Computerized Medical Imaging and Graphics 123 (2025) 102555

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

ICA-SAMv7: Internal carotid artery segmentation with coarse to fine network

Xiaotian Yan [a,b], Yuting Guo [a], Ziyi Pei [a], Xinyu Zhang [b], Jinghao Li [a], Zitao Zhou [b], Lifang Liang [a], Shuai Li [a,*], Peng Lun [c], Aimin Hao [a]

[a] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, 100191, China
[b] Qingdao Virtual Reality Institute Co., Ltd, Qingdao, Shandong, 266100, China
[c] Neurosurgery Department, Affiliated Hospital of Qingdao University, Qingdao, Shandong, 266555, China

ARTICLE INFO

Keywords:
Medical image segmentation
Segment anything model
Yolov7
Carotid artery stenosis

ABSTRACT

Internal carotid artery (ICA) stenosis is a life-threatening occult disease. Using Computed Tomography Angiography (CTA) to examine vascular lesions such as calcified and non-calcified plaques in cases of carotid artery stenosis is a necessary clinical step in formulating the correct treatment plan. Segment Anything Model (SAM) has shown promising performance in image segmentation tasks, but it performs poorly for carotid artery segmentation. Due to the small size of the calcification and the overlapping between the lumen and calcification, these challenges lead to issues such as mislabeling and boundary fragmentation, as well as high training costs. To address these problems, we propose a two-stage Carotid Artery lesion segmentation method called ICA-SAMv7, which performs coarse and fine segmentation based on the YOLOv7 and SAM model. Specifically, in the first stage (ICA-YOLOv7), we utilize YOLOv7 for coarse vessel recognition, introducing connectivity enhancement to improve accuracy and achieve precise localization of small target carotid artery. In the second stage (ICA-SAM), we enhance SAM through data augmentation and an efficient parameter fine-tuning strategy. This improves the segmentation accuracy of fine-grained lesions in blood vessels while saving training costs. Ultimately, the accuracy of lesion segmentation under the SAM model was increased from the original 48.62% to 83.69%. Extensive comparative experiments have demonstrated the outstanding performance of our algorithm. Our codes can be found at https://github.com/BessiePei/ICA-SAMv7.

1. Introduction

Carotid stenosis (CS) is defined as narrowing or blockage of the carotid artery caused by plaque accumulation, which increases the risk of cerebrovascular disease (Ismail et al., 2023). It is one of the main causes of acute ischemic stroke, accounting for approximately 20% of cases (Arasu et al., 2021). A study conducted in 2019 investigated the relationship between carotid stenosis and stroke incidence. Approximately half of the 2707 patients (50%) in the study experienced ischemic event. Among the 2707 patients examined, 99 cases (7.9%) showed symptomatic carotid stenosis (Cheng et al., 2019). Patients with carotid stenosis greater than 75% had an annual stroke rate of 3.3%. The incidence of cardiovascular events (8.3%) and mortality (6.5%) was higher in patients with severe carotid stenosis (Mohd et al., 2023). Early carotid occlusion may have no obvious symptoms, making it difficult for patients to detect and seek medical attention in a timely manner, which greatly increases the difficulty of treatment. The traditional method for detecting carotid artery stenosis is that doctors obtain patients' CTA images through computed tomography angiography and judge the degree of vascular stenosis based on experience. Doctors visually identify lesion areas in CTA images and manually record lesion types and distribution ranges. This relies too much on the individual doctor's experience, exhibits strong subjectivity, and consumes a considerable amount of time and effort.

In recent years, deep learning has become increasingly popular in medical image segmentation and is commonly used in areas such as breast tumor screening, brain tumor segmentation, skin lesion segmentation (Dhamija et al., 2022). However, blood vessel tissues are tiny and more difficult to identify compared to visceral organs, tumors, and other pathological tissues. Research on blood vessels mainly focuses on the detection of vascular stenosis areas, with few studies specifically addressing the segmentation of lesions within the vessels. For example, the 2D ResU-Net vascular stenosis automatic detection model and the 3D ResU-Net multi-feature sequence callback model developed by Fu et al. (2023) are used for vascular stenosis detection and plaque classification in head and neck CTA images. Yang et al. (2020) proposed a multi-scale level sets method based on Frangi to segment retinal vessels from fundus images. Chaudhry et al. (2013) utilized an active contour

* Corresponding author.
E-mail address: lishuai@buaa.edu.cn (S. Li).

https://doi.org/10.1016/j.compmedimag.2025.102555
Received 12 October 2024; Received in revised form 27 February 2025; Accepted 12 April 2025
Available online 25 April 2025

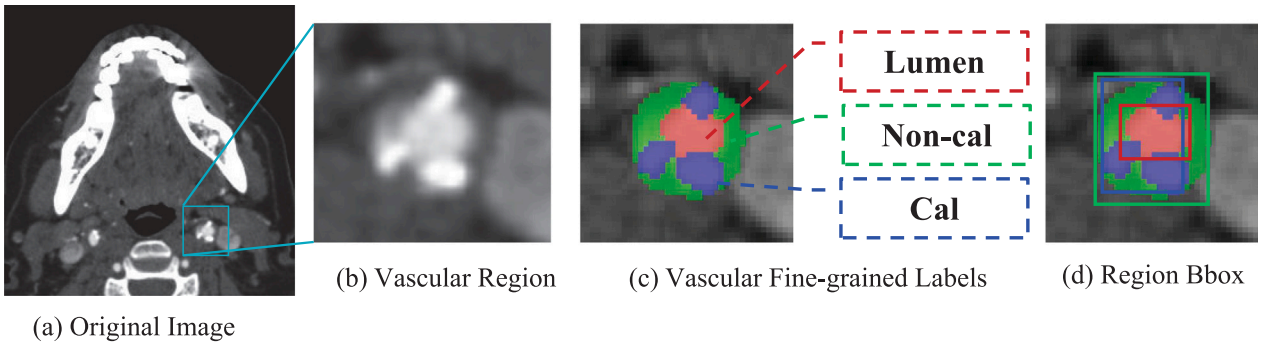

Fig. 1. The substances inside the vessel and their Bboxes prompts. We aim to develop a fine-grained segmentation algorithm for carotid artery vascular lesions that can segment lumen, non-calcified plaque and calcified plaque. The segmentation of vessels and calcified plaque in carotid artery CTA scans presents the following characteristics: (1) poor CTA image quality, (2) small vascular areas, and (3) overlapping bounding boxes between components. These factors all contribute to increased difficulty in image segmentation.

model algorithm for automatic segmentation of carotid artery images and used a Support Vector Machine (SVM) classifier for vascular image classification. Tetteh et al. (2018) employed a deep vessel network (DeepVesselNet) for vessel segmentation, which is a fully convolutional neural network composed of four convolutional layers and a sigmoid classifier. The deep learning algorithms developed for vascular diseases mentioned above lack fine-grained segmentation of vascular lesions and incur high training costs. Therefore, we aim to develop a fine-grained segmentation algorithm for carotid artery vascular lesions that can segment the specific lesion components in blood vessels (Fig. 1).

The segment anything model (SAM) (Kirillov et al., 2023) is a recently emerged foundational model used for image segmentation. It can take input from the original image with specific prompts (points, boxes, shadows, text) and produce different predicted segmentation results based on different hint information. Leveraging the advantages of SAM in achieving zero-shot segmentation tasks holds great promise for providing significant assistance in medical automated pre-diagnosis. Therefore, there is considerable attention on how to better apply the SAM model to the segmentation of carotid artery (Wu et al., 2023). However, training the SAM foundational model on traditional color image parameter sets and applying it to the CTA images of the carotid artery yields poor results. With the rapid growth of the SAM model size and the increasing number of model parameters, significant challenges are posed to computational power and storage space. In response to the above issues, we propose a two-stage segmentation network that progresses from coarse to fine for segmenting carotid artery lesions. In the first stage, the network locates the carotid artery vascular region to address inaccuracies in small target localization. In the second stage, we utilize an efficient parameter fine-tuning method to ensure precise segmentation of carotid artery calcification with minimal training resource expenditure for the SAM model.

