scientific reports

Check for updates

OPEN

A Two stage deep learning network for automated femoral segmentation in bilateral lower limb CT scans

Wenqing Xie[1,2,4], Peng Chen[1,2,4], Zhigang Li[1,2], Xiaopeng Wang[1,2], Chenggong Wang[1,2], Lin Zhang[3], Wenhao Wu[3], Junjie Xiang[3], Yiping Wang[3]✉ & Da Zhong[1,2]✉

This study presents the development of a deep learning-based two-stage network designed for the efficient and precise segmentation of the femur in full lower limb CT images. The proposed network incorporates a dual-phase approach: rapid delineation of regions of interest followed by semantic segmentation of the femur. The experimental dataset comprises 100 samples obtained from a hospital, partitioned into 85 for training, 8 for validation, and 7 for testing. In the first stage, the model achieves an average Intersection over Union of 0.9671 and a mean Average Precision of 0.9656, effectively delineating the femoral region with high accuracy. During the second stage, the network attains an average Dice coefficient of 0.953, sensitivity of 0.965, specificity of 0.998, and pixel accuracy of 0.996, ensuring precise segmentation of the femur. When compared to the single-stage SegResNet architecture, the proposed two-stage model demonstrates faster convergence during training, reduced inference times, higher segmentation accuracy, and overall superior performance. Comparative evaluations against the TransUnet model further highlight the network's notable advantages in accuracy and robustness. In summary, the proposed two-stage network offers an efficient, accurate, and autonomous solution for femur segmentation in large-scale and complex medical imaging datasets. Requiring relatively modest training and computational resources, the model exhibits significant potential for scalability and clinical applicability, making it a valuable tool for advancing femoral image segmentation and supporting diagnostic workflows.

Keywords Femur segmentation, Deep learning, Two-stage, YOLOv8, SegResNet

Abbreviations

ROI	Regions of interest
IoU	Intersection over Union
CNNs	Convolutional Neural Networks
ap	Average precision
mAP	Mean average precision
P–C Curve	Precision-recall curve
PA	Pixel accuracy
BMD	Bone mineral density
CAD	Computer-aided diagnosis
MONAI	Medical open network for artificial intelligence
ITK	Insight tool kit
GPU	Graphics processing unit
CPU	Central processing unit
CT	Computed tomography
MRI	Magnetic resonance imaging

[1]Deparment of Orthopedics, Xiangya Hospital, Central South University, Changsha 410008, Hunan, China. [2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changshan 410008, Huna, China. [3]Changzhou Jinse Medical Information Technology Co., Ltd, Changzhou 213000, Jiangsu, China. [4]Wenqing Xie and Peng Chen have contributed equally to this work. ✉email: yiping.wang@midivi.cn; zhongda@csu.edu.cn

The incidence of hip fractures has risen exponentially in recent years, with approximately 90% of cases attributed to age-related declines in Bone Mineral Density (BMD) at the proximal femur and an increased likelihood of falls[1]. Elderly hip fractures typically involve femoral neck, intertrochanteric, or subtrochanteric fractures, all of which are associated with high mortality and disability rates. These rates are further exacerbated by conservative treatments[2]. These outcomes are further exacerbated by the limitations of conservative treatments. Timely diagnosis and surgical intervention remain critical for managing such fractures effectively, with imaging serving as the primary diagnostic tool[3].

Amid rapid advancements in computational technology, the importance of computerized image processing methodologies in medical imaging has become increasingly prominent[4]. Using computers for medical image analysis facilitates the identification of clinical diseases and the formulation of treatment plans. Refining image quality through processes such as denoising[5], enhancement[6], and segmentation[7] accentuates key features and information within the images. Moreover, Computer-Aided Diagnosis (CAD) has become an indispensable tool in the medical field, assisting physicians in analyzing medical imagery, extracting relevant features, and segmenting lesions for classification or quantification. It provides invaluable support in decision-making, enhances diagnostic accuracy, and reduces both omission and false-positive errors[8].

The intricate anatomy of bone tissue, which often appears in a grayscale similar to that of adjacent tissues, muscles, and joints, makes its differentiation in medical images a challenging task. Achieving swift and accurate segmentation of bone tissue presents a significant challenge, and precision in bone tissue segmentation is crucial for computer-aided diagnosis and preoperative planning in various fracture scenarios[9], which makes it holds substantial practical importance within the clinical milieu.

Traditional segmentation methods

It often focuses on extracting low-level features from scenes. While traditional segmentation methods offer advantages in computational resources, efficiency, and simplicity, their performance often degrades when handling complex medical images, particularly those characterized by high noise, weak intensity variations, or blurred boundaries. Furthermore, these methods rely heavily on the expertise of physicians and radiologists, making them increasingly inadequate for the practical demands of medical image segmentation[10,11]. Traditional image segmentation methods[12,13] (Table 1) can be broadly categorized into edge-detection-based methods[14,15], threshold-based methods[16], region-based methods[17,18], graph-based segmentation methods[19,20], clustering-based methods[21,22], energy functional-based methods for medical image segmentation[23–28], atlas-based methods for medical image segmentation[29–32] and specific theories-basedmethods[33,34].

Deep learning-based algorithms

Traditional segmentation techniques struggle to meet the demands of high precision and automation. The diversity of medical imaging, along with the complexity of pathological features, unclear lesion boundaries, and noise interference, all contribute to challenges in achieving accurate and generalizable image segmentation. With the advancement and rise of artificial intelligence, deep learning algorithms have demonstrated impressive capabilities in medical image segmentation[35]. Deep learning, particularly Convolutional Neural Networks (CNNs), utilizes multi-layered neural network architectures to automatically learn and extract image features ranging from simple to complex. This approach effectively captures pathological changes and anatomical structures in medical images, thereby improving segmentation accuracy and robustness. Currently, deep learning-based medical image segmentation methods can be broadly categorized into five types: CNN-based

Segmentation technique	Explanation	Merits	Demerits
Edge-detection-based	Detecting the boundaries of the target region and connecting the edges to form a region	Simple and efficient	Lack of edge continuity and closure causes fragmented edges in detailed areas and difficulty balancing noise resistance and precision
Threshold-based	Set different thresholds to divide the pixels into background and target regions with distinct grayscale levels	It is easy to understand and implement, with low computational complexity and high efficiency	It lacks pixel spatial features, is sensitive to noise, and has poor robustness
Region-based	Divide the image into regions based on the similarity of pixels within the same category	It utilizes the local spatial information of the image, exhibiting favorable regional features	It has high computational costs and is prone to holes and over-segmentation
Graph-based	The segmentation of the graph is transformed into an optimization problem of an undirected weighted graph	It effectively preserves the image boundaries	It is slow, inefficient, computationally complex, and may exhibit "small cut" behavior and leakage
Clustering-based	Iterative clustering	The algorithm is conceptually simple	The segmentation results are highly sensitive to parameters
Energy-functional-based	Gradient information drives contour evolution via partial differential equations, with continuous curves representing the target boundary	Effectively separates the target region from the background in the image	The Snake model is noise-sensitive, with performance dependent on the initial contour. It struggles with curve topology changes, causing boundary leakage, and it has high computational complexity
Atlas-based	Construct a graph to apply prior knowledge for segmenting the target image	Make full use of the advantages of manual segmentation. Transform the image segmentation problem into an image registration problem	It requires significant memory and time. When there is a large morphological difference between the target image and the graph, registration alone is insufficient for accurate segmentation

Table 1. Explanation and merits/demerits of traditional image segmentation.

methods, Transformer-based methods, Mamba-based methods, semi-supervised learning methods, and weakly supervised learning methods[36].

