

基于SAM2实现CT影像上骨骼肌的少样本分割

宋阿伟¹, 蔡函达¹, 金柯佳²

¹温州大学计算机与人工智能学院, 浙江 温州

²温州医科大学第一附属医院医护中心科室, 浙江 温州

收稿日期: 2025年2月17日; 录用日期: 2025年3月26日; 发布日期: 2025年4月3日

摘要

肌少症对患者的身体健康有诸多不利影响, 而肌量的测量是诊断肌少症的重要手段之一。然而, 现有的手动测量方法效率较低, 限制了其在临床中的广泛应用。文章旨在提出一种适用于肌少症临床诊断和医学研究的自动分割方法, 该方法能够在少量训练数据的情况下实现高精度的骨骼肌分割。我们的方法首先使用少量数据对YOLOv10和SAM2模型进行微调。然后, 通过YOLOv10获取每个样本的矩形框坐标, 并将其作为SAM2的提示词进行骨骼肌分割。实验结果显示, 我们的方法在分割任务中的DSC评分达到了0.9025, 优于传统U-Net架构的0.8763和使用固定矩形区域作为提示词的0.8887。此外, 在测试数据集上的验证结果表明, 85%的分割结果在临床应用的可接受范围内。综上所述, 文章提出的方法在骨骼肌的少样本自动分割任务中表现出较高的精度和可靠性, 具有潜在的临床应用价值。

关键词

医学图像处理, 肌肉分割, 图像分割, 目标检测, 分割一切模型

Few-Shot Segmentation of Skeletal Muscle in CT Images Based on SAM2

Awei Song¹, Handa Cai¹, Kejia Jin²

¹College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

²Department of Medical Care Center, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou Zhejiang

Received: Feb. 17th, 2025; accepted: Mar. 26th, 2025; published: Apr. 3rd, 2025

Abstract

Sarcopenia has numerous adverse effects on patients' health, and muscle mass measurement is one of the important methods for diagnosing sarcopenia. However, the current manual measurement

文章引用: 宋阿伟, 蔡函达, 金柯佳. 基于SAM2实现CT影像上骨骼肌的少样本分割[J]. 人工智能与机器人研究, 2025, 14(3): 471-478. DOI: 10.12677/airr.2025.143046

methods are inefficient, limiting their widespread clinical application. This study aims to propose an automatic segmentation method suitable for both clinical diagnosis and medical research of sarcopenia, which can achieve high-precision skeletal muscle segmentation with a small amount of training data. Our method first fine-tunes the YOLOv10 and SAM2 models using a small amount of data. Then, the rectangular box coordinates of each sample are obtained through YOLOv10 and used as prompts for SAM2 to perform skeletal muscle segmentation. Experimental results show that our method achieves a DSC score of 0.9025 in the segmentation task, outperforming the traditional U-Net architecture's score of 0.8763 and the fixed rectangular region prompt's score of 0.8887. Additionally, validation on the test dataset indicates that 85% of the segmentation results fall within the clinically acceptable range. In conclusion, the method proposed in the article demonstrates high accuracy and reliability in the task of few-shot automatic segmentation of skeletal muscle, showing potential clinical application value.

Keywords

Medical Image Processing, Muscle Segmentation, Image Segmentation, Object Detection, Segment Anything Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 前言

肌少症是一种因持续骨骼肌量流失、强度和功能下降而引起的综合征，是老年人生理功能逐渐减退的重要原因和表现之一。近年来，肌少症的定义不断完善，现将其定义为随年龄增长，肌肉数量、质量及其力量与功能逐渐降低，肌肉质量常用肌肉脂肪变性的程度进行临床评估[1]。目前，常用的肌肉量检查方法包括双能 X 射线吸收法(DXA)、生物电阻抗分析(BIA)、超声(US)、计算机断层扫描(CT)和磁共振成像(MRI)等[2]。在 CT 图像中，CT 值(Hounsfield Unit, HU)常用于评估肌肉和肌肉脂肪的密度。CT 检查不同于 MRI，能够对肌肉的表面积、密度以及脂肪浸润程度进行精准定量测量，并且具有较高的可靠性。使用 CT 影像测量肌肉和肌肉脂肪的表面积和密度已得到广泛认可，是诊断肌肉质量的金标准[3]。

