress background interference, boosting segmentation performance in low-contrast MRI scenes [6,13]. Meanwhile, multi-scale structures including pyramid pooling, atrous spatial pyramid pooling (ASPP) and dilated convolution are capable of capturing multi-scale muscle features and adapting to large scale variations among different muscle tissues [7,14].

Recently, hybrid CNN–Transformer models have been developed to integrate local texture extraction and global context modeling. Despite their strong capability in modeling long-range dependencies, such methods typically come with high computational overhead and lack explicit boundary learning modules, which limits their ability to accurately separate adjacent lumbar muscles with ambiguous boundaries.

## 2.3 Boundary-Aware and Feature Fusion Strategies

Boundary-aware learning is a key research direction in medical image segmentation. Current methods introduce auxiliary edge detection branches, boundary supervision and boundary-aware loss functions to optimize network boundary localization capability [15]. However, these approaches usually require extra annotation costs and complex multi-task learning designs, which restrict their practical clinical deployment. In classic encoder–decoder networks, skip connections undertake core feature fusion tasks. However, direct concatenation of features tends to introduce redundant low-level information, leading to semantic ambiguity during the decoding process. To solve this problem, recent studies adopt gating mechanisms and selective fusion strategies to adaptively control the information transmission between encoder and decoder.

## 2.4 Motivation of This Work

In contrast to existing approaches, BSM-SegNet is designed from a synergistic perspective, explicitly integrating boundary-sensitive modeling, multi-scale structural

perception, and cross-stage gated feature fusion within a unified architecture. This design directly targets the intrinsic challenges of lumbar muscle MRI segmentation, enabling more accurate delineation of complex muscle boundaries and robust segmentation across diverse spatial scales.

### 3 METHOD

The proposed BSM-SegNet adopts an encoder–decoder U-shaped architecture to jointly enhance boundary sensitivity, multi-scale structural representation, and efficient feature fusion for lumbar muscle MRI segmentation. As illustrated in Fig. 1, we design a network that integrates an encoder and a decoder with three specialized modules: BSRM, MSCM, and CSG. Through a hierarchical learning process, the encoder captures features at different semantic levels, ranging from detailed texture information to high-level contextual representations. In contrast, the decoder gradually recovers spatial resolution to produce pixel-level segmentation outputs. In BSM-SegNet, the BSRM is embedded in each encoder stage to enhance boundary representation, the MSCM is incorporated into deeper encoder stages to strengthen multi-scale modeling, and the CSG module is applied to skip connections to selectively fuse encoder–decoder features and suppress redundancy. To minimize computational redundancy, the network accepts a single-channel axial lumbar MRI slice with a resolution of  $256 \times 256$  pixels as its input. A multi-class segmentation map is produced to identify six distinct lumbar muscle groups, namely the bilateral psoas major, erector spinae, and multifidus muscles. At the network output, a Softmax operation converts the predictions into class-wise probability maps, which support downstream morphological measurements.

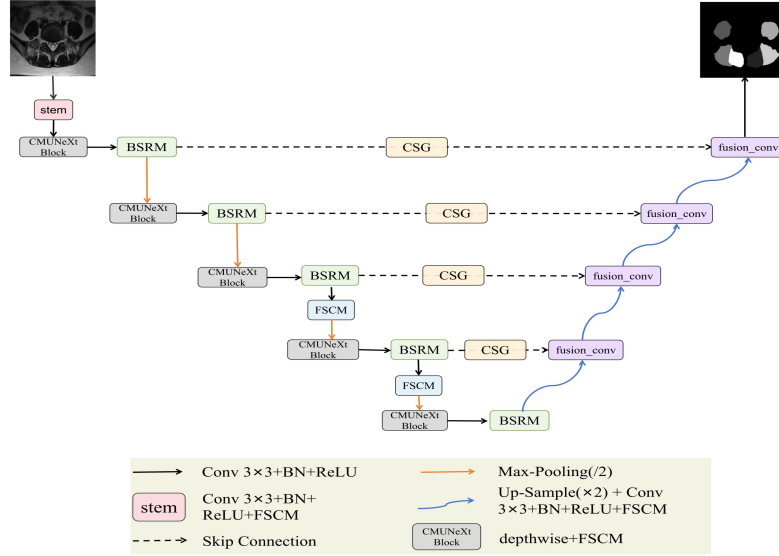
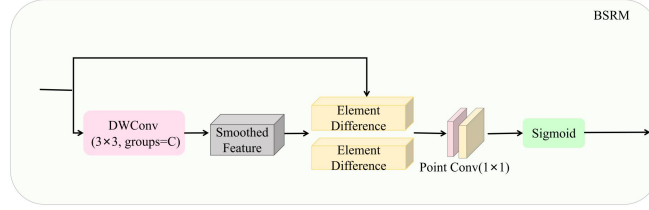


Fig. 1. Overall architecture of BSM-SegNet.