CNN-based methods
Prior to 2021, most networks employed the U-Net[37] architecture as the foundation for improving medical image segmentation, with variants such as V-Net[38], Unet3 + [39], Attention U-Net[40], nnU-Net[41], and Dense-U-Net[42]. These architectures primarily leverage Convolutional Neural Networks (CNNs) to extract inductive bias information from images. The key feature of U-Net is its distinctive encoder-decoder framework with skip connections, where the encoder effectively extracts hierarchical features and the decoder uses skip connections to reconstruct the segmented image. However, its focus on local features limits further improvements in segmentation accuracy.

Transformer-based methods
Unlike CNNs, Transformer architecture is a deep learning method based on self-attention mechanisms. It leverages multi-head self-attention layers in the encoder to capture long-range dependencies in the encoding process[43]. Examples include Convolution-Free Networks[44], such as Swin-Unet[45]. However, the performance of Transformer architecture is often limited by the insufficient availability of medical image data. This has led to the development of hybrid architectures that combine Transformer and CNNs, such as TransUNet[46], DS-TransUNet[47], and UNETR++[48].

Mamba-based methods
Recently, Mamba has emerged as a novel variant of the selective State Space Model (SSM), addressing the limitations of Convolutional Neural Networks (CNNs) in global modeling while maintaining linear complexity. It has demonstrated significant potential in medical image segmentation[49]. Examples include VM-UNet[50], LightM-UNet[51], and Swin-UMamba[52].

Semi-supervised and weakly supervised learning methods
To address the challenges posed by the limited availability of medical datasets and the need for large amounts of labeled data, semi-supervised and weakly supervised learning methods have been developed for medical image segmentation[53]. These methods train models using a small number of annotated samples. Semi-supervised learning methods include UA-MT[54], SASSNet[55], and ACTION++[56]. Weakly supervised learning methods include CCNN[57], DFTNet[58], and Cluster-Re-Supervision[59]. However, incomplete or inaccurate labeling can significantly impact model performance by leading to incorrect feature learning.

Femoral segmentation in medical imaging

Recent studies have attempted to address the challenges of femur segmentation in medical images. The primary challenge in bone tissue segmentation is the lack of contrast between bones across joints. The femur exhibits low contrast with adjacent joints and soft tissues, and its complex anatomy further complicates segmentation. As a result, traditional image processing techniques, such as threshold-based, edge-based, and region-based segmentation algorithms, face certain limitations[60]. Most femur segmentation methods rely on techniques such as atlas-based and/or Statistical Shape Models (SSM) and deformable models, which improve segmentation performance by iteratively adjusting numerous hyperparameters and shape priors[61–67]. Additionally, atlas-based methods have been used to achieve semi-automatic and unsupervised femur segmentation, although they still require some human intervention[68]. Machine learning approaches, such as Random Forest Classifiers[69], Decision Trees[70], and Bayes decision rule[71] decision rules, have also been applied for femur segmentation. Although these classifiers do not require prior anatomical knowledge, they typically rely on manually extracted features.

Deep learning, with its powerful feature extraction capabilities and ability to learn complex patterns, can adaptively learn and extract multi-level features from large datasets. This results in high accuracy, robustness, and scalability, making it well-suited to meet the demands of high dimensionality, complexity, and precision required for medical image segmentation. Numerous studies have attempted to address these challenges using deep learning approaches[72–81]. Overall, as deep learning-based segmentation models continue to evolve through the addition of modules and structural improvements, segmentation performance is progressively enhanced. However, this also places increasing demands on the training process and computational resources[82,83]. Therefore, we focus on improving segmentation performance with limited computational resources by altering the segmentation pipeline, particularly when handling large datasets (such as bilateral lower limb CT scans). This approach aims to enhance segmentation rather than solely relying on modifications and/or additional modules within the segmentation network.

Our contributions

This paper introduces a novel two-stage deep learning network for fully automated femoral segmentation in bilateral lower limb CT scans. The proposed pipeline integrates object detection and semantic segmentation to achieve end-to-end automation. By adopting a two-stage framework, the network achieves high segmentation accuracy with reduced computational overhead compared to single-stage architectures such as SegResNet. The dual advantage of enhanced precision and efficiency underscores its capability to handle complex segmentation tasks, making it a promising solution for clinical applications.

Methods

3D data processing in the first stage

In the first stage, we use Digitally Reconstructed Radiograph (DRR)[84] technology to process bilateral lower limb full-length CT scans, creating coronal (frontal) and axial (transverse) views, which capture the structure and intensity distributions within the volume. The proposed workflow consists of four components. The processing in the first stage is shown in Fig. 1.

Ray marching for 2D projection generation

The 3D CT data is processed to generate 2D coronal and axial projection images using ray marching. These projections are essential for capturing structural information from different orientations, enabling effective visualization and subsequent analysis.

Object detection using YOLOv8

YOLOv8, an object detection framework, is applied to the generated 2D images. Separate models or pipelines process the coronal and axial projections to detect and localize regions of interest, producing bounding boxes for each projection.

3D bounding box construction

The 2D bounding boxes obtained from the coronal and axial views are geometrically fused to derive a 3D bounding box. This process involves calculating the intersection and extrapolation of the detected regions in the two views to ensure precise spatial localization within the 3D volume.

Cropping the original CT data

The computed 3D bounding box is applied to the original CT data to extract a subvolume containing the region of interest. This cropping step reduces the data size and isolates relevant structures for subsequent analysis, such as segmentation, classification, or quantitative assessment.

Segmentation of Femur in the second stage

In the first stage, YOLOv8's primary objective is to accurately detect approximate bounding boxes around the Regions of Interest (ROIs) in femur medical images. In the second stage, SegResNet utilizes the highlighted femur region generated by YOLOv8 to perform the segmentation task and output the predicted femur mask. The abstract structure of the Second stage is shown in Fig. 2. By leveraging YOLOv8 to eliminate unnecessary regions, the SegResNet segmentation task experiences reduced memory usage and are less influenced by negative samples from irrelevant regions.