Our two-stage segmentation algorithm for internal carotid artery vascular lesions (ICA-SAMv7) involves two steps. The first step utilizes the carotid artery localization algorithm (ICA-YOLOv7) to magnify the localization of carotid artery vessels in CTA images. The second step employs the vascular lesion segmentation algorithm (ICA-SAM) to perform precise segmentation tasks on specific lesion tissues within the vessels. To tackle the problem of the large size of the SAM model, efficient parameter fine-tuning methods such as LoRA parameter adjustment (Hu et al., 2021) are incorporated.

The contributions of this work are as follows:

1. We propose a two-stage segmentation algorithm for internal carotid artery vascular lesions (ICA-SAMv7) based on SAM. Firstly, use the YOLOv7 algorithm with the improvement of 3D connectivity to achieve the localization and coarse segmentation of the bilateral carotid arteries in CTA images. Then, utilize parameter-efficient fine-tuning with SAM to precisely segment the lesion tissues within the vascular lumen.
2. We propose a bilateral carotid artery vessel coarse localization algorithm (ICA-YOLOv7) and perform connectivity verification using Connectivity Enhancement Model (CEM). The ICA-YOLOv7 algorithm provides two types of prompts, automatic prompts from YOLOv7 and manual prompts from doctors, for the ICA-SAMv7 large model, enabling the algorithm to adapt to clinical applications and achieve precise localization of bilateral carotid artery vessels and accurate segmentation of tiny lesions.
3. We propose an efficient parameter fine-tuning method based on SAM. We introduce the LoRA fine-tuning for SAM's image encoder, which is cost-effective and yields good results, enhancing the diagnostic accuracy of fine segmentation of lesions after enlarging tiny vascular areas.

2. Related work

2.1. Medical image segmentation

Medical image segmentation research plays a crucial role in the rapid diagnosis of diseases. Since the emergence of the concept of deep learning, many deep learning algorithms have been applied to medical image segmentation, such as FCN (Long et al., 2015), U-Net (Ronneberger et al., 2015), UNet++ (Ronneberger et al., 2015), and others. Based on these networks, several variants have been proposed. R2U-Net architecture was proposed by Alom et al. (2018), which integrates the structures of U-Net, ResNet, and RCNN (Liang and Hu, 2015). It has achieved excellent experimental results in multiple medical image segmentation tasks such as vessels, lungs, and retinas. BCDU-Net (Azad et al., 2019) incorporates bidirectional ConvLSTM into the skip connection path to produce finer segmentation results in medical images. In 2019, Perslev et al. (2019) proposed a MPUNet segmentation model based on multi-view data augmentation, which accurately completes 13 medical image segmentation tasks without the need for hyperparameter tuning after training. HyperDenseNet architecture (Dolz et al., 2018) fused multi-modal images for medical image segmentation. Although the above methods have been applied in the field of medical image segmentation, they are mostly suitable for tumor segmentation of the top ten human organs such as the brain, liver, kidney, cardiopulmonary system, and spleen in CTA images (Gibson et al., 2018). These algorithms still lack the ability to accurately locate tiny blood vessels in CTA medical images and precisely segment tiny lesions within vascular cross-sections.

2.2. YOLO

In 2016, Redmon (2016) proposed a one-stage detection network named YOLO. Compared with the traditional two-stage detection algorithms represented by RCNN and Faster R-CNN, YOLO has a very high detection speed. It can process 45 frames of images per second and

is capable of detecting targets in real time while ensuring accuracy. The YOLOv7 algorithm (Wang et al., 2023a), introduced in 2023, integrates flexible and efficient training tools with a novel architecture and a compound scaling method, providing a trainable, bag-of-freebies-oriented solution. By the time we started this research, YOLOv7 had surpassed all known object detectors in both speed and accuracy within the range of 5 FPS to 120 FPS, achieving the highest accuracy of 56.8%.

YOLOv7 can be applied to detect blood vessels in medical images. However, it fails to achieve fine-grained lesion segmentation. Still, it can be used to pinpoint carotid arteries. For examples, Wu et al. (2024) proposed a deep learning-based vascular recognition system for ultrasound-guided venous puncture. They designed a lightweight vascular ultrasound network (UV-YOLOv7) to identify puncture veins, achieving a recognition accuracy with an mAP value of 86.2% and an inference time of 0.6 ms. To assist doctors in quantifying blood vessels in pathological images, Jiale et al. (2024) proposed an enhanced YOLOv7-based detection network named PI-YOLO for pathological image vascular detection, achieving a recognition accuracy with an mAP value of 87.48%. Wang et al. (2024) developed an integrated deep learning model for the localization and classification of coronary artery stenosis in coronary angiography images. This model successfully detected and classified the severity of coronary artery stenosis, with an mAP value of 87.5%.

2.3. SAM in medical imaging

Kirillov et al. (2023) proposed the SAM (Segment anything model) image segmentation foundational model, which differs from traditional segmentation models. This model can predict segmentation results without the need for training on the target dataset, solely based on prompting information. SAM is designed to take both the original image and specific prompts (points, boxes, shadows, text) as input and generate different segmentation result images based on different prompts. It is suitable for various scenarios, including interactive segmentation, boundary detection, super-resolution, object generation, foreground segmentation, semantic segmentation, instance segmentation, and panoramic segmentation. SAM was demonstrated with high accuracy, real-time performance, and significant potential for application.

Since the rise of the SAM model, many researchers have attempted to apply the SAM mega-model to medical image segmentation. For example, Ma et al. (2024) trained the SAM model on a dataset containing 33 medical image segmentation tasks. To reduce computation, they froze the image encoder and prompt encoder of SAM, and only updated the parameters of the mask decoder, resulting in the MedSAM model. This improved the accuracy of SAM model in mask prediction on medical images, but this model only supports prediction using Bounding box (Bbox) prompts. Some researchers have also attempted to fine-tune the SAM model. For instance, Cheng et al. (2023) proposed the SAM-Med2D model, which adds an Adapter Layer to the Image Encoder part of SAM, enabling the integration of medical domain knowledge into the image encoder at a lower cost. SAM-CLIP (Wang et al., 2023b) attempted to combine the SAM model with the CLIP mega-model. ClipSAM (Li et al., 2024) attempted to use CLIP to obtain coarse segmentation results and then use SAM for fine segmentation to obtain the model. However, with the rapid increase in the size of the SAM mega-model, the number of parameters that need to be updated continues to increase, posing a significant challenge to both computational power and storage space.

2.4. Parameter-efficient fine-tuning

Since the advent of large-scale models, the number of parameters has continued to increase, reaching billions or even trillions of parameters (Shoeybi et al., 2019), making it difficult to train on GPUs due to their limited memory capacity. Parameter-Efficient Fine-Tuning (PEFT) (Liu et al., 2022) addressed these issues by training only a small number of parameters, such as a subset of existing model parameters or newly added parameters, thereby providing higher memory efficiency, training speed, model quality, and inference cost efficiency. In recent years, researchers have proposed various PEFT methods, which can be classified into methods involving adding extra parameters, selecting parameter update methods, reconstructing the original model parameters, and hybrid methods (Lialin et al., 2023). Houlsby et al. (2019) proposed adding an additional fully connected layer after the attention and feed-forward layers in the Transformer. BitFit (Zaken et al., 2021), fine-tuned only the biases of the network, meaning that for each linear layer or convolutional layer, the weight matrix W remains unchanged, and only the bias vector b is optimized. Sung et al. (2021) proposed a sparse update method that selects the top p parameters for parameter updates using the model parameter's Fisher information. LoRA (Hu et al., 2021) introduced a simple low-rank fine-tuning method, which reconstructs the original model parameters, decomposing the parameter update of the weight matrix into the product of two low-rank matrices. They greatly reduce the number of model update parameters, improve computational efficiency, and maintain high accuracy, making them suitable for large-scale models.