目前，大部分临床实践或科学研究高度依赖于研究者的手动分割，手工分割骨骼肌是一项繁琐且耗时的工作，除了受到操作者主观因素的影响外，低下的效率同样限制了其在常规临床实践中的应用[4]。近年来，深度卷积神经网络在图像自动分割任务中表现出色，其中 U-Net 系列模型[5]-[9]在医学图像处理领域被广泛应用。然而，随着分割一切模型(Segment Anything Model, SAM)的出现[10]，研究者们开始关注新的方向。SAM 通过提示词辅助，实现了对图像中任意目标的交互式分割，并且由于其庞大的训练数据，在少样本分割任务中也表现出色，这对于数据获取困难的医学图像分割任务具有重要意义。在不断改进和迭代后，分割一切模型 2 (Segment Anything Model 2, SAM2)在 2024 年被提出[11]，SAM2 是 SAM 的改进版本，增加了对 3D 图像的分割功能。

然而，SAM 系列模型(SAM, SAM2)在医学图像分割上的表现并不理想，因此一些研究者利用现有的公共医学影像数据集对 SAM 进行了微调，开发出了 MedSAM [12]。尽管如此，SAM 系列模型和 MedSAM 模型在骨骼肌分割方面仍存在局限性，且目前尚无完成标注工作的骨骼肌分割数据集被公开。因此，实现骨骼肌的自动分割仍需进行人工标注，并对模型进行训练和微调。标记医学影像数据是一个高成本的工作，且受限于数据的隐私控制问题，模型的训练还需要在少样本条件下进行。此外，SAM 系列模型和

MedSAM 都是交互式分割模型，在实际应用中仍需一定的人工辅助，难以满足大规模数据批量处理的需求，例如在相关医学研究中对大规模人群数据进行分析的场景。

为了在少样本条件下训练出高精度的骨骼肌全自动分割模型，我们充分利用了大模型在这方面的优势，且考虑到研究的可持续性，我们选择了 SAM 系列大模型中的 SAM2 作为基础模型进行改进，因为 SAM2 同时具备在 2D 影像和 3D 影像上进行分割的能力。综上所述，我们在本文中提出了一种基于 SAM2 的少样本骨骼肌全自动分割方法。

2. 材料与方法

2.1. 数据集

我们使用了来自温州地区单一医疗中心在 2016 年至 2020 年间招募的参与者的胸部常规检查 CT 影像。本研究已通过该医院伦理委员会批准，并严格按照《赫尔辛基宣言》进行，以确保研究过程的合规性和研究结果的可靠性。所有入选患者均签署了知情同意书。

从这些患者的影像检查中，我们挑选出具有较高诊断价值的 T12 水平面图像。所有 T12 水平图像均由两位经过专业训练的临床医学学生手动筛选，并对这些图像进行了骨骼肌与肌间脂肪的手动标记，作为训练和测试自动分割模型的必要材料。参与标记工作的研究人员对这些患者的临床信息完全不知情。标记工作完成后，由一位资历深厚的专业副主任医师进行检查，未达到要求的样本会重新进行标注，以确保数据标注的准确性。

2.2. 图像预处理

根据医学成像方面的先验知识，骨骼肌与肌间脂肪的 HU 取值范围介于-190 至 150 之间[13]，因此在整个实验过程中仅观察窗位为 0、宽度为 400 的窗口，即-200~200 HU 的部分。预处理前后图像效果见图 1。