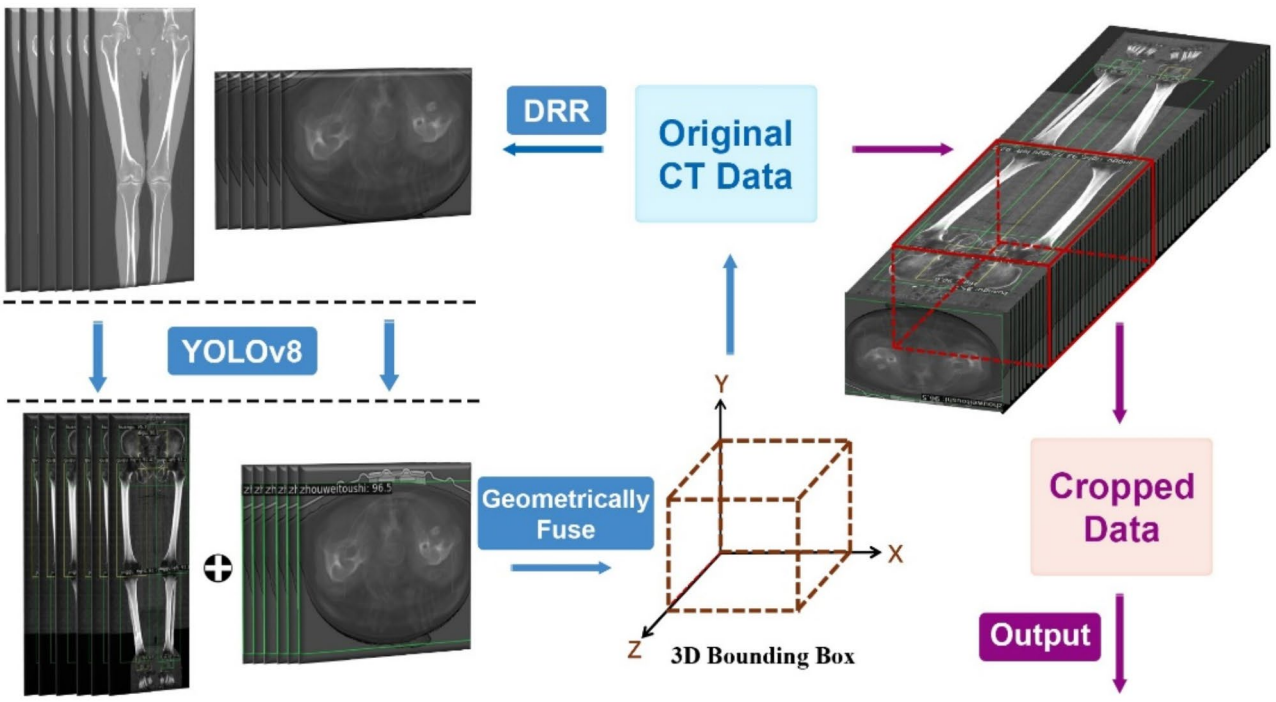

Fig. 1. The original CT data is divided into two 2D coronal and axial projection images using DRR techniques. These images are processed by YOLOv8 to generate two bounding boxes, which are then fused to create a 3D bounding box. The 3D bounding box is used to crop the original CT data, extracting the region of interest (ROI) subset as the output.

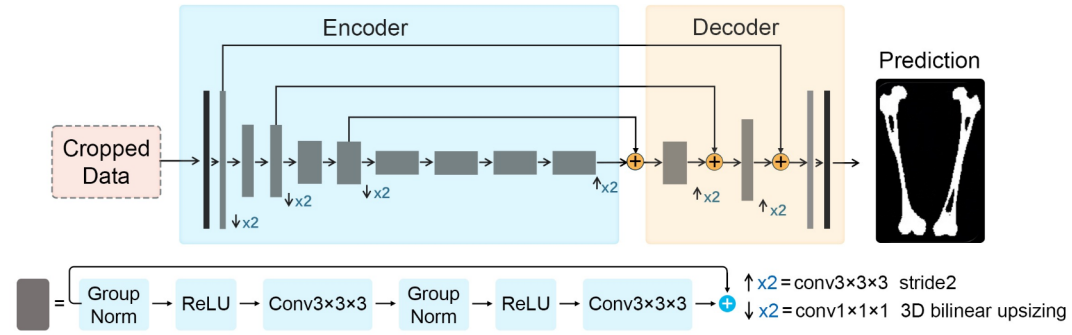

Fig. 2. The second stage performs semantic segmentation based on SegResNet. The cropped data from the first stage (YOLOv8) is used as input, and the decoder of the SegResNet architecture gradually extracts image features. These features are passed through the encoder to restore the image's spatial resolution and generate the femur mask as the segmentation output.

We employ YOLOv8[85] to detect the femoral region. The advantages of YOLOv8 as a target detection framework are evident. YOLOv8 is a representative model for one-stage object detection, utilizing a single neural network to make predictions on the entire image. A single forward pass generates detection results for all objects, making it faster compared to traditional algorithms like Fast R-CNN[86]. The simplicity of YOLOv8's overall network structure greatly facilitates the training process and parameter tuning. YOLOv8 introduces the C2f. module[87], which enhances the network's feature fusion ability, increasing inference speed and achieving further lightweight performance. SegResNet[88] combines the residual structure of ResNet[89] with the encoding–decoding structure of U-Net to enhance the accuracy and effectiveness of semantic segmentation tasks. We employed it as the network for the second stage of femur segmentation.

Experiments and results

Dataset

The experimental dataset used in this paper is from the Hospital. There are 100 sets of experimental data sets used in this research in total: 85 sets for training sets, 8 sets for validation, and 7 sets for testing. The study was approved by the Ethics Committee of Xiangya Hospital, Central South University (No: 202310926). A requirement for informed consent was waived for this study by the approval of the Ethics Committee of Xiangya Hospital, Central South University. Clinical data was de-identified and analyzed anonymously. The study was compliant with the ethical guidelines of the Declarations of Helsinki.

The resolution of the data varies, with values including 0.76 mm × 0.76 mm × 0.8 mm, 0.85 mm × 0.85 mm × 1 mm, and 0.8262 mm × 0.8262 mm × 0.75 mm. To standardize the input data to a unified size, a resampling layer is applied in SegResNet as part of the data processing pipeline. The size of each data layer is 512 × 512, and the number of slices ranges from 1123 to 1769. The original dataset format is DICOM, which is then converted to nii.gz to ensure patient privacy. The resolution and image size are preserved during this conversion, remaining consistent with the original DICOM format.ITK-SNAP is the software used to display the original image, examine the labeled femur regions, and present the final 3D reconstruction results. The system operates on a 64-bit Windows 11 platform with a built-in CPU frequency of 3.20 GHz. The training and prediction processes primarily rely on the graphics card, an NVIDIA GTX A4000 with 16 GB of memory. The programming environment utilizes PyCharm as the development IDE, with Python 3.9 as the programming language. The deep learning segmentation algorithm is based on the Medical Open Network for Artificial Intelligence (MONAI), while the deep learning object detection algorithm is implemented using MMYOLOv8. The Insight Toolkit (ITK) is employed for data processing.

Model training and prediction

Dimensionality reduction is performed on the original CT data using Digitally Reconstructed Radiograph (DRR) technology. The processed data are then used to train the YOLOv8 model, which is responsible for cropping the CT images and highlighting the femur regions.

The outputs from the first stage serve as input for training the SegResNet model, which generates precise femur masks for both sides. During the training process, a validation step is performed every five iterations to evaluate the predictions on the validation set. The validation loss value is used as a metric to guide the optimization process. The training concludes once the specified number of iterations is reached, and the model with the best performance on the validation set is saved locally. The model parameters for YOLOv8 and SegResNet are detailed in Tables 2 and 3, respectively. The model prediction process follows the same two-stage framework. First, the YOLOv8 model predicts the femur regions, and second, the SegResNet model generates the femur masks. All data processing steps are seamlessly integrated within the network pipeline. During prediction, the algorithm takes the original CT scan images as input and outputs the predicted femur masks, which maintain the same resolution and size as the input images.

Training Parameters	Parameter Value
Max epochs	120
Training Batch Size per GPU	32
Number of Workers for Training	4
Minimum Area Ratio	0.01

Table 2. YOLOv8 training parameters table.

Training Parameters	Parameter Value
Max epochs	300
Patch size	(160, 160, 160)
Batch size	4
Resampled spacing	(1.0, 1.0, 1.0)
Normalize intensity range	(-100, 500)
Blocks down	(1, 2, 2, 4)
Blocks up	(1, 1, 1)
Initialization Filter	16
In channels	1
Out channels	2
Dropout rate	0.2
Loss function	Dice loss

Table 3. Two-stage SegResNet training parameters table.

Subject no	IoU	AP (IoU ≥ 0.3)	Averaged IoU	mAP
1	0.97	0.9510	0.9671	0.9656
2	0.96	0.9600		
3	0.96	0.9460		
4	0.99	0.9710		
5	0.96	0.9620		
6	0.96	0.9810		
7	0.97	0.9880		

Table 4. Evaluation table of YOLOv8 on test sets.

Evaluation metrics

Given the two-stage architecture of our model, the evaluation metrics have been correspondingly divided. In the first stage, the precision of object localization is assessed using the Intersection over Union (IoU) metric, while the accuracy of object category prediction is evaluated through a suite of measures: Precision, Recall, the Precision-Recall (P-R) curve, Average Precision (AP), and mean Average Precision (mAP). The performance of our segmentation model was thoroughly evaluated using the Dice coefficient, along with metrics for sensitivity, specificity, and pixel accuracy.