3. Method

3.1. Overview

The method architecture we proposed mainly consists of two stages (Fig. 2). Firstly, CTA images of the carotid artery are input into the YOLOv7 image detection network. Then, the 3D information is introduced for connectivity enhancement, focusing the detection targets on the area surrounding the carotid artery. This helps eliminate interference from other morphologically similar irrelevant blood vessels for fine-grained recognition of carotid artery lesions. Furthermore, we introduce the SAM model in stage two for target vessel fine-grained segmentation. In this process, considering the model's large parameter size and training cost, we introduce the efficient parameter fine-tuning strategy LoRA into the image encoder.

3.2. ICA-YOLOv7

To solve the problem of localizing tiny carotid arteries in large CTA images, we use YOLOv7 to detect their positions. Then, with the connectivity enhancement model, we supplement missing vessels, remove interfering ones, and improve the localization accuracy. First, we start by using the YOLOv7 algorithm for initial blood vessel detection. The backbone employs CSPDarknet, while the head utilizes depthwise separable convolution, reducing computational cost while preserving feature extraction capability. Given input CTA image $I \in \mathbb{R}^{H\times D}$, we use the vanilla YOLOv7 network to obtain the vessel bounding box $B = \{x1, y1, x2, y2\}$, where $(x1, y1)$ and $(x2, y2)$ are the coordinates of the vertices in the upper left and lower right respectively.

After using the YOLOv7 network for segmentation, we found that there are some mislabeling areas to locate other blood vessels as carotid arteries, and missing carotid artery blood vessels (Fig. 3). Therefore, we further improve the output data by leveraging the 3D connectivity of the carotid artery vessels.

Consecutive CTA of the same carotid artery exhibit minimal displacement and a high Intersection over Union (IoU). We calculate the IoU of YOLOv7-generated vessel prediction boxes across adjacent CTAs to identify the primary carotid artery chain. Missed detection boxes within the main chain are supplemented, and incorrectly labeled branch vessels are removed. This process corrects YOLOv7's mislabeling errors and improves carotid artery localization accuracy.

Specifically, regarding the prediction boxes, i.e. the vessel bounding box $B_{b_i} = \{x1, y1, x2, y2\}$, of bilateral carotid arteries in CTA automatically identified by YOLOv7 in the first stage. We design a 2D–3D fusion

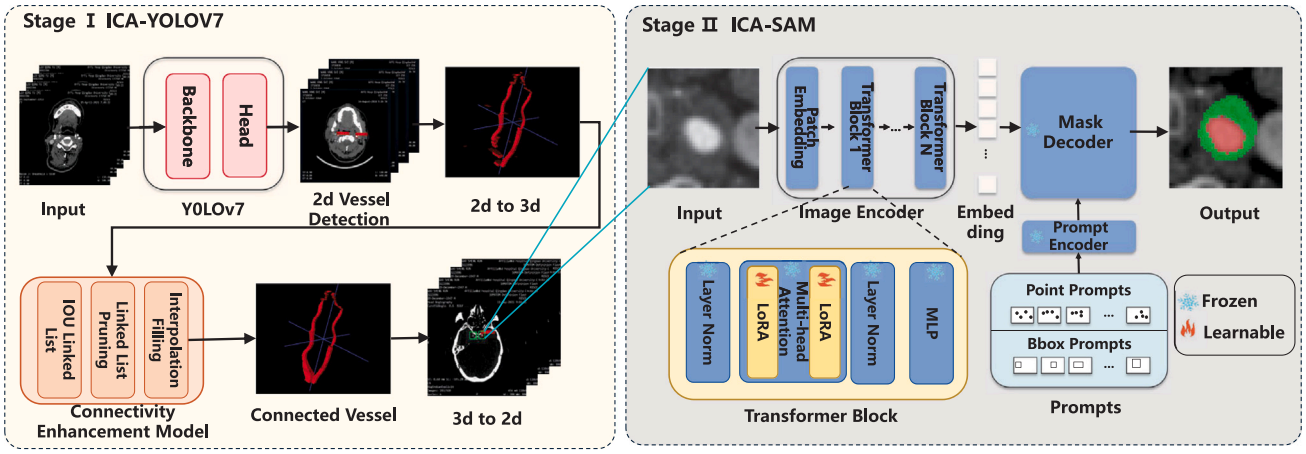

Fig. 2. Whole structure of our method. In stage I, we use YOLOv7 for carotid artery detection. The Connectivity Enhancement Model addresses missed detections and filters interfering vessels, improving YOLOv7's accuracy in vessel localization. In stage II, on the basis of the ICA-YOLOv7 prediction results and point prompt learning, we use Low-Rank Adaptation (LoRA) to fine-tune the Segment Anything Model (SAM) to obtain a suitable model for our project.

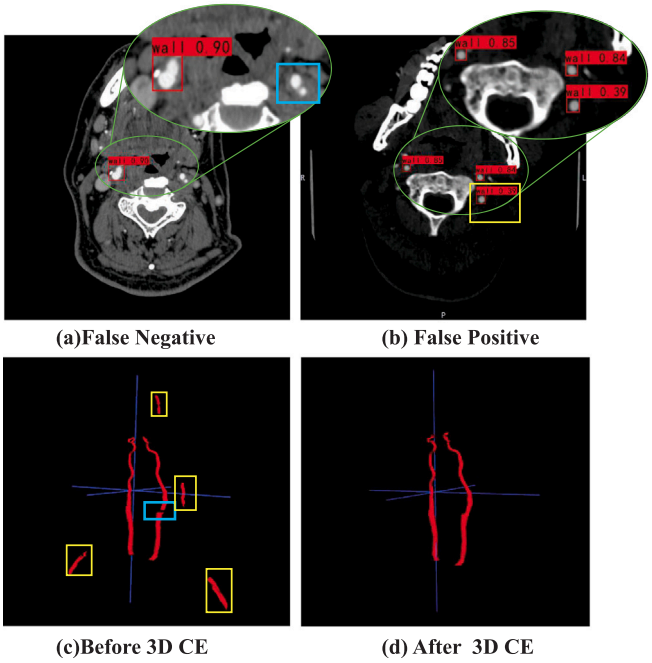

Fig. 3. Issues in carotid artery lesion segmentation. The numbers in the red boxes represent the confidence levels of the prediction boxes. In figure (a), the blue box marks the missed vessels, while in figure (b), the yellow box marks the wrong vessels. The 3D Connectivity Enhancement Model (CE) helps connect missed carotid artery vessels and eliminate incorrectly detected branches. Figure (c) shows the result before 3D CE, and figure (d) shows the improved result after applying 3D CE.

method to enhance the connectivity of blood vessels. The set of all bounding boxes (bboxes) in all slices is $\mathcal{B} = \{B_1, B_2, \ldots, B_{b_i}, \ldots, B_n\}$, where $B_{b_i} = \{B^1_{b_i}, B^2_{b_i}, \ldots, B^t_{b_i}, \ldots\}$ represents the bbox set of the b_ith slice. Since there is no overlap in the axial direction between vessels on different sides, i.e., the IOU between adjacent slices is 0, $\mathcal{B}$ is comprised of two graph structures where single-side bboxes act as nodes and the IOU between adjacent slices' bboxes serve as edges. We illustrate how to find an appropriate bbox path as the vessel path from one of these graphs. Specifically, we set a threshold τ, and for each bbox in B_{b_i}, we calculate the IOU with all bboxes in the next k slices:

$$IOU_{b_i b_j} = \frac{B^t_{b_i} \cap B^t_{b_j}}{B^t_{b_i} \cup B^t_{b_j}}; \quad 0 < |b_i - b_j| < k, \tag{1}$$

where b_i and b_j represent the b_ith and b_jth slices, respectively. This determines whether there is an edge between two bboxes. Then, based on the physical characteristic of the carotid artery being the longest, we use topological sorting to find the longest path containing the most nodes in the graph $\{B^t_{b_1}, B^t_{b_2}, \ldots\}$, where $B^t_{b_i} \in B_{b_i}$ and b_i and b_{i+1} are not necessarily adjacent. For non-adjacent bboxes, we interpolate to supplement the undetected vessels in the middle:

$$\hat{B}_{(b_i)+1} = Interpolate(B^t_{b_i}, B^t_{b(i+1)}). \tag{2}$$

This results in the final complete sequence $\{\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_{b_i}, \ldots\}$, where $\hat{B}_{b_i}$ represents the retained bboxes after keeping only one bbox per slice.