### 3.1 Boundary-Sensitive Residual Module (BSRM)

Lumbar muscle MRI images often exhibit indistinct boundaries and subtle local intensity variations, which significantly hinder accurate segmentation. To address this issue, BSM-SegNet incorporates a BSRM as a core component for boundary enhancement. The structure of the BSRM is detailed in **Fig. 2**. The BSRM explicitly emphasizes muscle boundary regions by leveraging residual differencing between original features and locally smoothed features.



**Fig. 2.** Structure of the BSRM module.

Specifically, the BSRM first applies a  $3 \times 3$  depthwise convolution to the input feature map to obtain locally smoothed response features. The residual difference between the original features and the smoothed features is then computed, which amplifies spatially variant regions corresponding to muscle boundaries while suppressing homogeneous interior regions. Subsequently, a Sigmoid-based gating mechanism is employed to generate adaptive weights that are used to reweight the original features. This process enhances the network's sensitivity to muscle boundaries and fine structural details while reducing redundant responses within uniform regions, thereby providing more discriminative features for subsequent multi-scale modeling and feature fusion.

The mathematical formulation of the BSRM is defined as follows:

$$\mathbf{b} = \text{DWConv}_{3 \times 3}(\mathbf{x}), \quad (1)$$

$$\Delta \mathbf{b} = |\mathbf{b} - \mathbf{x}|, \quad (2)$$

$$\mathbf{g} = \sigma(\text{PWConv}_{1 \times 1}(\Delta \mathbf{b})), \quad (3)$$

$$\mathbf{y} = \mathbf{x} \otimes (1 + \mathbf{g}), \quad (4)$$

where  $\mathbf{x} \in \mathbb{R}^{B \times C \times H \times W}$  denotes the input feature map of the BSRM, with  $B$ ,  $C$ ,  $H$  and  $W$  representing the batch size, number of channels, height, and width, respectively.  $\mathbf{b}$  represents the locally smoothed response extracted by depthwise convolution, and  $\Delta \mathbf{b}$  denotes the residual boundary response obtained by feature differencing.  $\mathbf{y}$  is the output feature of the module. DWConv and PWConv denote depthwise and pointwise convolutions, respectively,  $\sigma(\cdot)$  is the Sigmoid activation function, and  $\otimes$  indicates element-wise multiplication. The gating weight  $\mathbf{g}$  adaptively emphasizes boundary-sensitive regions.

By adopting a residual connection design, the BSRM effectively mitigates gradient vanishing while preserving the original semantic information of the features, ensuring that boundary enhancement does not disrupt the overall feature representation.

### 3.2 Frequency–Scale Coupling Module (MSCM)

To bolster BSM-SegNet’s ability to capture lumbar muscle features across different spatial resolutions, we design a Multi-Scale Context Module (MSCM) and embed it within the encoder. Fig. 3 presents the detailed structure of the MSCM. This module functions as the key building block for multi-scale feature modeling, by combining multi-scale pooling operations with channel-wise attention mechanisms. This design allows the network to adaptively capture the scale differences that exist between different lumbar muscle groups.

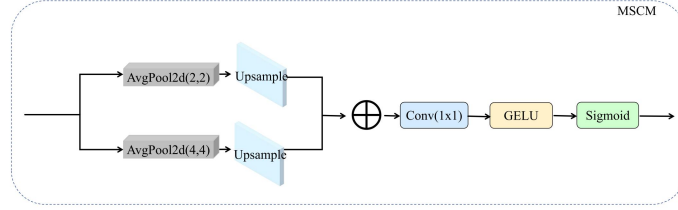


Fig. 3. Structure of the MSCM module.