Experimental results

YOLOv8

The results of the experiment are shown in Table 4. The stabilized average Intersection over Union (IoU) of 96.71% indicates that the femur region is correctly framed, and the mean Average Precision (mAP) reaches 0.9656. These results demonstrate that YOLOv8 can effectively recognize the femoral region with high performance, providing accurate data for the subsequent SegResNet step. This ensures that the final femur segmentation remains unaffected. During training, we found that setting the threshold to 0.3 effectively distinguished the bone tissue from the surrounding soft tissue, and thus, the threshold was set to 0.3.

Two-stage SegResNet

The specific evaluation metrics of Two-stage SegResNet based on test sets are shown in Table 5. The table shows that Two-stage SegResNet has large differences in sensitivity and Dice coefficient metrics and that Dice coefficient and sensitivity also differ between subjects, mainly due to differences in segmentation of the femoral canal. However, such differences do not affect the primary objective of segmenting the femoral region from its

Subject no	Dice	Sensitivity	Specificity	Pixel Accuracy
1	0.9327	0.9758	0.9957	0.9950
2	0.9092	0.9800	0.9937	0.9932
3	0.9722	0.9619	0.9992	0.9976
4	0.9624	0.9462	0.9992	0.9973
5	0.9667	0.9691	0.9984	0.9972
6	0.9634	0.9603	0.9989	0.9975
7	0.9644	0.9647	0.9985	0.9971
Averaged	0.953	0.965	0.998	0.996

Table 5. Evaluation table of Two-stage SegResNet on test sets.

Model	Training time (s)	Averaged inference time in test data (s)	Best dice at convergence
Two-stage SegResNet	9755	61	0.9621
Non-Two-Stage SegResNet	21,251	87	0.7899

Table 6. Comparison of two-stage and non-two-stage SegResNet.

adjacent tissues. The specificity and pixel accuracy evaluation metrics are stable with less variation and higher values, indicating that the Two-stage SegResNet segmentation network is accurate for femoral segmentation.

Comparison of segmentation models of the femoral region

The same dataset is used to train and test both the two-stage semantic segmentation and the SegResNet models. A comparison is made between the two-stage model presented in this paper and the original SegResNet model. Since the input data for the non-two-stage SegResNet is larger compared to the two-stage SegResNet, a higher maximum number of epochs is set to ensure convergence.

The comparison of training time and inference time between the two models is presented in Table 6. The two-stage model not only drastically reduces training time and average inference time on test data but also significantly improves the best Dice coefficient. Figures 3 and 4 illustrate the visual performance of segmentation for both models.

The results of the comparison of accuracy and loss values between the two models are shown in Fig. 5. The horizontal axis represents the number of network epochs, while the vertical axis displays the loss per epoch on the validation set. The visuals demonstrate that as the number of epochs increases, the overall loss for the two-stage model is lower than that of the Non-Two-Stage model, with the two-stage model showing a faster reduction in loss. Moreover, the overall accuracy of the two-stage model is higher and increases and converges at a faster rate compared to the non-two-stage model.

Compare with state-of-the-art works

We compared the proposed SegResNet model with the TransUNet model using the same GPU and test set (CT data cropped by YOLOv8). The results are presented in Table 7 below. After both models underwent the first-stage cropping, SegResNet outperformed TransUNet in both averaged inference time on test data and the Dice coefficient. This is further analyzed in section "Analysis of the results compared with TransUNet".

Discussion

Automating femur segmentation in medical imaging presents a significant challenge, yet it holds substantial practical importance. The diversity of imaging modalities, coupled with inherent challenges such as irregular grayscale values, low contrast, and high noise levels, makes bone tissue segmentation a complex task. This complexity is further compounded by the considerable inter-individual anatomical variability.

This paper presents a two-stage, deep learning-based semantic segmentation network, which has been meticulously trained, with finely tuned parameters and the optimal model carefully preserved. Empirical experiments were conducted using this advanced network, and the results were rigorously analyzed. The findings reveal that the proposed segmentation method achieves a mean Dice coefficient of 0.953 and a pixel accuracy of 0.996 on the test set. Notably, the segmentation network operates without the need for manual intervention in image processing, resulting in a workflow characterized by both high efficiency and accuracy.

The use of object detection to quickly and accurately delineate the femoral region in complex medical imagery while simultaneously excluding irrelevant areas is crucial for reducing memory demands and minimizing the impact of excessive negative samples. YOLOv8 is employed as the object detection model in the first stage, providing fast detection speed and high accuracy, making it well-suited for the initial femur recognition task. The SegResNet network, used for the semantic segmentation stage, combines a residual architecture with an encoder-decoder structure, effectively enhancing segmentation accuracy and performance. The residual connections facilitate better training of deep networks by addressing issues like vanishing and exploding gradients, while the encoder-decoder structure integrates low-level details with high-level semantic information[90]. Encoder-

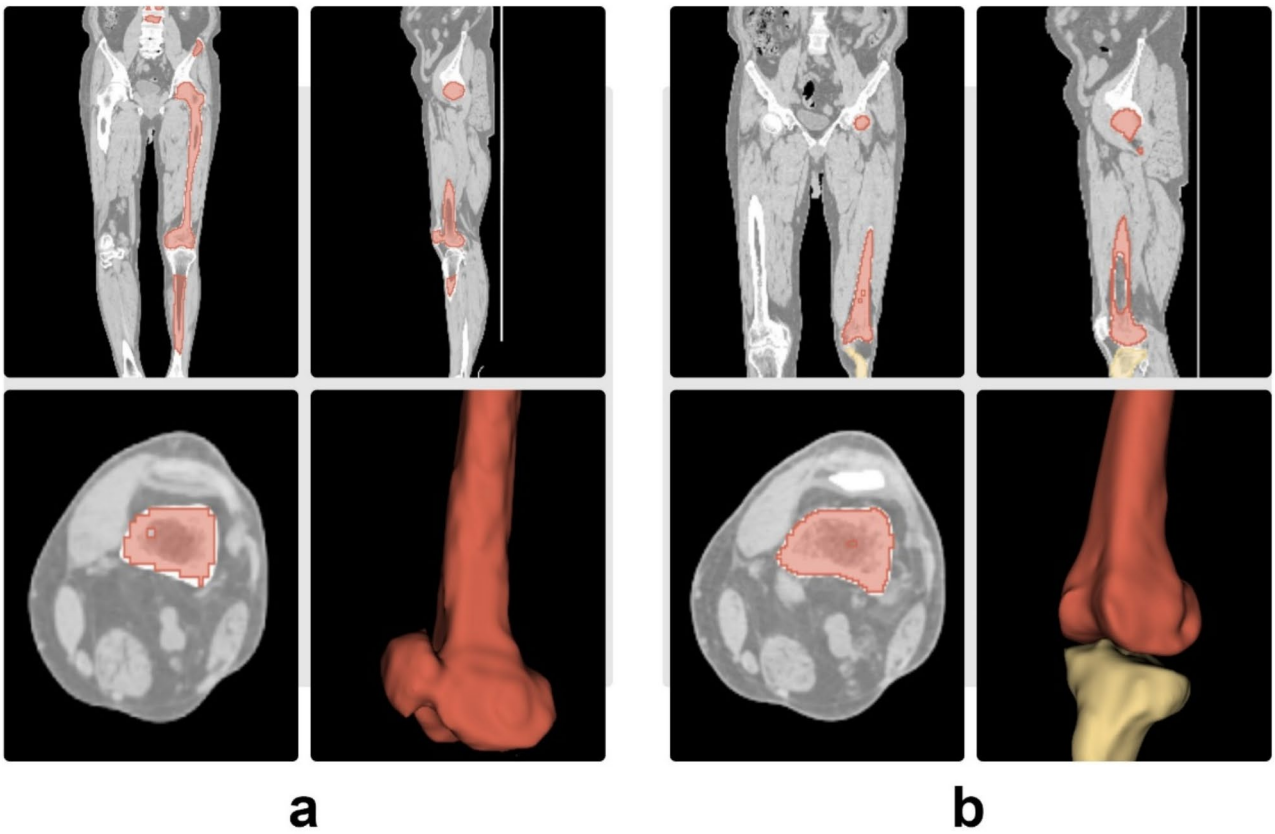

Fig. 3. Femur masks from Two-stage SegResNet (**b**) and Non-Two-Stage model (**a**). The Non-Two-Stage model, without first-stage cropping, incorrectly and incompletely segments the femur mask (as the tibia), while the Two-Stage model segments the femur mask accurately and precisely.