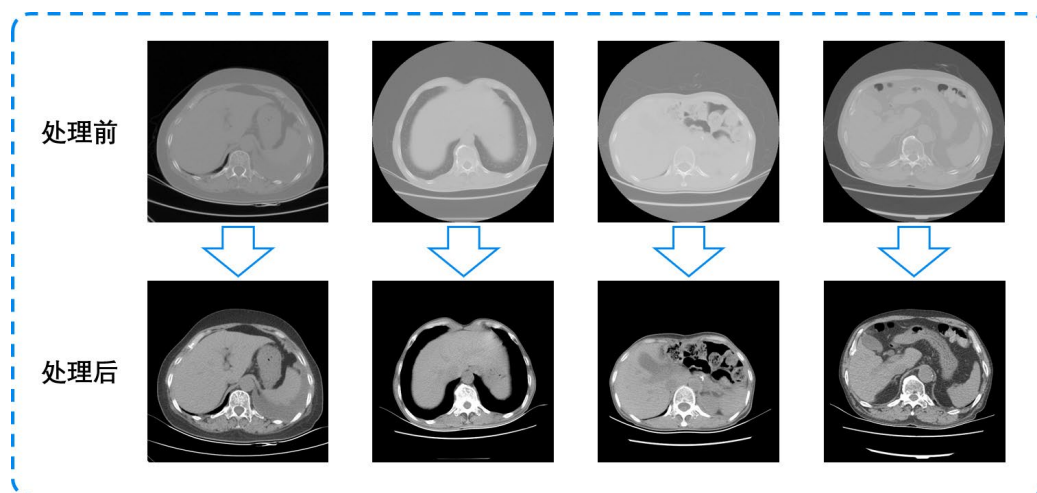


Figure 1. Instances of the comparison before and after image preprocessing
图 1. 图像预处理前后的对比示例

2.3. SAM2

SAM 是一种用于通用图像分割的大模型，能够在各种场景和任务中实现高效且准确的图像分割。通过在大规模、多样化的数据集上进行预训练，SAM 能够泛化到多种不同的图像分割任务中，包括语义分

割、实例分割和全景分割等，并具备出色的零样本分割能力。SAM2 是 SAM 的改进版本，相较于 SAM，SAM2 增加了对视频处理的支持，并引入了记忆模块以有效处理视频。这一特性也使得 SAM2 能够应用于 3D 医学图像的分割任务。

我们利用标记好的 T12 水平 CT 影像对 SAM2 的 Mask 解码器进行微调，训练过程中编码器的参数全部冻结。SAM2 的 Mask 解码器结构如图 2 所示。