Finally, the main chain of the carotid artery vessels is processed. Using linear interpolation, the coordinate values of the predicted CTA vessel bounding boxes from adjacent frames are utilized to reconstruct the missing vessel segments. Meanwhile, the prediction boxes of the branch vessels that are mislabeled outside the main chain are also removed at this stage. Ultimately, only the vessel prediction boxes of the bilateral continuous carotid arteries are retained in the two-dimensional CTA images, ensuring the accuracy and continuity of carotid artery detection. The whole process can be divided into three parts: three-dimensional vascular connectivity verification, vascular screening, and vascular completion, as shown in Fig. 4

3.3. ICA-SAM

We optimize the SAM model in three aspects, reducing the number of model-trainable parameters and further improving the detection accuracy of object detection data.

3.3.1. Image augmentation

First, we utilize YOLO segmentation and connectivity enhancement to improve the data and crop the head CTA images to obtain image slices focused on the carotid artery vessels. The resulting cropped images have fewer pixels, with an average of approximately 40 × 40 pixels per 2D vessel image, accounting for only 2.5% of the original input head CTA image, as shown in Fig. 1(a). Our goal is to enhance these images while preserving the original information to better suit the requirements of our fine-grained segmentation model. Since CTA images are grayscale with a single channel and pixel values ranging from 0 to 255. The default image processing approach of SAM-Med2D (Cheng et al., 2023) involves copying the original grayscale values into three channels and applying normalization using 'pixel_mean' and 'pixel_std', which results in the visualization shown in Fig. 5(b). It is clear that this processing method is not appropriate for our data. Therefore,

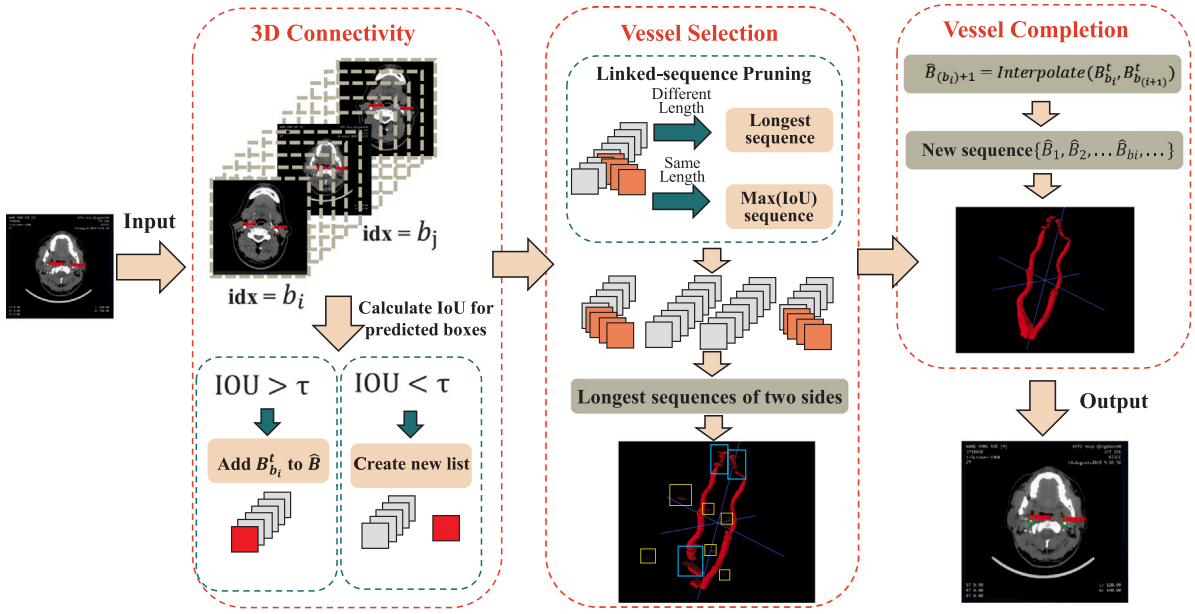

Fig. 4. Details of connectivity enhancement. We design a 2D–3D fusion method that constructs graph structures from bounding boxes, where edges represent the IOU between adjacent slices. By applying topological sorting and setting a threshold τ for edge connections, we identify the longest path in the graph, which corresponds to the carotid artery's physical characteristic of being the longest vessel. This technique ensures accurate detection and interpolation of missing vessels, resulting in a complete and connected vascular structure.

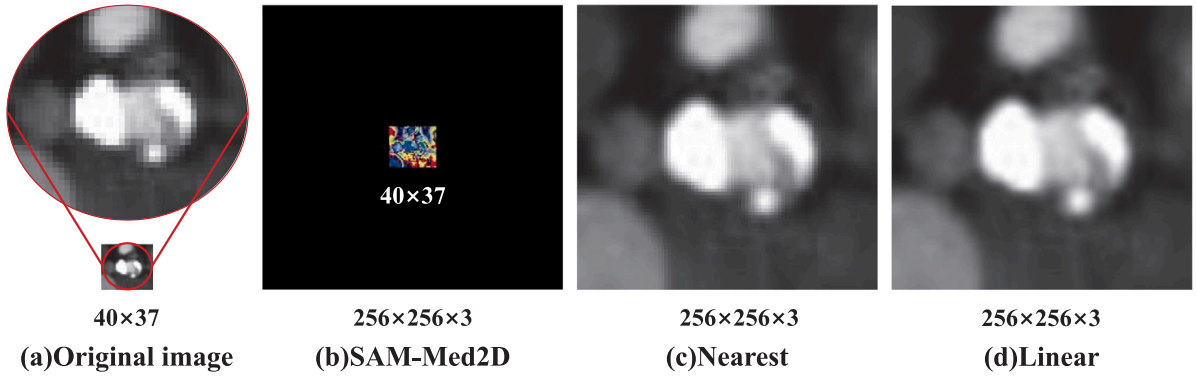

Fig. 5. Visualization of different augmentation results. Figure (b) shows the visualization result using SAM-Med2D's default image augmentation method. As can be observed, our images differ significantly in size from the default dimensions, requiring resize and interpolation. Figures (c) and (d) display the augmented images obtained using different interpolation methods. Additionally, the image normalization approach in Figure (b) is not suitable for our samples; therefore, we removed the normalization step and simply converted the grayscale images into three-channel RGB images for input.

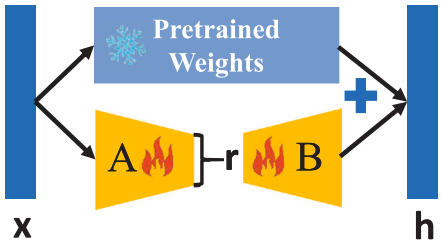

Fig. 6. Low rank adaptation (LoRA) structure.