Decoder structure preserves semantic accuracy and improves the recovery of object boundaries and detailed features, ultimately enhancing segmentation results and making it ideal for medical image segmentation[91].

Under the same GPU and dataset conditions, the two-stage segmentation network achieved a Dice coefficient of 0.9621 at convergence, significantly higher than SegResNet's 0.7899. Additionally, its inference time and training speed outperformed those of SegResNet. Comparing the validation loss and accuracy between the two models, the proposed network reached stable accuracy and lower loss more quickly, demonstrating its superior feature learning capability, higher efficiency, and ease of training. This discrepancy can be attributed to the fact that two-stage training effectively reduces redundant pixel areas in the predicted images, leading to a smaller denominator in the Dice coefficient calculation. While the numerator, representing the number of labeled pixels, remains relatively constant, this difference creates a notable gap in the Dice coefficient and accuracy calculations.

Moreover, the entire segmentation process does not require image cropping and is capable of automatically segmenting batch femur regions, thereby reducing the need for manual intervention. The dataset used in this study consists of comprehensive lower limb CT scans, with each slice being a 512×512-pixel matrix and the number of slices ranging from 1123 to 1769. The dataset represents a large volume of input data, which necessitates the use of high-performance computing resources. However, the model presented in this paper, even under less demanding computational conditions, significantly reduces both training and inference times while maintaining impressive segmentation precision. This evinces the Two-stage network process's robust adaptability and a pronounced level of universality.

Comparison with related work

X-rays lack spatial and structural details of the skeleton, limiting the information available for femur segmentation[92]. Zhang et al.[81]. employed a two-stage convolutional network for femoral region segmentation from X-ray images, achieving end-to-end batch processing with relatively stable segmentation metrics. However, their network structure is complex, involving numerous parameters. Additionally, the limited size of the input block requires sliding window sampling for larger input data during image preprocessing, thereby increasing the computational load. Moreover, their 3D reconstruction based on X-rays shows deficiencies in spatial structure and detail. In response, the method proposed in this paper utilizes whole lower limb CT data without requiring any manual preprocessing steps or computationally demanding network architectures. This makes our approach more widely applicable. However, it is important to note that directly comparing the Dice coefficient between different studies is not scientifically valid, as the training and validation datasets may differ. CT imaging offers

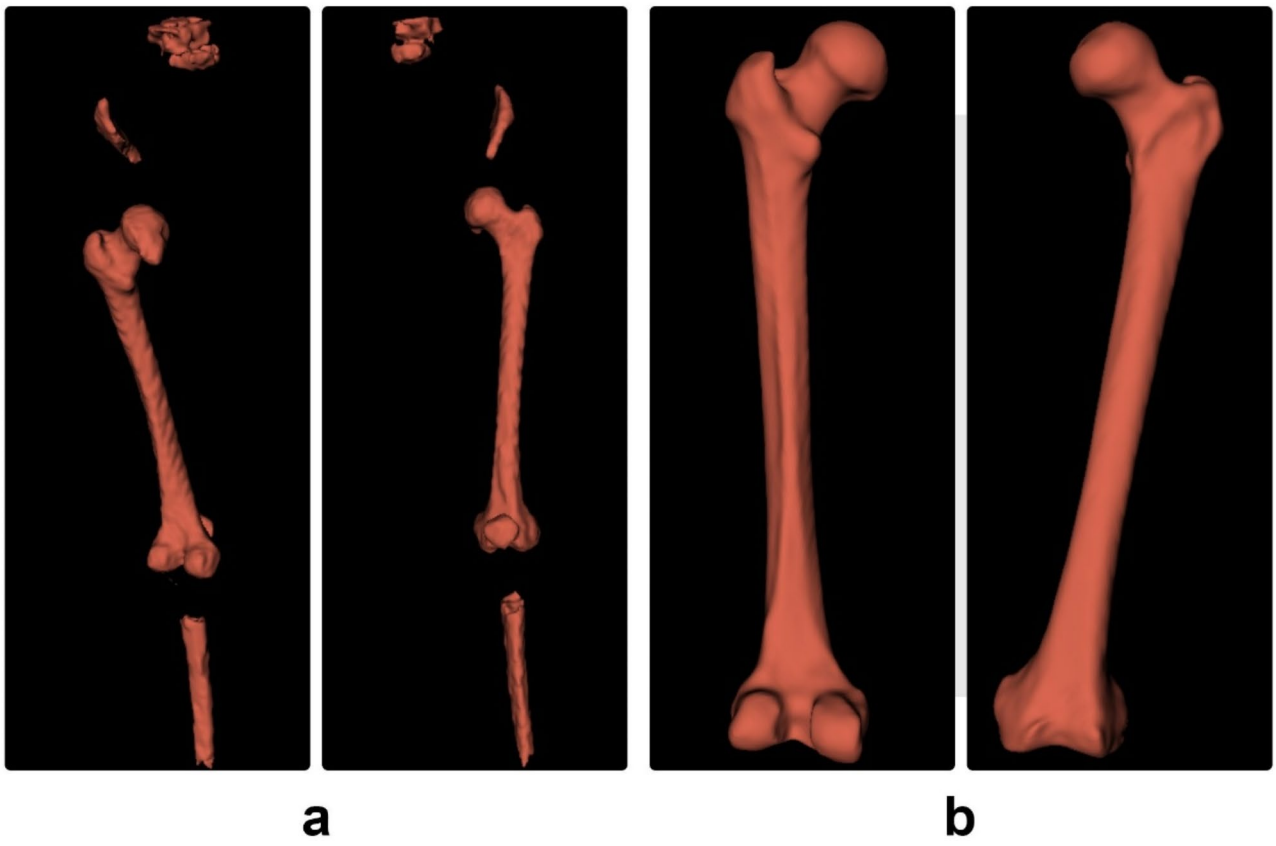

Fig. 4.. 3D reconstruction based on femur masks from Two-Stage SegResNet (**b**) and Non-Two-Stage model (**a**). Since the quality of femur masks determines the performance of 3D reconstruction, the non-two-stage model exhibits poor femur continuity, incomplete reconstruction, and significant non-femur structures (tibia and pelvic tissue). In contrast, the two-stage model generates a smooth and complete femur in the 3D reconstruction.

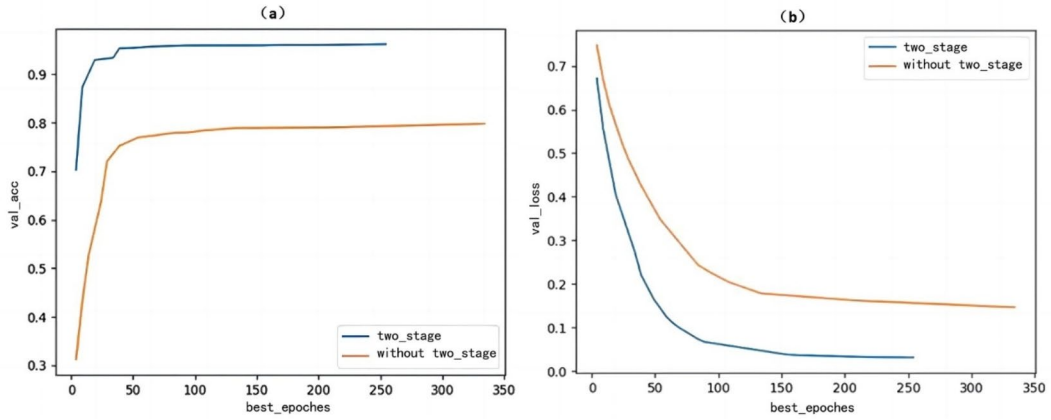

Fig. 5. The accuracy (**a**) and loss values (**b**) of the two networks. As epochs increase, the Two-Stage model shows a lower and faster-reducing loss, along with higher and faster-converging accuracy compared to the Non-Two-Stage model.