Specifically, the MSCM module leverages a multi-scale average pooling scheme to model encoder features at two distinct spatial scales. The  $2 \times 2$  pooling operation captures high-frequency local cues, including fine boundaries and textures of small muscles, whereas the  $4 \times 4$  pooling focuses on low-frequency global structures, such as the overall morphology of larger muscle groups. Subsequently, these pooled feature representations are upsampled back to their original spatial dimensions and merged with the input feature map. This mechanism allows for the synergistic fusion of high-resolution local details and low-resolution global contextual cues.

The resulting multi-scale fused features are then refined through a channel attention mechanism. This component generates learnable, channel-specific weights, which are then used to recalibrate the original feature map. This adaptive modulation enables BSM-SegNet to dynamically emphasize scale-relevant features according to the characteristics of different muscles, thereby enhancing segmentation performance, especially for small and spatially complex muscle structures.

The mathematical formulation of the MSCM is given as follows:

$$S = \text{Up}(\text{AvgPool}_{2 \times 2}(x), (H, W)), \quad (5)$$

$$I = \text{Up}(\text{AvgPool}_{4 \times 4}(x), (H, W)), \quad (6)$$

$$f = \text{Concat}(S, I), \quad (7)$$

$$w = \sigma(\text{GELU}(\text{Conv}_{1 \times 1}(f))), \quad (8)$$

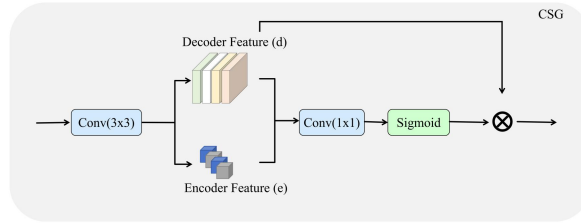
$$y = x \otimes (1 + w), \quad (9)$$

where  $x \in \mathbb{R}^{C \times H \times W}$  denotes the input feature map of the MSCM, with  $C$ ,  $H$  and  $W$  representing the number of channels, height, and width, respectively. AvgPool denotes average pooling, and  $\text{Up}(\cdot)$  represents bilinear upsampling to the target spatial resolution ( $H$ ,  $W$ ).  $S$  corresponds to the small-scale detail features obtained via  $2 \times 2$  pooling, while  $I$  represents the large-scale structural features captured by  $4 \times 4$  pooling.  $\text{Concat}(\cdot)$  denotes channel-wise concatenation, and  $f$  is the fused multi-scale feature representation.  $\text{Conv}_{1 \times 1}$  indicates a pointwise convolution, and  $\text{GELU}(\cdot)$  and  $\sigma(\cdot)$  denote the GELU and Sigmoid activation functions, respectively.  $w$  represents the adaptive channel-wise weighting coefficients, and  $y$  is the output feature map of the MSCM.

By coupling multi-scale pooling with channel attention, the MSCM enables efficient multi-scale feature modeling without introducing excessive computational overhead, allowing adaptive representation of lumbar muscles with pronounced scale variations and enhancing the multi-scale segmentation capability of BSM-SegNet.

### 3.3 Cross-Stage Gating Module (CSG)

Traditional U-shaped networks often introduce redundant information when fusing encoder and decoder features through skip connections, which degrades semantic consistency and negatively affects segmentation performance. To address this issue, BSM-SegNet incorporates a CSG module to suppress redundant interference during cross-stage feature fusion. The CSG module's internal structure is depicted in Fig. 4.



**Fig. 4.** Structure of the CSG module.

The CSG module utilizes decoder-derived high-level semantic information to adaptively generate weighting coefficients that modulate the corresponding encoder feature maps. The process involves two steps: first, a  $1 \times 1$  convolution aligns the decoder features' channel dimensions, then a Sigmoid activation produces the final gating weights. These weights are subsequently multiplied with the encoder features, enabling selective amplification of semantically relevant information.

The mathematical formulation of the CSG module is presented below:

$$h = \sigma(\text{Conv}_{1 \times 1}(d)), \quad (10)$$

$$\mathbf{e}' = \mathbf{e} \otimes (1 + \mathbf{h}), \quad (11)$$

$$\mathbf{f}_{\text{fusion}} = \text{Concat}(\mathbf{e}', \mathbf{d}), \quad (12)$$

where  $\mathbf{d}$  denotes the upsampled decoder feature,  $\mathbf{h}$  represents the decoder-guided gating weight, and  $\mathbf{e}$  denotes the corresponding encoder feature.  $\mathbf{e}'$  is the gated encoder feature after adaptive filtering, and  $\mathbf{f}_{\text{fusion}}$  denotes the fused cross-stage feature.  $\text{Conv}_{1 \times 1}$  indicates a  $1 \times 1$  convolution,  $\sigma(\cdot)$  in which the Sigmoid function is used as the activation operation, and  $\otimes$  refers to element-wise multiplication.