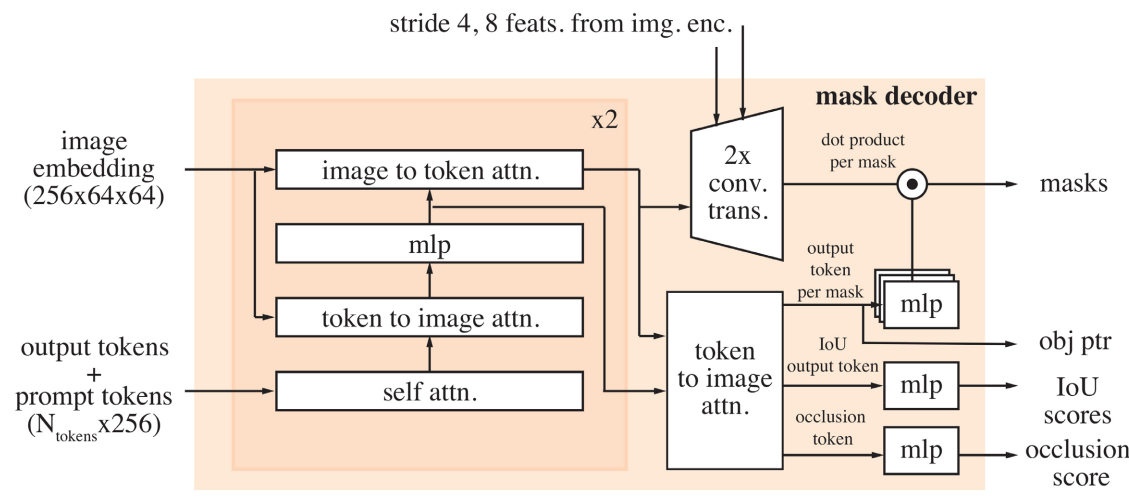


Figure 2. Architecture of the Mask decoder
图 2. Mask 解码器架构图

2.4. YOLOv10

You Only Look Once (YOLO)广泛应用于实时目标检测任务，通过神经网络结构定位和分类图像中的物体，并生成矩形框标识目标物体的位置。YOLO 的运行效率极高，几乎不需要额外的时间成本。YOLO 已经迭代了多个版本，其中 YOLOv10 通过改进模型结构，并利用双标签分配和一致的匹配度量代替冗余的后处理过程，有效提高了检测效率和精度。YOLOv10 在各种目标检测任务中的出色表现，使其广泛应用于监控系统和自动驾驶系统中。我们计划将其应用于医学影像分割任务，为 SAM2 提供分割提示词。

2.5. 提示词生成

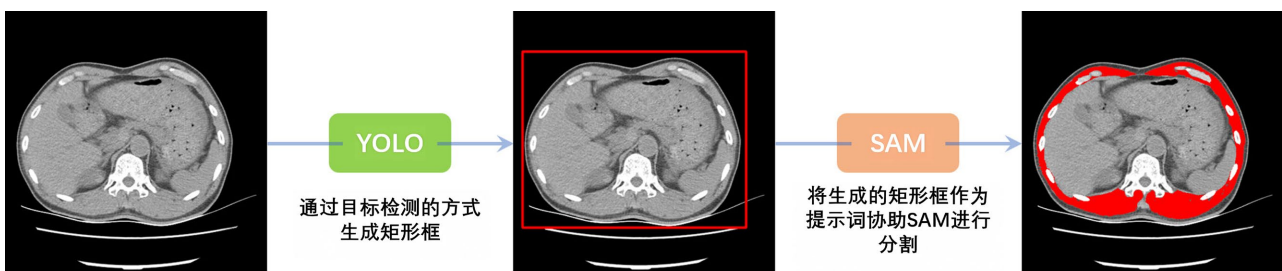


Figure 3. Flowchart of the proposed methodology
图 3. 提出的方法流程图

作为交互式分割模型，SAM 需要用户输入提示词，提示词包括坐标点、矩形区域或粗分割掩膜。由于点提示词对分割精度的提升有限，而粗分割掩膜需要耗费较多的人力，我们选择使用矩形区域作为提示词。提示词的矩形区域由微调后的 YOLOv10 [14]生成，微调过程中仅使用训练集数据。对于 YOLO 未

识别到的样本，根据骨骼肌与肌间脂肪的 HU 值自动生成，将图像中符合阈值范围的所有像素点包裹在内。最后，我们根据这些生成的提示词对骨骼肌进行分割。总之，我们采用的方法如图 3 所示。

3. 模型与评估指标

3.1. 模型选择

SAM2 具备强大的 3D 分割能力，同时支持 2D 分割。鉴于后续工作将扩展至骨骼肌的 3D 分割，本次评估直接采用 SAM2。此外，使用 U-Net 的分割结果作为基线进行测试。用于生成提示词的模块选择了易于获取的 YOLOv10。

3.2. 损失函数

我们采用交叉熵损失(Cross-Entropy)和 Dice 损失组成的复合函数作为最后训练 SAM2 的损失函数。

CE 损失函数是分类问题中常用的一种损失函数，通过衡量预测分布与真实分布之间的差异，指导模型的参数更新。它在二分类和多分类问题中都有广泛应用，并且在神经网络训练、语言模型等领域中发挥重要作用。分割问题可以视为对图像进行前景和背景进行分类，因此 CE 损失函数也常被用于分割模型的训练当中。CE 损失函数表达式为：

$$\text{CELoss} = \frac{1}{H \times W} \sum_i^H \sum_j^W - \left[y_{ij} \times \log p_{ij} + (1 - y_{ij}) \times \log (1 - p_{ij}) \right]$$