Model	Model file size (MB)	Maximum patch size available on the disk	Averaged inference time in test data (s)	Dice
SegResNet	18	(224, 224, 224)	4.2	0.9523
TransUNet	354	(192, 192, 192)	10.7	0.9378

Table 7. SegResNet and TransUNet comparison test results.

superior spatial resolution, particularly for calcified tissues such as bone, thereby providing a more detailed representation and facilitating easier acquisition[93]. This advantage is especially beneficial for bone tissue segmentation and significantly enhances the precision of our model.

Analysis of the results compared with TransUNet

The suboptimal Dice coefficient observed in TransUNet can primarily be attributed to its reliance on patch-based operations, which divide the original image into smaller regions for network input. In femur imaging, these localized patches often exhibit structural similarities, limiting the network's ability to learn meaningful distinctions. In contrast, the SegResNet architecture leverages residual structures to address vanishing gradient issues during backpropagation, facilitating deeper network training and improving representational capacity. Residual blocks also enhance feature reuse, essential for capturing intricate details critical to medical image segmentation. While TransUNet excels at global context modeling, its patch-based strategy limits its applicability in domains with high local similarity, such as femur imaging. SegResNet's robust residual structure offers improved segmentation accuracy, making it more suitable for such tasks. Future research could explore hybrid approaches that combine the strengths of both architectures.

Advantages of the proposed pipeline

The proposed approach capitalizes on the computational efficiency and reduced data requirements of 2D object detection models compared to 3D counterparts. Additionally, the pipeline effectively compensates for the loss of spatial information inherent in 2D projections by combining outputs from coronal and axial detections. This workflow is particularly valuable in medical imaging, where rapid and accurate anomaly localization is critical for diagnosis and treatment planning. Subsequent semantic segmentation and 3D reconstruction further enhance the utility of the approach. Future research could focus on optimizing the ray marching step for real-time processing and integrating advanced 3D fusion techniques to improve bounding box accuracy. The modular nature of the pipeline also allows for incorporating more sophisticated detection models or multimodal imaging data, extending its applicability in clinical workflows.

Limitations and future directions

This study has several limitations. First, the number of models and comparison tests is insufficiently diverse. Second, the dataset comprises only normal physiological femoral images, lacking data from pathological conditions such as femoral fractures or hip dysplasia, which limits the clinical applicability of the model. Additionally, as with many supervised deep learning models, the proposed approach requires manual annotation during the early stages, increasing labor and time costs. Future work will involve collaboration with hospital imaging centers to collect images of relevant pathological conditions and the integration of advanced modules (e.g., D-LKA[94]) or networks (e.g., SNN[95]) to further enhance the generalizability and practical utility of the two-stage segmentation process.

Conclusion

Image segmentation plays a critical role in medical image processing and quantitative analysis. However, medical image segmentation is challenged by several limitations inherent in imaging technologies, including variations in grayscale values, reduced contrast, high noise levels, and indistinct boundaries. This paper proposes a deep learning-based, two-stage segmentation network process specifically designed for femur segmentation in comprehensive lower limb CT scans. The results demonstrate both satisfactory performance and broad applicability. Segmentation of bone tissue is crucial for clinical assessment and management of fractures, underscoring its significant practical implications in medical practice.

Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the protection of personal information and limitation of ethical approvement but are available from the corresponding author on reasonable request.

Received: 20 December 2024; Accepted: 12 March 2025
Published online: 17 March 2025

References

1. Cummings, S. R. & Melton, L. J. Epidemiology and outcomes of osteoporotic fractures. *Lancet* **359**, 1761–1767 (2002).
2. Della Rocca, G. J. & Crist, B. D. Hip fracture protocols. *Orthopedic Clin. N. Am.* **44**, 163–182 (2013).
3. O'Connor, M. I. & Switzer, J. A. AAOS clinical practice guideline summary: Management of hip fractures in older adults. *J Am Acad Orthop Surg* **30**, e1291–e1296 (2022).
4. Rogers, W. et al. Radiomics: From qualitative to quantitative imaging. *BJR* **93**, 20190948 (2020).