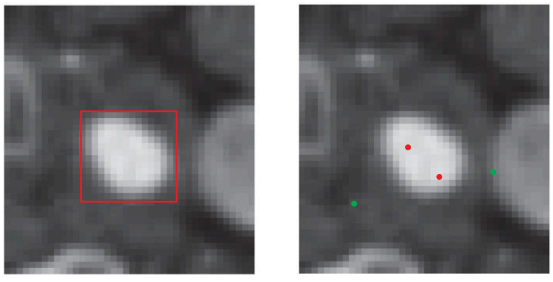

Fig. 7. Bbox prompt and point prompt examples.

we just copy the grayscale channel to RGB channels to adapt the input of SAM-Med2D but remove the normalization. For improve the image resolution, we use interpolation to increase the resolution of the cropped images to 256 × 256, which clarifies the edges of the structures within the vessels, thereby improving segmentation accuracy.

3.3.2. *Parameter-efficient fine-tuning*

Considering that SAM has a large number of parameters and needs high training costs, we choose to introduce a parameter-efficient fine-tuning strategy. Based on the experiments and theory in Aghajanyan

et al. (2020), it shows that pre-trained models exhibit a low-rank adaptation for a specific task. This means that the model parameters can be projected onto a smaller subspace, allowing the model to adapt to the task through learning. Based on this idea, we apply low-rank adaptation (LoRA (Hu et al., 2021)) to adjust the original model and achieve efficient parameter fine-tuning for large models. In LoRA, the weight matrices W of the attention heads in the Transformer (Vaswani et al., 2017) are replaced with low-rank matrices A and B. This can be expressed by the following equation:

$$h = W_0 x + \Delta W x = W_0 x + BAx \tag{3}$$

In the equation provided: x represents the input, h represents the output, W_0 represents the weights of the pre-trained model. If the weight matrix before transformation, denoted as $W \in \mathbb{R}^{d_{in} \times d_{out}}$, then $W_0 \in \mathbb{R}^{d_{in} \times d_{out}}$. Additionally, $B \in \mathbb{R}^{d_{in} \times r}$ and $A \in \mathbb{R}^{r \times d_{out}}$, where $r \ll \min(d_{in}, d_{out})$.

Specifically, as shown in Fig. 6, we replace the weight factors W of attention heads of Transformer Blocks in the Image Encoder with low-rank adapter layers B and A. Each adapter layer is a Multi-Layer Perception (MLP). The specific process can be formulated as follows:

$$H'^{(l)}_X = H^{(l-1)}_X + Adapter(MSA(LN(H^{(l-1)}_X))) \tag{4}$$

$$H^{(l)}_X = H'^{(l)}_X + MLP(LN(H'^{(l)}_X)) \tag{5}$$

$$+ \; r \cdot Adapter(LN(H'^{(l)}_X)) \tag{6}$$

where $H^{(l-1)}_X$ and $H^{(l)}_X$ are the output of the $(l-1)^{th}$ and lth Transformer Block, r is a scaling factor that regulates the influence of the adapter's output weight. MSA stands for Multi-Scale Attention, and LN refers to Layer Normalization. We freeze the other parameters of the original model, aiming to preserve the advantages of the large-scale model in image feature extraction. Only the parameters in the low-rank matrices, i.e. $Adapter$ in equations, are updated during fine-tuning. This approach reduces the parameter update volume to 4.9% of the original model, achieving efficient parameter fine-tuning while leveraging the benefits of the large-scale model.

In addition, we also optimize the prompt input for the model, compared to MedSAM (Ma et al., 2024), which only supports Bbox prompt input. Our model, in the prompt encoder part, supports two types of prompt information input: Bbox prompt and point prompt. For Bbox prompt information, we extract the position information of the top-left and bottom-right corners of the square box formed by selecting the outermost region based on the mask. This information is used as the prompt. For point prompt, we select foreground points within the detected object region and background points outside the detected object region within the image pixels. We assign them labels of 1 and 0, respectively, as shown in Fig. 7. The coordinates of these points along with their corresponding labels are used as the prompt. This enhancement allows our model to handle both Bbox prompt and point prompt information, providing more flexibility in capturing relevant cues for the task.

In our approach, we initialize the SAM-Med2D model with pre-trained model parameters. We then utilize LoRA to modify the model, allowing it to learn only the replaced parameters, i.e., the parameters B and A in Eq. (3). We fine-tune this modified model on our target dataset. Regarding the acquisition of prompt information, during training, we randomly select either point or Bbox prompts with a probability of 1:1 and input them into the prompt encoder for guidance. During inference, we generate prompt Bboxes or points based on the actual masks, according to the testing requirements. To simulate the uncertainty in the process of doctors annotating prompt information, we introduce random perturbation factors when generating prompt information.

4. Experiment

4.1. Datasets

Due to the absence of publicly available datasets for carotid artery lesion segmentation, this study used a custom dataset. Over 400 cases of angiographic CTA scans from patients with carotid artery stenosis were collected from five affiliated hospitals of Qingdao University. Our team, led by neurosurgeons and medical postgraduates, spent six months annotating 150 cases. The patient ages ranged from 24 to 88 years, with most cases aged 55 to 80. To address variations in imaging quality caused by differences in hospital campuses and equipment, we standardized the DICOM images by adjusting grayscale and brightness (window level: 100, window width: 500) before exporting them as JPG files. Using the Labelme tool, we annotated the carotid arteries from the thoracic cavity to the cranial region, covering the common and internal carotid arteries. Annotations included four categories: vascular wall(Wall), vascular lumen(Lumen), calcified plaques(Cal), and non-calcified(Non-Cal) plaques. The dataset comprised 64,469 CTA images, but only 16,362 images with labeled lesions at stenotic sites were used for training. These were divided into a training set (120 cases, 13,216 images) and a test set (30 cases, 3146 images) in a 4:1 ratio. The original image resolution was 256 × 256 pixels, later resized to 512 × 512 pixels for algorithm input. This dataset was used for all methods and experiments employing large image training, enabling fair comparison and validation.

Our team also completed the annotation of 20 cases of intact carotid arteries (including healthy blood vessels). Using the annotated vessel wall (wall) labels, we trained a YOLOv7-based coarse vessel segmentation algorithm (ICA-YOLOv7) to automatically detect and predict the vessel walls of bilateral carotid arteries. Based on the predicted vessel bounding boxes, the images are cropped with an outward extension of 10 pixels, providing input for the SAM model to perform fine lesion segmentation. After cropping, 21,135 vascular images with lesion labels were obtained. These were divided into a training set (120 cases, 16,877 images) and a test set (30 cases, 4258 images) at an 8:2 ratio. The cropped images ranged from 30 × 30 to 50 × 50 pixels in size, displaying the characteristics of low resolution and mosaic effects. During preprocessing, a linear algorithm was applied to enhance and enlarge the images. Finally, using three label categories: vessel lumen (lumen), calcified plaque (cal), and non-calcified plaque (non-cal), we trained the ICA-SAM fine-grained lesion segmentation algorithm, with the input image size being 256 × 256 pixels.

4.2. Implementation details & evaluation metrics

Our algorithm is implemented under the PyTorch framework and trained on the NVIDIA GeForce RTX 4080 GPU. It can be configured for use on desktop computers and laptops with GPUs, or configured on servers for use by networked computers in hospitals via the web terminal. Among them, for the algorithms trained with uncropped large CTA images, such as ICA-YOLOv7, U-Net, and FCN, the input image size is 512 × 512. Training is conducted for 400 epochs with a batch size of 8, reduced to 4 during the unfreezing stage. The initial learning rate is set to 1×10^{-3}, and then decayed using a cosine annealing strategy to 1×10^{-5}, with a weight decay rate of 1×10^{-4}. For the algorithms trained with the cropped small vascular images, including the main ICA-SAM algorithm and the ablation experiments, the original images have a size ranging from 30 × 30 to 50 × 50 pixels with low resolution and mosaic phenomena. After undergoing high-definition processing through linear image interpolation, they are enlarged to 256 × 256 for training. The model selects ViT-B (base resolution) (Dong et al., 2022), which is a medium-sized model with a base resolution among the variants of the Vision Transformer model. During the training stage, the Batch size was set to 8, the dimension of the low-rank matrix was set to $r = 16$, and

Table 1
Comparison with Other Methods. The first four methods are fully fine-tuning methods, and others are SAM and its improvement methods. The red letters indicate the best results among different methods for the metric and the blue letters represent the second-best results. The upward arrows signify that higher values are better, while downward arrows indicate the opposite.