其中， H 、 W 分别为图像的高度、宽度， y 为真实值， p 为预测值。

Dice 损失函数是一种常用于图像分割任务的损失函数，特别是在医学图像分割这类需要精确分割的领域。其原理是通过最大化预测分割与真实分割之间的重叠来优化模型，与交叉熵损失函数不同的是，Dice 损失函数更关注分割结果的整体形状和区域重叠，更适合用于不平衡数据和需要精确分割的任务。Dice 损失函数表达式为：

$$\text{DiceLoss} = 1 - \frac{2 \times \sum_{i=1}^W \sum_{j=1}^H p_{ij} \times y_{ij}}{\sum_{i=1}^W \sum_{j=1}^H (p_{ij} + y_{ij})}$$

式中， H 、 W 分别为图像的高度、宽度， y 、 p 分别为真实值、预测值。

最终我们使用的损失函数为 CELoss 与 DiceLoss 之和，表达式为：

$$L = \text{CELoss} + \text{DiceLoss}$$

3.3. 评估方法及指标

评估结果依据肌肉 HU 阈值范围的先验知识进行后处理，预测结果中-29 至 150 HU 的区域被认定为肌肉，其余区域被排除。

本研究在验证集上使用 Dice 相似度系数(Dice Similarity Coefficient, DSC)为主要指标。Dice 系数公式为：

$$\text{DSC} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}$$

其中，TP 表示真正例像素的数量。FP 和 FN 分别表示假正例和假负例像素的数量。

由于最终的分割结果主要用于获取骨骼肌面积，因此我们对平均面积偏差和最大面积偏差也进行了评估。平均面积偏差和最大面积偏差表达式分别为：

$$\bar{d} = \sum_{n=1}^N \frac{|S_n - \hat{S}_n|}{N}$$

$$d_{\max} = \max(S_1 - \hat{S}_1, S_2 - \hat{S}_2, \dots, S_n - \hat{S}_n)$$

其中, N 表示总样本数量, S_n 表示实际面积, $\hat{S}_n$ 表示预测面积。

最后, 我们将测试集的分割结果提供给专家进行评价, 评价分为三个等级:

- 1) 分割结果可接受, 达到专家水平;
- 2) 分割结果在临床上可接受, 与专家分割的面积有 10% 以内的差异;
- 3) 分割结果不可接受, 与专家分割的面积差异超过 10%。

4. 实验

4.1. 实验方法

实验在配备 4 个 RTX4090 GPU、2 个 Intel Xeon Platinum 8336C CPU 和 256 GB RAM 的设备上进行, 使用 Python 3.11.5。数据集在样本有限的条件下平分为训练集和验证集, 各包含 16 个样本, 并提供 180 个样本的测试集。最终取验证集上骨骼肌与肌间脂肪混合区域 DSC 得分最高的一轮训练结果进行测试。

实验中, YOLOv10 选择预训练模型版本为 YOLOv10-S, 微调时仅使用训练集中的 16 个样本。SAM2 微调时的 Batch size 设置为 2, 最多训练 300 轮, 学习率设置为 $6e-5$ 。无提示词对照组采用固定的矩形框作为提示词, 如图 4 所示。