Through the proposed CSG module, irrelevant feature interference is effectively suppressed, while the semantic consistency of cross-stage feature fusion is significantly enhanced, providing reliable support for the high-precision segmentation performance of BSM-SegNet.

## 4 EXPERIMENTS

### 4.1 Datasets and evaluation metrics

To rigorously assess the performance and generalization ability of the proposed BSM-SegNet, comprehensive experiments were conducted using two representative MRI datasets tailored for lumbar muscle segmentation: a publicly available multi-center dataset and a private clinical dataset constructed specifically for sarcopenia evaluation [3].

The public dataset consists of 290 carefully annotated cases derived from 515 patients with symptomatic low back pain, with all MRI scans obtained using a standard 1.5T MRI system. The private dataset includes 562 cases collected from a tertiary hospital between 2020 and 2024, acquired on a high-resolution 3.0T MRI platform. This dataset comprises 321 male and 241 female participants, with ages ranging from 45 to 88 years. Both datasets were annotated with six-class labels corresponding to bilateral psoas major, erector spinae, and multifidus muscle groups in the lumbar region. All annotations were performed by two experienced radiologists using ITK-SNAP, yielding an inter-rater Dice similarity coefficient greater than 0.93 to ensure high annotation reliability. Each dataset was randomly split, with 70% of the samples allocated for training and the remaining 30% used for validation.

A unified preprocessing pipeline was applied to all data, including conversion from DICOM to PNG format, N4ITK bias field correction to reduce MRI inhomogeneity artifacts, Z-score intensity normalization, and spatial resizing to a uniform resolution of  $256 \times 256$  pixels. To improve the model's robustness and generalization capability, a diverse set of data augmentation strategies was applied exclusively during the training phase, such as random horizontal flipping, rotation, scaling, and contrast adjustment.

The experimental evaluation was carried out using a workstation equipped with an Intel Core i9-12900K CPU, an NVIDIA RTX 3090 graphics card (24 GB VRAM), and 64 GB RAM. The software environment was based on Ubuntu 20.04, Python 3.8, and PyTorch 1.12.0. The network was trained using the AdamW optimizer with an initial learning rate of  $1 \times 10^{-4}$ , combined with a cosine annealing learning rate scheduler to



ensure stable convergence. The batch size was set to 16, and an early stopping mechanism was adopted to monitor validation performance and mitigate overfitting.

To quantitatively evaluate performance, we use six widely accepted medical image segmentation metrics: IoU, Dice coefficient, Sensitivity (SE), Precision (PC), F1 score, and Accuracy (ACC). TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives [16]. Additionally, we report computational complexity metrics—parameter count, FLOPs, and inference speed—to assess the model’s clinical deployment potential.

## 4.2 Comparison with Other UNet-Based Models

To validate the superiority of BSM-SegNet, six mainstream UNet-based medical image segmentation models are selected for comparison, including U-Net [17], Attention U-Net (AttU-Net) [6], UNet++ [18], MobileUViT [8], CMUNeXt [19], and CMUNet. To guarantee a rigorous and unbiased comparison, to ensure consistent conditions, every model in this study was trained and validated with a unified experimental configuration and hyperparameter setup.

The quantitative segmentation outcomes of different comparative models on the private clinical dataset and public multi-center dataset are summarized in Table 1 and Table 2, respectively. As reflected in the comprehensive evaluation metrics, the proposed BSM-SegNet method outperforms all competing mainstream segmentation models on both datasets. It achieves optimal scores across all quantitative evaluation indices, including IoU, Dice coefficient, precision and sensitivity, exhibiting stable and outstanding segmentation performance in different data scenarios. The consistent superiority across both private clinical and external multi-center datasets fully confirms the reliability and generalization ability of our designed model. The experimental findings further verify that the boundary perception optimization and multi-scale collaborative feature fusion strategies in BSM-SegNet can effectively tackle the segmentation difficulties of lumbar muscle tissues, leading to substantial improvements in the accuracy and robustness of multi-class lumbar muscle segmentation tasks.