Models	Years	Params (M)↓	Dice (Total)↑	IoU (Total)↑	Dice (Non-Cal) ↑	Dice (Cal)↑	Dice (Lumen) ↑
FCN (Long et al., 2015)	2015	35.3	0.6835	0.5192	0.5858	0.7413	0.7322
U-Net (Ronneberger et al., 2015)	2015	24.89	0.7292	0.5738	0.6207	0.7879	0.7421
UNet++ (Zhou et al., 2019)	2019	8.74	0.6347	0.4646	0.5051	0.7605	0.6760
SAM (Kirillov et al., 2023)	2023	87.12	0.4862	0.3212	0.2508	0.5014	0.6885
SAMed (Zhang and Liu, 2023)	2023	4.02	0.4158	0.2624	0.3329	0.3740	0.5193
SAM-Med2D (Cheng et al., 2023)	2023	175.95	0.4822	0.3177	0.2424	0.4314	0.7341
ICA-SAMv7(Ours)	2024	**3.93**	**0.8369**	**0.7386**	**0.7773**	**0.8178**	**0.9028**

Table 2
Impact of different prompts.

Prompt	Parameters		Dice (Total)	IoU (Total)	Dice (Non-Cal)	Dice (Cal)	Dice (Lumen)
Bboxes_Prompt	Default		0.5903	0.4854	0.4858	0.5435	0.7143
	Points_Num	Iter_Point	Dice (Total)	IoU (Total)	Dice (Non-cal)	Dice (Cal)	Dice (Lumen)
Points_Prompt	1	1	0.6156	0.4985	0.5055	0.5755	0.7407
	1	2	0.6429	0.5291	0.5401	0.5895	0.7695
	1	3	0.6624	0.5492	0.5598	0.6130	0.7863
	1	10	0.7326	0.6208	0.6363	0.7064	0.8367
	1	20	0.7562	0.6467	0.6631	0.7377	0.8525
	1	50	0.7800	0.6743	0.6930	0.7641	0.8691
	3	1	0.5872	0.4714	0.4911	0.5395	0.7043
	5	1	0.6402	0.5197	0.5504	0.5717	0.7640
	9	1	0.6644	0.5431	0.5717	0.5905	0.7945

the number of prompting points during training was set to 1 by default, with a perturbation factor set to 0.1.

The evaluation metrics for the ICA-YOLOv7 algorithm use AP value. The evaluation metrics for segmentation algorithms include the Dice coefficient, and Intersection over Union (IoU). Higher evaluation values indicate better results.

The calculation methods for these evaluation metrics are as follows:

$$Dice = \frac{2TP}{2TP + FP + FN}, \tag{7}$$

$$IOU = \frac{TP}{TP + FP + FN}, \tag{8}$$

where TP (True Positives) are the samples predicted as positive correctly. TN (True Negatives) are the samples predicted as negative correctly. FP (False Positives) are the samples predicted as positive incorrectly. FN (False Negatives) are the samples predicted as negative incorrectly.

4.3. Comparison with SOTA models

We utilize 2D head CTA images as input to train several models for achieving precise segmentation of carotid artery lesions. The first 120 cases are used as training data, while the remaining 30 cases are used for testing. We compare our model with traditional image segmentation algorithms, as well as with SAM and its improvement models.

Traditional deep learning algorithms require learning all model parameters, resulting in high learning costs. However, our method only requires learning a subset of the model parameters. Experimental results revealed that traditional image segmentation models perform well in fine-grained detection of small targets. However, the SAM model and its improved versions do not perform well in fine-grained detection of medical images with significant interference information. Our model surpasses traditional detection models with a notable improvement of 11%–20% in the Dice coefficient. Compared to SAM and its enhanced variants, our model achieves an even greater improvement of approximately 35%. Furthermore, in fine-grained detection tasks, our model demonstrates superior performance in accurately detecting both non-calcified plaque (Non-cal) and calcified plaque (Cal) regions. Table 1 illustrates the comparison.

Table 3
Impact of data augmentation.

Augment	Dice (Total)	IoU (Total)	Dice (Non-Cal)	Dice (Cal)	Dice (Lumen)
Normalization	0.5903	0.4854	0.4858	0.5435	0.7143
Nearest	0.8161	0.7119	0.7401	0.8042	0.8918
Linear	0.8369	0.7386	0.7773	0.8178	0.9028

In our experiments, we also utilize the MedSAM (Ma et al., 2024) method with our data. However, we find that due to the MedSAM model simultaneously outputting masks for three fine-grained types based on the bounding boxes (Bbox) and the arrangement of substances inside the vessels, which exhibit inclusion relationships, there is a high degree of overlap between the bounding boxes. This high overlap can be seen in Fig. 1(d), resulting in highly overlapped predicted masks. After fine-tuning the MedSAM model using the training data and testing it on the testing data, the experimental results show that even the highest accuracy achieved for the fine-grained substance "Lumen" recognition is only 27%. The accuracy for other substances recognition is below 5%. These results demonstrate that the approach of using Bbox prompts and multi-label inputs in MedSAM is not suitable for the fine-grained detection task of carotid artery vessels, as it exhibits very low recognition accuracy.

In Fig. 8, we illustrate the image segmentation results using different models. It can be observed that traditional detection algorithms such as FCN and U-Net often suffer from false negatives. In Fig. 8, the FCN algorithm missed the left carotid artery lesion and the U-Net algorithm missed the right one, which is false negative. The SAM model, due to the ambiguity of the prompt information, tends to produce overlapping detection results, as seen in the "Sample 2", leading to lower accuracy in fine-grained detection. In comparison to FCN, Unet, and SAM methods, our model provides more accurate fine-grained detection with clearer boundaries.

4.4. Ablation studies

To demonstrate the effectiveness of the technique, we conducted ablation experiments in five aspects: the impact of connectivity enhancement, the impact of YOLO cropping, the impact of different