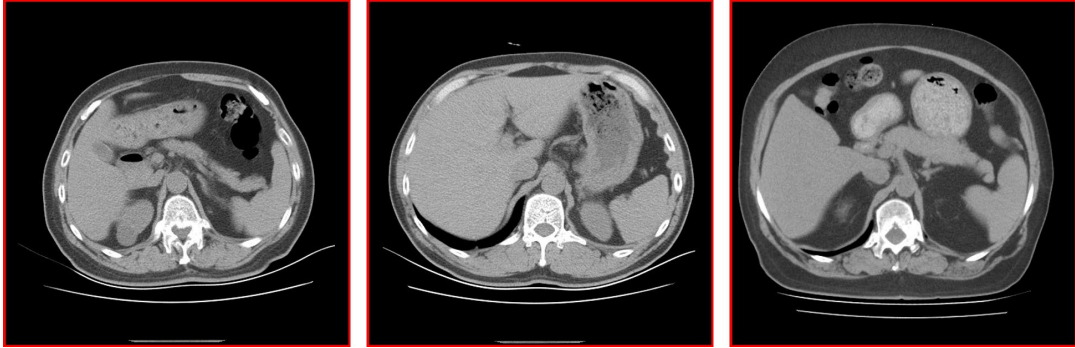


Figure 4. Examples of a sample prompt word with a fixed rectangular box

图 4. 采用固定矩形框提示词的样本示例

4.2. 实验结果

分割结果评分见表 1。传统 U-Net 结构下 DSC 得分为 0.8763, 平均面积偏差为 5.96, 最大面积偏差为 25.09 cm^2 。采用固定矩形框作为提示词的方法时, SAM 分割的 DSC 评分为 0.8887, 平均面积偏差为 5.62 cm^2 , 最大面积偏差达到 21.03 cm^2 。而通过 YOLO 生成的矩形框作为提示词时, DSC 评分提高为 0.9025, 平均面积偏差减少到 5.12 cm^2 , 最大面积偏差减少到 16.43 cm^2 。

分割结果表明, 传统 U-Net 结构在样本较少的情况下, 对于骨骼肌的分割存在一定的误差, 尤其是在最大面积偏差方面, 可能会影响临床应用的精度。相比于传统 U-Net, 固定矩形框提示词的方法在 DSC 评分和面积偏差上都有所改进, 尤其是最大面积偏差减少了约 4 cm^2 , 表明该方法在一定程度上提高了分割的精度和稳定性。另外, 与前两种方法相比, 使用 YOLO 生成的矩形框提示词显著提高了分割的准确性, 具体来说, DSC 评分提高了约 0.014~0.026, 平均面积偏差减少了约 0.5 cm^2 至 0.84 cm^2 , 最大面积偏

差减少了约 4.6 cm² 至 8.66 cm²。这些改进表明，通过 YOLO 生成的矩形框提示词能够更准确地定位骨骼肌区域，从而提高分割结果的精度和一致性。

Table 1. Results of the evaluation of the segmentation
表 1. 分割评估结果

方法	DSC	平均面积偏差(cm ²)	最大面积偏差(cm ²)
U-Net	0.8763	5.96	25.09
SAM (Fixed box)	0.8887	5.62	21.03
SAM (YOLO box)	0.9025	5.12	16.43

经过专家的评估后，测试集评分结果见表 2。

Table 2. Evaluation results of the test set
表 2. 测试集评价结果

可接受程度	数量	比例
可接受(<5%)	93	51.67%
临床可接受(5%~10%)	60	33.33%
不可接受(≥10%)	27	15.00%

评估结果表明，测试集上超过半数的分割结果可以达到专家的分割水平，而有 85%的分割结果在临床诊断中可以被接受，但也有 15%的分割结果存在着较大误差，无法被使用。

综上所述，通过 YOLO 生成的矩形框提示词的方法在骨骼肌分割任务中表现出显著的优势，不仅提高了分割的精度，还减少了面积偏差，特别是最大面积偏差的显著减少，表明该方法在进行骨骼肌的少样本分割训练时具有更高的鲁棒性和可靠性。

5. 结论

本研究提出了一种基于 SAM2 的胸部 CT 影像骨骼肌全自动分割方法，仅需少量人工标注即可实现高精度分割。实验结果显示，在小样本数据条件下，我们的方法分割精度超过了传统 U-Net 架构的 0.8763。固定矩形框模式下，SAM2 分割结果的 DSC 得分为 0.8887，而使用 YOLOv10 提供的矩形框后，DSC 得分提升至 0.9025。相比固定矩形框，通过 YOLOv10 生成的矩形框作为提示词能够提高 SAM2 的分割精度。此外，该方法在批量处理大规模数据时减少了人力开销，适用于快速分析庞大人群影像数据的场景。

尽管目前模型在 T12 平面上实现了高精度骨骼肌分割，但由于不同平面之间的差异，完成胸部 CT 其他平面的骨骼肌分割仍面临巨大挑战。我们接下来的计划是通过选取尽可能少的平面实现高精度的 3D 分割。