## 4.3 Results

As seen from the results, BSM-SegNet consistently surpasses all other UNet-based models on both datasets. On the private dataset, it obtains an IoU of 86.68% and a Dice score of 92.85%, whereas on the public dataset, it achieves an IoU of 86.16% and a Dice score of 92.50%. These results demonstrate that BSM-SegNet achieves higher segmentation accuracy and stronger generalization capability in lumbar muscle MRI segmentation tasks [20].

Fig. 5 provides a comprehensive visual comparison of segmentation outputs between the proposed BSM-SegNet and several widely adopted state-of-the-art models. In these visualizations, the red outlines denote the manually annotated ground truth masks, while the cyan outlines correspond to the segmentation predictions generated by each model. As clearly observed, the predictions produced by BSM-SegNet demonstrate the most consistent alignment with the reference annotations, characterized by remarkably

smooth, well-defined, and anatomically accurate boundaries. Notably, the model excels at capturing fine-grained anatomical structures and intricate muscle contours, with no evident signs of under-segmentation, over-segmentation, or inter-muscle adhesion. Conversely, competing models frequently exhibit noticeable boundary deviations, spurious segmentation artifacts, and morphological distortions, particularly at the boundaries of small or adjacent muscle groups. These qualitative findings strongly corroborate the superior performance of BSM-SegNet and highlight its enhanced clinical suitability for accurate and reliable lumbar muscle analysis.

**Table 1.** Comparison results on the private dataset.

| Network                 | IoU(%)       | Dice(%)      | SE(%)        | PC(%)        | F1(%)        | ACC(%)       |
|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CMUNet                  | 86.25        | 92.60        | 92.92        | 92.17        | 92.54        | 97.72        |
| AttU-Net                | 82.60        | 90.40        | 90.23        | 90.64        | 90.42        | 97.23        |
| U-Net                   | 80.81        | 89.33        | 89.44        | 89.35        | 89.39        | 96.88        |
| UNetPlusPlus            | 76.71        | 86.70        | 86.93        | 86.69        | 86.79        | 96.28        |
| MobileUViT              | 82.78        | 90.53        | 91.01        | 89.90        | 90.44        | 97.20        |
| CMUNeXt                 | 84.55        | 91.57        | 91.79        | 91.24        | 91.51        | 97.47        |
| <b>Ours(BSM-SegNet)</b> | <b>86.68</b> | <b>92.85</b> | <b>93.21</b> | <b>92.39</b> | <b>92.79</b> | <b>97.74</b> |

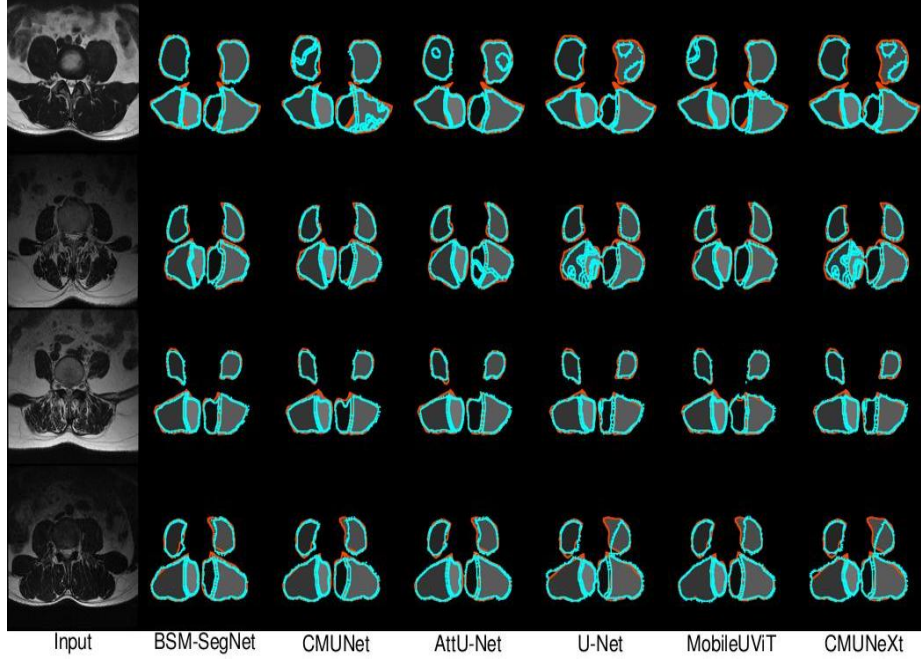
**Table 2.** Comparative performance on the public dataset.