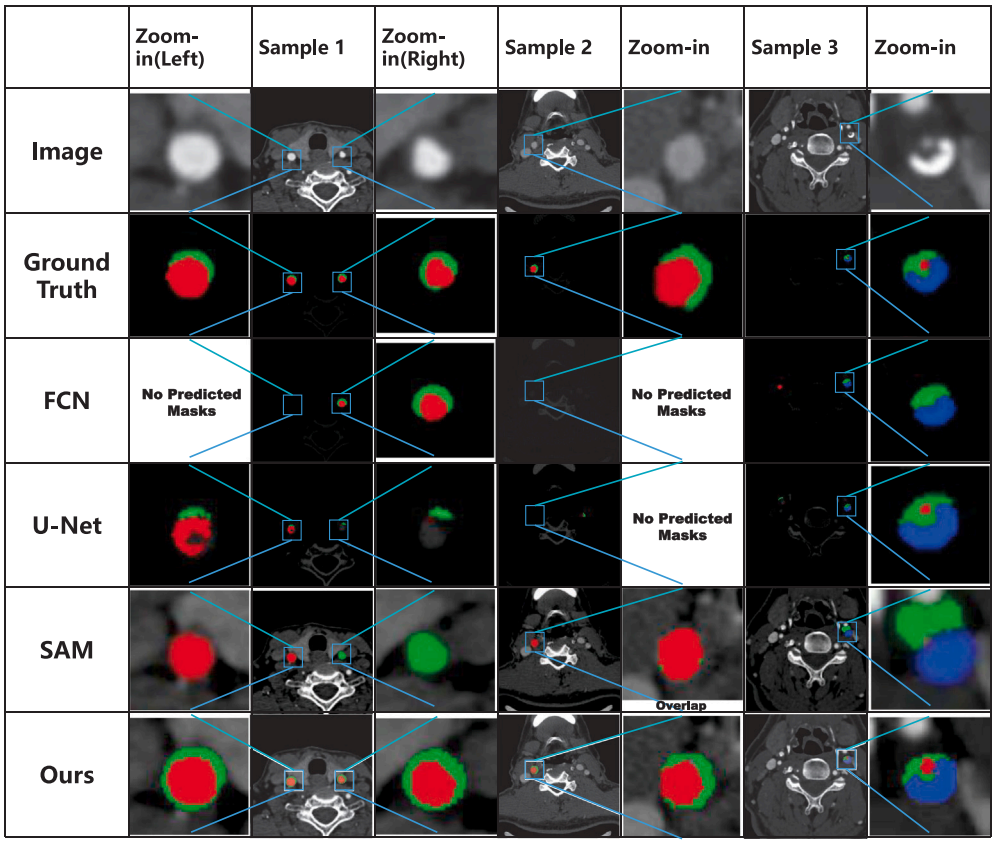

Fig. 8. Comparison of different SOTA methods. Horizontally, there are three test images. Vertically, there are the lesion segmentation result images of various algorithms. The blue area represents Cal, the red area represents Lumen, and the green area represents Non-cal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Impact of connectivity enhancement.

	ICA-YOLOv7	YOLOv7
Number of CTA Slicers	64 469	
Amount of Correct Annotations	62 194	53 352
Dice Per Detection	96.47%	82.76%
AP of Vessel Detecting Boxes	98.24%	91.38%

prompts, the impact of data augmentation, and the impact of LoRA fine-tuning.

4.4.1. Impact of connectivity enhancement

To investigate the impact of the connectivity enhancement step on the experimental results, we compare the results of using ICA-YOLOv7 with the results obtained by only using YOLOv7 for vessel segmentation, followed by prediction on unlabeled CTA images. The comparison results are shown in Table 4, where it can be observed that the 3D connectivity enhancement method enhances the accuracy of Dice Per Detection by 14%. In addition, from Fig. 3(c) and (d), it can be seen that before the 3D connectivity enhancement, the vessels detected by YOLO contained some false positives and missed detections, resulting in disconnected or noisy vessels. After the 3D connectivity enhancement, we connect missed carotid artery vessels and eliminate incorrectly detected branches, improving the coherence and accuracy of the vessels.

4.4.2. Impact of YOLO cropping

To evaluate the effectiveness of YOLO cropping, we compared the results of directly detecting using head CTA images with those of performing fine-grained detection on images segmented by YOLO. We use the SAM-Med2D (Cheng et al., 2023) as the base model and train and test it using both the original un-cropped 512 × 512 head CTA image data (referred to as "All") and the segmented data obtained after YOLO cropping and connectivity enhancement (referred to as "Seg"). The results are shown in Table 5. From the table, it can be observed that using the YOLO segmented and connectivity enhancement dataset for detection improves the accuracy by approximately 26% compared to using the original un-cropped head CTA image data. This demonstrates the effectiveness of utilizing YOLO segmentation and connectivity enhancement for fine-grained recognition, allowing for more accurate detection of specific regions of interest within the head CTA images.

4.4.3. Impact of different prompts

To investigate the impact of different prompt types on detection accuracy, we explored the effects of different numbers of points and different iterations for point prompts. We compare these results with the performance obtained using Bbox prompts. The results are shown in Table 2. In the table, "Bboxes_prompt" represents the Bbox prompt, and "Points_prompt" represents the point prompt. We vary the number of points selected("points_num") and the number of iterations for point prompts("Iter_point"). For point prompts, the first iteration randomly selects the points, and subsequent iterations select points from the regions where the model made incorrect predictions. The model used in this comparison is based on SAM-Med2D with the parameters replaced by LoRA, and the input data are obtained after ICA-YOLOv7 cropping without data augmentation. It can be observed that point prompts achieve higher accuracy compared to Bbox prompts. Furthermore, as the number of points and iterations increases, the accuracy further improves. We believe that using 9 points with 1 iteration and 1 point with 50 iterations already simulates the maximum number of external prompts from assistance in fine-grained detection.

Table 5
Impact of YOLO cropping.

Dataset	Dice (Total)	IoU (Total)	Dice (Non-Cal)	Dice (Cal)	Dice (Lumen)
All	0.3513	0.2131	0.2016	0.3353	0.4989
Seg	0.6197	0.5157	0.5107	0.6239	0.7167

Table 6
Impact of LoRA fine-tuning.

Model	Dice (Total)	IoU (Total)	Dice (Non-Cal)	Dice (Cal)	Dice (Lumen)
SAM-Med2D	0.7437	0.6345	0.6032	0.7755	0.8269
ICA-SAMv7	0.8369	0.7386	0.7773	0.8178	0.9028

Table 7
Five-fold cross-validation results.

Augmentation	Linear		Nearest	
Metrics	IoU	Dice	IoU	Dice
Fold0	0.7115	0.8161	0.7136	0.8182
Fold1	0.7509	0.8464	0.7017	0.8086
Fold2	0.7488	0.8449	0.7174	0.8214
Fold3	0.7268	0.8286	0.7024	0.8087
Fold4	0.7550	0.8485	0.7244	0.8255
Average	0.7386	0.8369	0.7119	0.8161

Table 8
Sensitivity and specificity of stenosis rate prediction.

Stenosis rate	Sensitivity	Specificity
<30%	0.6667	0.9444
30%–70%	1	0.7368
>70%	0.6875	1
Avg.	0.7847	0.8938

4.4.4. Impact of data augmentation

Compared to traditional segmentation algorithms such as U-Net and FCN, we focused on the prompt-based learning method of the SAM large model to first locate vessels in large CTA images and then perform fine lesion segmentation. Since vascular structures are very small, averaging only about 40 × 40 pixels, we explored the original image processing method (Normalization) used in SAM-Med2D. The Normalization operation involved filling all edge pixels of the input image with a value of 1 to create 256 × 256 normalized training data. However, the training results were poor, achieving a lesion segmentation accuracy (Dice coefficient) of only 59.03%. To address this, we compared different image processing methods: nearest-neighbor interpolation (Nearest) and bilinear interpolation (Linear). The experimental results showed that when training with 256 × 256-sized images, both Linear and Nearest methods achieved similar performance, improving segmentation accuracy by 20% compared to the original Normalization method, yielding highly effective results. The experimental outcomes are shown in Table 3. It is worth noting that since the prompts were randomly generated, accuracy fluctuations within 1% are considered normal. Visualization can be seen at Fig. 9. Evidently, compared with the original (c) Normalization processing method of SAM-Med2D, the (e) Linear processing method has witnessed a significant improvement and achieved the optimal lesion segmentation effect.

In order to achieve the goal of assessing the model's generalization ability and reducing overfitting, we conducted five-fold cross-validation using both Linear processing and Nearest methods. As shown in Table 7.

4.4.5. Impact of LoRA fine-tuning

To investigate the impact of parameter-efficient fine-tuning improvements on the experimental results, we train and test models using the data obtained after YOLO segmentation and connectivity enhancement. The results are shown in Table 6. In the table, "SAM-Med2D" represents the results obtained using the SAM-Med2D model (Cheng et al., 2023) for training and prediction, while "ICA-SAMv7" represents the results obtained using our method for training and prediction. Both methods used the Linear method for data augmentation. It can be observed that our method, ICA-SAMv7, achieves an 8% improvement in accuracy compared to the SAM-Med2D method. Furthermore, our method demonstrates improvements across all categories in fine-grained detection. This highlights the effectiveness of our parameter-efficient fine-tuning improvements in enhancing the accuracy of the detection model. The comparison of visualization effects is shown in Fig. 10. Our method (d) is significantly superior to the effect of SAM-Med2D (c) in terms of visualization and recognition effect, and is close to the effect of the ground truth masks(b).