5. Kaur, P., Singh, G. & Kaur, P. A review of denoising medical images using machine learning approaches. *CMIR* **14**, 675–685 (2018).
6. Lee, M. S., Park, C. H. & Kang, M. G. Edge enhancement algorithm for low-dose X-ray fluoroscopic imaging. *Comput. Methods Programs Biomed.* **152**, 45–52 (2017).
7. Weaver, A. A., Nguyen, C. M., Schoell, S. L., Maldjian, J. A. & Stitzel, J. D. Image segmentation and registration algorithm to collect thoracic skeleton semilandmarks for characterization of age and sex-based thoracic morphology variation. *Comput. Biol. Med.* **67**, 41–48 (2015).
8. Yanase, J. & Triantaphyllou, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst. Appl.* **138**, 112821 (2019).
9. Sugano, N. Computer-assisted orthopedic surgery. *J. Orthopaedic Sci.* **8**, 442–448 (2003).
10. Xu, Y. et al. Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering* **11**, 1034 (2024).
11. Wang, Y., Ahsan, U., Li, H. & Hagen, M. A comprehensive review of modern object segmentation approaches. *Found. Trends® Comput. Graph. Vis.* **13**, 111–283 (2022).
12. Bhairnallykar, S. T. & Narawade, V. Exploration of image segmentation: A comprehensive review of traditional segmentation techniques. In *2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)* 1–6 (2023). https://doi.org/10.1109/IEMENTech60402.2023.10423439.
13. Bali, A. & Singh, S. N. A review on the strategies and techniques of image segmentation. In *2015 Fifth International Conference on Advanced Computing & Communication Technologies* 113–120 (2015). https://doi.org/10.1109/ACCT.2015.63.
14. Jing, J., Liu, S., Wang, G., Zhang, W. & Sun, C. Recent advances on image edge detection: A comprehensive review. *Neurocomputing* **503**, 259–271 (2022).
15. Rong, W., Li, Z., Zhang, W. & Sun, L. An improved canny edge detection algorithm. In *2014 IEEE International Conference on Mechatronics and Automation* 577–582 (2014). https://doi.org/10.1109/ICMA.2014.6885761.
16. Jardim, S., António, J. & Mora, C. Image thresholding approaches for medical image segmentation - short literature review. *Procedia Comput. Sci.* **219**, 1485–1492 (2023).
17. Adams, R. & Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 641–647 (1994).
18. Ramli, M. F. & Tahar, K. N. Homogeneous tree height derivation from tree crown delineation using seeded region growing (SRG) segmentation. *Geo-spatial Inf. Sci.* **23**, 195–208 (2020).
19. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 888–905 (2000).
20. Chen, X. & Pan, L. A survey of graph cuts/graph search based medical image segmentation. *IEEE Rev. Biomed. Eng.* **11**, 112–124 (2018).
21. Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. & Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* **622**, 178–210 (2023).
22. Sisodia, D., Singh, L. & Sisodia, S. Clustering techniques: A brief survey of different clustering algorithms (2012).
23. Kass, M., Witkin, A. & Terzopoulos, D. Snakes: Active contour models. *Int. J. Comput. Vision* **1**, 321–331 (1988).
24. Caselles, V., Kimmel, R. & Sapiro, G. Geodesic active contours. *Int. J. Comput. Vision* **22**, 61–79 (1997).
25. Chan, T. F. & Vese, L. A. Active contours without edges. *IEEE Trans. Image Process.* **10**, 266–277 (2001).
26. Ji, Z., Xia, Y., Sun, Q., Cao, G. & Chen, Q. Active contours driven by local likelihood image fitting energy for image segmentation. *Inf. Sci.* **301**, 285–304 (2015).
27. Zhang, K., Zhang, L., Song, H. & Zhou, W. Active contours with selective local or global segmentation: A new formulation and level set method. *Image Vision Comput.* **28**, 668–676 (2010).
28. Sun, L., Meng, X., Xu, J. & Zhang, S. An image segmentation method based on improved regularized level set model. *Appl. Sci.* **8**, 2393 (2018).
29. Zhuang, X., Rhode, K. S., Razavi, R. S., Hawkes, D. J. & Ourselin, S. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans Med Imaging* **29**, 1612–1625 (2010).
30. Iglesias, J. E. & Sabuncu, M. R. Multi-atlas segmentation of biomedical images: A survey. *Med. Image Anal.* **24**, 205–219 (2015).
31. Wachinger, C. & Golland, P. Atlas-based under-segmentation. *Med. image comput. comput.-assist. interv.: MICCAI Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* **17**, 315–322 (2014).
32. Bach Cuadra, M., Duay, V. & Thiran, J.-Ph. Atlas-based segmentation. In *Handbook of Biomedical Imaging: Methodologies and Clinical Research* (eds. Paragios, N., Duncan, J. & Ayache, N.) 221–244 (Springer US, Boston, MA, 2015). https://doi.org/10.1007/978-0-387-09749-7_12.
33. Maulik, U. Medical image segmentation using genetic algorithms. *IEEE Trans. Inf. Technol. Biomed.* **13**, 166–173 (2009).
34. Mandal, B. Optimization of quadratic curve fitting from data points using real coded genetic algorithm. In *Emerging Technologies in Data Mining and Information Security* (eds. Hassanien, A. E., Bhattacharyya, S., Chakrabati, S., Bhattacharya, A. & Dutta, S.) 419–428 (Springer, Singapore, 2021). https://doi.org/10.1007/978-981-15-9927-9_41.
35. Panayides, A. S. et al. AI in medical imaging informatics: Current challenges and future directions. *IEEE J. Biomed. Health Inform.* **24**, 1837–1857 (2020).
36. Xia, Q. et al. A comprehensive review of deep learning for medical image segmentation. *Neurocomputing* **613**, 128740 (2025).
37. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer, Cham, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
38. Milletari, F., Navab, N. & Ahmadi, S.-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* 565–571 (2016). https://doi.org/10.1109/3DV.2016.79.
39. Huang, H. *et al.* UNet 3+: A full-scale connected UNet for medical image segmentation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1055–1059 (2020). https://doi.org/10.1109/ICASSP40776.2020.9053405.
40. Oktay, O. *et al.* Attention U-net: Learning where to look for the pancreas. Preprint at https://doi.org/10.48550/arXiv.1804.03999 (2018).
41. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203–211 (2021).
42. Cai, S. et al. Dense-UNet: A novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. Quant. *Imaging Med. Surg.* **10**, 1275–1285 (2020).
43. Vaswani, A. et al. Attention is all you need. in Proceedings of the 31st International Conference on Neural Information Processing Systems 6000–6010 (Curran Associates Inc., Red Hook, NY, USA, 2017).
44. Karimi, D., Vasylechko, S. D. & Gholipour, A. Convolution-free medical image segmentation using transformers. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (eds. de Bruijne, M. et al.) 78–88 (Springer International Publishing, Cham, 2021). https://doi.org/10.1007/978-3-030-87193-2_8.
45. Cao, H. *et al.* Swin-unet: Unet-like pure transformer for medical image segmentation. In *Computer Vision – ECCV 2022 Workshops* (eds. Karlinsky, L., Michaeli, T. & Nishino, K.) 205–218 (Springer Nature Switzerland, Cham, 2023). https://doi.org/10.1007/978-3-031-25066-8_9.
46. Chen, J. *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. Preprint at https://doi.org/10.48550/arXiv.2102.04306 (2021).