| Network                 | IoU(%)       | Dice(%)      | SE(%)        | PC(%)        | F1(%)        | ACC(%)       |
|-------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CMUNet                  | 80.19        | 88.90        | 88.08        | 89.52        | 88.79        | 97.35        |
| AttU-Net                | 84.14        | 91.30        | 90.72        | 91.73        | 91.22        | 97.86        |
| U-Net                   | 82.11        | 90.09        | 89.51        | 90.65        | 90.08        | 97.66        |
| UNetPlusPlus            | 74.37        | 85.25        | 83.73        | 86.18        | 84.94        | 96.75        |
| MobileUViT              | 77.68        | 87.38        | 85.83        | 88.42        | 87.10        | 96.83        |
| CMUNeXt                 | 83.61        | 91.03        | 89.80        | 92.20        | 90.98        | 97.60        |
| <b>Ours(BSM-SegNet)</b> | <b>86.16</b> | <b>92.50</b> | <b>92.32</b> | <b>92.57</b> | <b>92.44</b> | <b>98.08</b> |

#### 4.4 Ablation study

To evaluate both the standalone performance and the complementary effects of the three core components—BSRM, MSCM, and CSG—we perform ablation studies on

the private dataset, with CMUNeXt [14] serving as the baseline architecture. Adopting a control-variable design, we construct seven distinct experimental setups, including the baseline model with each module added individually, pairwise combinations of two modules, and the full configuration incorporating all three modules. The corresponding results are summarized in Table 3.



**Fig. 5.** Segmentation results of MRI slices from a 68-year-old sarcopenia patient.

Ablation experiments verify that each embedded module can independently boost segmentation accuracy. Among all components, the CSG module achieves the most prominent performance increment, with IoU rising by 0.75% and Dice by 0.42%, revealing that high-level semantic guidance can effectively optimize low-level encoder features. The BSRM improves boundary localization precision, bringing IoU and Dice gains of 0.37% and 0.18% respectively, which verifies its superiority in capturing blurred and intricate muscle boundaries. By contrast, the MSCM provides steady marginal gains via enhancing multi-scale feature modeling ability. Collectively, each module targets different inherent defects of the baseline structure, proving the rationality of our model design [8,11].

In pairwise combination experiments, any two modules can further elevate segmentation metrics, indicating good mutual complementarity and collaboration. In particular, the combination of BSRM and MSCM reaches 85.32% IoU and 92.02% Dice. When BSRM, MSCM and CSG are integrated into the complete BSM-SegNet, the optimal performance is obtained with IoU of 86.68% and Dice of 92.85%. Compared with the baseline CMUNeXt, the corresponding gains reach 2.60% and 1.67%. Such results fully demonstrate the synergistic effect of boundary-aware

modeling, multi-scale context extraction and cross-stage feature adjustment, which jointly tackle the main difficulties in lumbar muscle MRI segmentation, such as serious class imbalance, diverse spatial scales and vague tissue boundaries.

**Table 3.** Ablation experiment results on the private dataset.

| <b>BSRM</b> | <b>MSCM</b> | <b>CSG</b> | <b>IoU(%)↑</b> | <b>DICE(%)↑</b> |
|-------------|-------------|------------|----------------|-----------------|
| No          | No          | No         | 84.09          | 91.32           |
| Yes         | No          | No         | 84.46          | 91.50           |
| No          | Yes         | No         | 84.20          | 91.38           |
| No          | No          | Yes        | 84.84          | 91.74           |
| Yes         | Yes         | No         | 85.32          | 92.02           |
| Yes         | No          | Yes        | 85.17          | 91.96           |
| No          | Yes         | Yes        | 85.21          | 91.97           |
| Yes         | Yes         | Yes        | <b>86.68</b>   | <b>92.85</b>    |