参考文献

[1] Kim, S.H., Hong, C.H., Shin, M., Kim, K.U., Park, T.S., Park, J.Y., *et al.* (2024) Prevalence and Clinical Characteristics of Sarcopenia in Older Adult Patients with Stable Chronic Obstructive Pulmonary Disease: A Cross-Sectional and Follow-Up Study. *BMC Pulmonary Medicine*, **24**, Article No. 219. <https://doi.org/10.1186/s12890-024-03034-5>

[2] Li, L., Xia, Z., Zeng, X., Tang, A., Wang, L. and Su, Y. (2024) The Agreement of Different Techniques for Muscle Measurement in Diagnosing Sarcopenia: A Systematic Review and Meta-Analysis. *Quantitative Imaging in Medicine and Surgery*, **14**, 2177-2192. <https://doi.org/10.21037/qims-23-1089>

[3] Zhang, S., Chen, H., Xu, H., Yi, Y., Fang, X. and Wang, S. (2022) Computed Tomography-Based Paravertebral Muscle

- Density Predicts Subsequent Vertebral Fracture Risks Independently of Bone Mineral Density in Postmenopausal Women Following Percutaneous Vertebral Augmentation. *Aging Clinical and Experimental Research*, **34**, 2797-2805. <https://doi.org/10.1007/s40520-022-02218-5>
- [4] Cao, J., Zuo, D., Han, T., Liu, H., Liu, W., Zhang, J., *et al.* (2022) Correlation between Bioelectrical Impedance Analysis and Chest CT-Measured Erector Spinae Muscle Area: A Cross-Sectional Study. *Frontiers in Endocrinology*, **13**, Article 923200. <https://doi.org/10.3389/fendo.2022.923200>
 - [5] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
 - [6] Liu, X., Gao, P., Yu, T., Wang, F. and Yuan, R. (2025) CSWin-UNet: Transformer UNet with Cross-Shaped Windows for Medical Image Segmentation. *Information Fusion*, **113**, Article ID: 102634. <https://doi.org/10.1016/j.inffus.2024.102634>
 - [7] Saikia, F.N., Iwahori, Y., Suzuki, T., Bhuyan, M.K., Wang, A. and Kijisirikul, B. (2023) MLP-UNet: Glomerulus Segmentation. *IEEE Access*, **11**, 53034-53047. <https://doi.org/10.1109/access.2023.3280831>
 - [8] 徐旺旺, 许良凤, 李博凯, 等. TransAS-UNet: 融合 Swin Transformer 和 UNet 的乳腺癌区域分割[J]. 中国图象图形学报, 2024, 29(3): 741-754.
 - [9] Xia, W., Fortin, M., Ahn, J., Rivaz, H., Battié, M.C., Peters, T.M., *et al.* (2019) Automatic Paraspinal Muscle Segmentation in Patients with Lumbar Pathology Using Deep Convolutional Neural Network. In: Shen, D., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*, Springer, 318-325. https://doi.org/10.1007/978-3-030-32245-8_36
 - [10] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
 - [11] Ravi, N., Gabeur, V., Hu, Y.T., Hu, R.H., Ryali, C.K., Ma, T.Y., *et al.* (2024) SAM 2: Segment Anything in Images and Videos. arXiv: 2408.00714.
 - [12] Cheng, J.L., Ye, J., Deng, Z.Y., Chen, J.P., Li, T.X., Wang, H., *et al.* (2023) SAM-Med2D. arXiv: 2308.16184.
 - [13] Ahmad, N., Strand, R., Sparresäter, B., Tarai, S., Lundström, E., Bergström, G., *et al.* (2023) Automatic Segmentation of Large-Scale CT Image Datasets for Detailed Body Composition Analysis. *BMC Bioinformatics*, **24**, Article No. 346. <https://doi.org/10.1186/s12859-023-05462-2>
 - [14] Wang, A., Chen, H., Liu, L.H., Chen, K., Lin, Z.J., Han, J.G. and Ding, G.G. (2024) YoLOv10: Real-Time End-to-End Object Detection. arXiv: 2405.14458.