47. Lin, A. et al. DS-TransUNet: Dual swin transformer U-net for medical image segmentation. *IEEE Trans. Instrum. Meas.* **71**, 1–15 (2022).
48. Shaker, A. et al. UNETR++: Delving into efficient and accurate 3D medical image segmentation. *IEEE Trans. Med. Imag.* **43**, 3377–3390 (2024).
49. Gu, A. & Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. Preprint at https://doi.org/10.48550/arXiv.2312.00752 (2024).
50. Ruan, J., Li, J. & Xiang, S. VM-UNet: Vision mamba UNet for medical image segmentation. Preprint at https://doi.org/10.48550/arXiv.2402.02491 (2024).
51. Liao, W. *et al.* LightM-UNet: mamba assists in lightweight UNet for medical image segmentation. Preprint at https://doi.org/10.48550/arXiv.2403.05246 (2024).
52. Liu, J. *et al.* Swin-UMamba: Mamba-based UNet with ImageNet-based pretraining. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024* (eds. Linguraru, M. G. et al.) 615–625 (Springer Nature Switzerland, Cham, 2024). https://doi.org/10.1007/978-3-031-72114-4_59.
53. Peng, J. & Wang, Y. Medical image segmentation with limited supervision: A review of deep network models. *IEEE Access* **9**, 36827–36851 (2021).
54. Yu, L., Wang, S., Li, X., Fu, C.-W. & Heng, P.-A. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (eds. Shen, D. et al.) 605–613 (Springer International Publishing, Cham, 2019). https://doi.org/10.1007/978-3-030-32245-8_67.
55. Li, S., Zhang, C. & He, X. Shape-aware semi-supervised 3D semantic segmentation for medical images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (eds. Martel, A. L. et al.) 552–561 (Springer International Publishing, Cham, 2020). https://doi.org/10.1007/978-3-030-59710-8_54.
56. You, C. *et al.* ACTION++: Improving semi-supervised medical image segmentation with adaptive anatomical contrast. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023* (eds. Greenspan, H. et al.) 194–205 (Springer Nature Switzerland, Cham, 2023). https://doi.org/10.1007/978-3-031-43901-8_19.
57. Pathak, D., Krähenbühl, P. & Darrell, T. Constrained convolutional neural networks for weakly supervised segmentation. In *2015 IEEE International Conference on Computer Vision (ICCV)* 1796–1804 (2015). https://doi.org/10.1109/ICCV.2015.209.
58. Cai, W. et al. DFTNet: Dual-path feature transfer network for weakly supervised medical image segmentation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **20**, 2530–2540 (2023).
59. Kuang, Z., Yan, Z., Zhou, H. & Yu, L. Cluster-re-supervision: Bridging the gap between image-level and pixel-wise labels for weakly supervised medical image segmentation. *IEEE J. Biomed. Health Inform.* **27**, 4890–4901 (2023).
60. Besler, B. A. et al. Bone and joint enhancement filtering: Application to proximal femur segmentation from uncalibrated computed tomography datasets. *Medical Image Analysis* **67**, 101887 (2021).
61. Xie, W. et al. Statistical model-based segmentation of the proximal femur in digital antero-posterior (AP) pelvic radiographs. *Int. J. Comput. Assisted Radiol. Surg.* **9**, 165–176 (2014).
62. Xia, Y. et al. Automated bone segmentation from large field of view 3D MR images of the hip joint. *Phys. Med. Biol.* **58**, 7375–7390 (2013).
63. Ben Younes, L., Nakajima, Y. & Saito, T. Fully automatic segmentation of the femur from 3D-CT images using primitive shape recognition and statistical shape models. *Int J Comput. Assist. Radiol. Surg.* **9**, 189–196 (2014).
64. Arezoomand, S., Lee, W.-S., Rakhra, K. S. & Beaulé, P. E. A 3D active model framework for segmentation of proximal femur in MR images. *Int. J. Comput. Assisted Radiol. Surg.* **10**, 55–66 (2015).
65. Chandra, S. S. et al. Focused shape models for hip joint segmentation in 3D magnetic resonance images. *Med. Image Anal.* **18**, 567–578 (2014).
66. Yokota, F. *et al.* Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* (eds. Mori, K., Sakuma, I., Sato, Y., Barillot, C. & Navab, N.) 190–197 (Springer, Berlin, Heidelberg, 2013). https://doi.org/10.1007/978-3-642-40763-5_24.
67. Yokota, F. *et al.* Automated CT segmentation of diseased hip using hierarchical and conditional statistical shape models. *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* **16**, 190–197 (2013).
68. Aldieri, A. et al. Development and validation of a semi-automated and unsupervised method for femur segmentation from CT. *Sci Rep* **14**, 7403 (2024).
69. Damopoulos, D. et al. Segmentation of the proximal femur in radial MR scans using a random forest classifier and deformable model registration. *Int J Comput Assist Radiol Surg* **14**, 545–561 (2019).
70. Hussain, D., Al-Antari, M. A., Al-Masni, M. A., Han, S.-M. & Kim, T.-S. Femur segmentation in DXA imaging using a machine learning decision tree. *J Xray Sci Technol* **26**, 727–746 (2018).
71. Cheng, Y. et al. Automatic segmentation technique for acetabulum and femoral head in CT images. *Pattern Recognit.* **46**, 2969–2984 (2013).
72. Zhao, C. et al. ST-V-net: Incorporating shape prior into convolutional neural networks for proximal femur segmentation. *Complex Intell Syst.* **9**, 2747–2758 (2023).
73. Chen, F., Liu, J., Zhao, Z., Zhu, M. & Liao, H. Three-dimensional feature-enhanced network for automatic femur segmentation. *IEEE J. Biomed. Health Inform.* **23**, 243–252 (2019).
74. Zhu, L. et al. An automatic classification of the early osteonecrosis of femoral head with deep learning. *Curr. Med. Imaging* **16**, 1323–1331 (2020).
75. Zeng, G. *et al.* 3D U-net with Multi-level Deep Supervision: Fully Automatic Segmentation of Proximal Femur in 3D MR Images. In *Machine Learning in Medical Imaging* (eds. Wang, Q., Shi, Y., Suk, H.-I. & Suzuki, K.) vol. 10541 274–282 (Springer International Publishing, Cham, 2017).
76. Zhang, D. et al. Using 2D U-net convolutional neural networks for automatic acetabular and proximal femur segmentation of hip MRI images and morphological quantification: A preliminary study in DDH. *Biomed. Eng. Online* **23**, 98 (2024).
77. Kim-Wang, S. Y. et al. Auto-segmentation of the tibia and femur from knee MR images via deep learning and its application to cartilage strain and recovery. *J Biomech* **149**, 111473 (2023).
78. Zeng, G. & Zheng, G. Deep learning-based automatic segmentation of the proximal femur from MR images. *Adv Exp Med Biol* **1093**, 73–79 (2018).
79. Deng, Y. et al. A deep learning-based approach to automatic proximal femur segmentation in quantitative CT images. *Med Biol Eng Comput* **60**, 1417–1429 (2022).
80. Apivanichkul, K., Phasukkit, P., Dankulchai, P., Sittiwong, W. & Jitwatcharakomol, T. Enhanced deep-learning-based automatic left-femur segmentation scheme with attribute augmentation. *Sens. (Basel Switz.)* **23**, 5720 (2023).
81. Zhang, X. et al. Femoral image segmentation based on two-stage convolutional network using 3D-DMFNet and 3D-ResUnet. *Comput. Methods Programs Biomed.* **226**, 107110 (2022).
82. Cai, L., Gao, J. & Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **8**, 713 (2020).
83. Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J Digit Imaging* **32**, 582–596 (2019).

84. Staub, D. & Murphy, M. J. A digitally reconstructed radiograph algorithm calculated from first principles. *Med. Phys.* **40**, 11902 (2013).
85. Shah, S. M. A. H., Rizwan, A., Atteia, G. & Alabdulhafith, M. CADFU for dermatologists: A novel chronic wounds and ulcers diagnosis system with DHuNeT (dual-phase hyperactive UNet) and YOLOv8 algorithm. *Healthcare* **11**, 2840 (2023).
86. Adegun, A. A., Fonou Dombeu, J. V., Viriri, S. & Odindi, J. State-of-the-art deep learning methods for objects detection in remote sensing satellite images. *Sensors* **23**, 5849 (2023).
87. Terven, J. R. & Esparza, D. M. C. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv.org* (2023).
88. Myronenko, A. 3D MRI brain tumor segmentation using autoencoder regularization. Preprint at http://arxiv.org/abs/1810.11654 (2018).
89. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90.
90. Agarwal, R., Ghosal, P., Sadhu, A. K., Murmu, N. & Nandi, D. Multi-scale dual-channel feature embedding decoder for biomedical image segmentation. *Comput. Methods Programs Biomed.* **257**, 108464 (2024).
91. Agarwal, R. *et al.* Deep quasi-recurrent self-attention with dual encoder-decoder in biomedical CT image segmentation. *IEEE J. Biomed. Health Inform.* (2024).
92. Pilgram, R., Walch, C., Kuhn, V., Schubert, R. & Staudinger, R. Proximal femur segmentation in conventional pelvic x ray. *Med Phys* **35**, 2463–2472 (2008).
93. Carrino, J. A. et al. CT in musculoskeletal imaging: Still helpful and for what?. *Skeletal Radiol.* **53**, 1711–1725 (2024).
94. Azad, R. *et al.* Beyond self-attention: deformable large kernel attention for medical image segmentation. Preprint at https://doi.org/10.48550/arXiv.2309.00121 (2023).
95. Agarwal, R., Ghosal, P., Murmu, N. & Nandi, D. Spiking neural network in computer vision: Techniques, tools and trends. In *Advanced Computational and Communication Paradigms* (eds. Borah, S., Gandhi, T. K. & Piuri, V.) 201–209 (Springer Nature, Singapore, 2023). https://doi.org/10.1007/978-981-99-4284-8_16.

Author contributions

W.X., P.C, Z.L., X.W. and C.W. contributed to the overall study design; L.Z., W.W. and J.X. contributed to clinical sample collection; L.Z., and J.X. contributed to experiments; L.Z., W.W. and J.X. contributed to data analysis; X.W. and C.W. contributed to data curation; W.X. and P.C contributed to original draft writing; Y.W. and D.Z. contributed to draft review; D.Z. contributed to project administration; Y.W. and D.Z. contributed to funding acquisition. All authors read, discussed, and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81974360), the Innovation Foundation of National Orthopedics and Sports Rehabilitation Clinical Medicine Research Center (No. 2021-NCRC-CXJJ-PY-37), the Independent Exploration and Innovation Project for Postgraduate Students of Central South University (2024ZZTS0163) and the Key Technology R and D Program of Jiangsu (No. BE2022718).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

The study was approved by the Ethics Committee of Xiangya Hospital, Central South University (No: 202310926). A requirement for informed consent was waived for this study by the approval of the Ethics Committee of Xiangya Hospital, Central South University. Clinical data was de-identified and analyzed anonymously. The study was compliant with the ethical guidelines of the Declarations of Helsinki.

Additional information

Correspondence and requests for materials should be addressed to Y.W. or D.Z.