## 5 CONCLUSION

This work introduces BSM-SegNet, a boundary-scale cooperative architecture designed for automatic lumbar muscle MRI segmentation in sarcopenia evaluation. By integrating three purpose-built modules with a weighted loss function, the proposed framework successfully tackles the core obstacles in this task, such as indistinct anatomical boundaries, inter-muscle scale discrepancies, and redundant feature information. Evaluations on both private and publicly available datasets show that BSM-SegNet achieves superior performance over leading segmentation models in both segmentation accuracy and generalization capability. The network can produce anatomically precise, smooth muscle outlines while retaining subtle structural details, thus meeting the clinical standards for robust sarcopenia quantification and morphological analysis. In summary, BSM-SegNet offers an accurate, efficient, and clinically practical solution for lumbar muscle MRI segmentation. Looking ahead, we aim to expand this framework to volumetric (3D) MRI segmentation and explore multi-modal data fusion techniques to further advance sarcopenia diagnosis and evaluation workflows.

## 6 Acknowledgment

This work is supported by the Natural Science Foundation Innovation and Development Joint Fund Project of Shandong Province under Grant NO.ZR2023LZH009.

## References

1. Cruz-Jentoft, A.J., Bahat, G., Bauer, J., et al. Sarcopenia: Revised European consensus on definition and diagnosis. *Age and Ageing*, 48(1), 16–31 (2019).
2. Lundberg, T.R., Fernandez-Gonzalo, R., Gustafsson, T. Muscle imaging techniques: MRI for assessment of skeletal muscle morphology and composition. *European Journal of Applied Physiology*, 118(6), 1125–1136 (2018).
3. Niemeyer, F., Zanker, A., Jonas, R., et al. An externally validated deep learning model for accurate segmentation of the lumbar paravertebral muscles. *European Spine Journal*, 31(8), 2156–2164 (2022).
4. Wesselink, E.O., Elliott, J.M., Coppieters, M.W., et al. Convolutional neural networks for automatic segmentation of lumbar paraspinal muscles in people with low back pain. *Scientific Reports*, 12(1), 13485 (2022).
5. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. In: *Proc. DLMIA/ML-CDS*, pp. 3–11. Springer (2018).
6. Islam, M.Z., Ali, A.B.M.S., Ali, F. Multi-scale deep learning architectures for medical image segmentation: A survey. *IEEE Access*, 9, 43857–43883 (2021).
7. Liu, X., Wang, H., Zhang, Y., et al. MobileUViT: Lightweight U-ViT for efficient medical image segmentation. In: *Proc. ICASSP*, pp. 1235–1239. IEEE (2024).
8. Chaurasia, A., Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In: *Proc. IVMSP*, pp. 1–6. IEEE (2017).
9. Huang, H., Lin, L., Tong, R., et al. UNet 3+: A full-scale connected UNet for medical image segmentation. In: *Proc. ICASSP*, pp. 1055–1059. IEEE (2020).
10. Sudre, C.H., Li, W., Vercauteren, T., et al. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis*, pp. 240–248. Springer (2017).
11. Valentinič, A., Schlaeger, S., Lorbeer, R., et al. Automated segmentation of skeletal muscle at lumbar level using deep learning. *European Radiology*, 29(8), 3910–3919 (2019).
12. Drozdal, M., Vorontsov, E., Chartrand, G., et al. The importance of skip connections in biomedical image segmentation. In: *Proc. DLMIA*, pp. 179–187. Springer (2016).
13. Jetley, S., Lord, N.A., Lee, N., Torr, P.H.S. Learn to pay attention. In: *Proc. ICLR* (2018).
14. Chen, L.-C., Papandreou, G., Kokkinos, I., et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE TPAMI*, 40(4), 834–848 (2018).
15. Kervadec, H., Bouchtiba, J., Desrosiers, C., et al. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67, 101851 (2021).
16. Isensee, F., Jaeger, P.F., Kohl, S.A.A., et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211 (2021).
17. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In: *Proc. MICCAI*, pp. 234–241. Springer (2015).
18. Beaudart, C., McCloskey, E., Bruyère, O., et al. Sarcopenia in daily practice: Assessment and management. *BMC Geriatrics*, 16(1), 170 (2016).
19. Lu, Y., Li, Y., Wang, Z., et al. CMUNeXt: Lightweight context-mining U-Net for fast medical image segmentation. *Computers in Biology and Medicine*, 153, 106228 (2023).
20. Li, X., Chen, Y., Wei, X., et al. Frequency domain feature learning for robust medical image segmentation. *IEEE Transactions on Medical Imaging*, 39(12), 4142–4153 (2020).