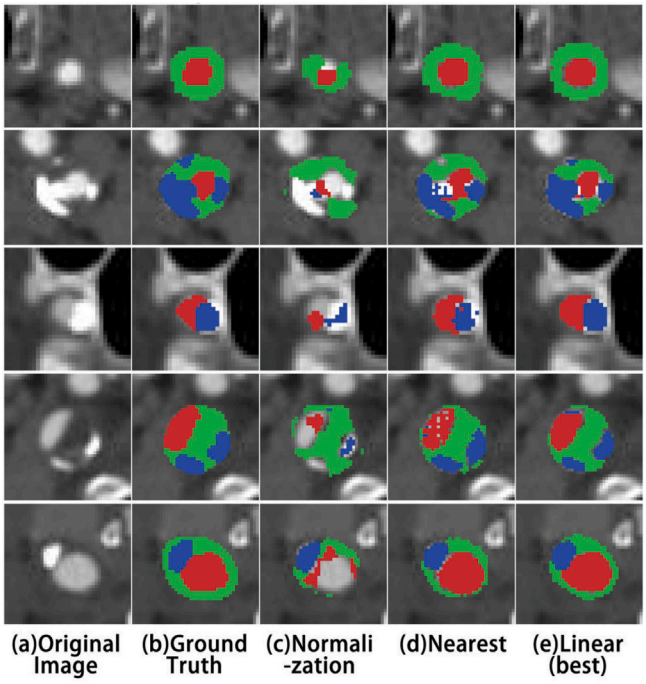

Fig. 9. Image processing renderings. The blue area represents Cal, the red area represents Lumen, and the green area represents Non-cal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.5. User study

To validate the clinical applicability of our method, we designed an experiment to compare the results of the ICA-SAMv7 algorithm with physicians' diagnostic reports. Physicians' CTA diagnostic reports typically only indicate the degree of carotid artery stenosis. Following the methodology for evaluating carotid artery stenosis outlined in Jensen (1991) and effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis (1991), stenosis rates were categorized into three levels: mild stenosis (<30%), moderate stenosis (30%–70%), and severe stenosis (>70%). In this experiment, the ICA-SAMv7 algorithm was applied to perform fine segmentation of carotid artery lesions in 20 new cases, including calcified (Cal), non-calcified (Non-Cal), and lumen components. The vascular stenosis rate was then calculated as the ratio of the lesion area to the vascular cross-sectional area. These results were classified into three levels (<30%, 30%–70%, >70%) and compared with the stenosis severity (mild, moderate, severe) recorded in physicians' diagnostic reports. As shown in Table 8 and Fig. 11, the model achieved a sensitivity of 0.78 and a specificity of 0.89, meeting the basic requirements for assisting clinical diagnosis. Notably, errors were primarily concentrated in samples with stenosis rates of

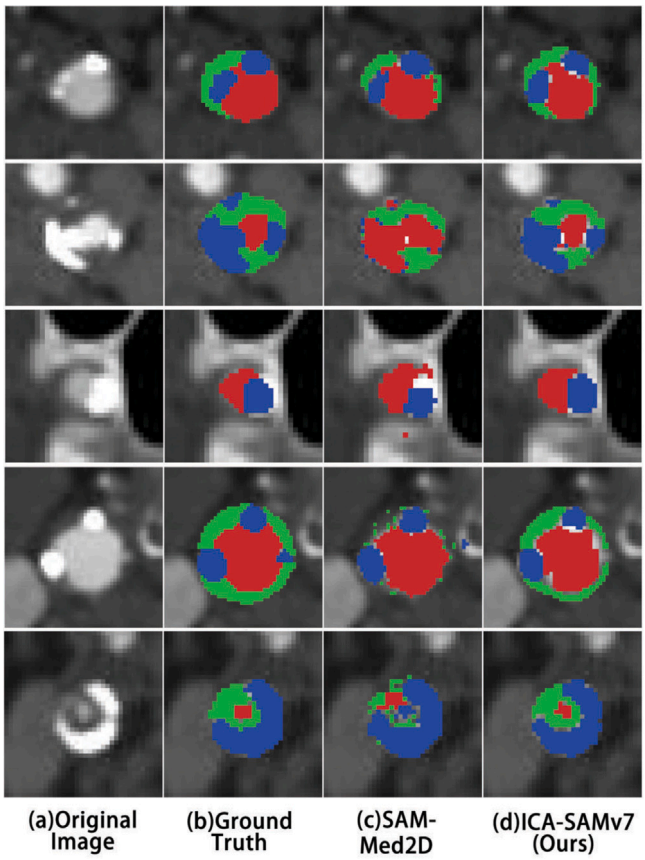

Fig. 10. Visualization of PEFT impact. The blue area represents Cal, the red area represents Lumen, and the green area represents Non-cal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

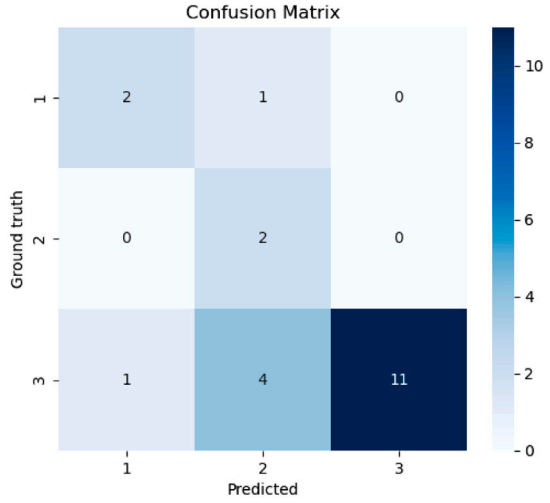

Fig. 11. Confusion matrix for the prediction of sample stenosis degrees. Here, label 1, 2, and 3 correspond to <30%, 30%–70%, and >70% stenosis rate respectively.

30%–70%, as the evaluation of these intermediate cases is strongly influenced by physicians' subjective judgment.

5. Conclusion

In this paper, we introduce a novel method, ICA-SAMv7, for segmenting carotid artery vascular lesions. This two-stage algorithm uses the YOLOv7 model with the 3D Connectivity Enhancement Model to initially locate and roughly segment carotid arteries. Subsequently, it employs parameter-efficient fine-tuning with SAM to precisely segment vascular lesion tissues. The lesion segmentation accuracy under the SAM model rises from 48.62% to 83.69%. A profusion of comparative experiments demonstrate the algorithm's excellent performance.

The proposed ICA-SAMv7 achieves an 11%–20% improvement in the Dice coefficient versus traditional segmentation algorithms and a 35% improvement over the original SAM model, as shown in Table 1. Qualitatively, Fig. 8 shows that ICA-SAMv7 provides more precise lesion segmentation than both traditional methods and the original SAM model. Ablation studies and a user study confirmed the accuracy and effectiveness of our data processing and model improvements. In clinical practice, ICA-SAMv7 achieved precise segmentation of carotid artery lesions and accurate assessment of the stenosis rate, validating its applicability and the potential for broader clinical applications.

CRediT authorship contribution statement

Xiaotian Yan: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Yuting Guo:** Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis. **Ziyi Pei:** Writing – original draft, Visualization, Validation, Methodology, Investigation. **Xinyu Zhang:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Jinghao Li:** Supervision, Software, Project administration, Methodology, Data curation, Conceptualization. **Zitao Zhou:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis. **Lifang Liang:** Visualization, Validation, Methodology, Data curation, Conceptualization. **Shuai Li:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peng Lun:** Writing – original draft, Validation, Supervision, Data curation, Conceptualization. **Aimin Hao:** Writing – review & editing, Visualization, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the National Key Research and Development Program of China NO. 2023YFF1203803, the Demonstration Project in Shandong Province No. 2022SFGC0601, the National Natural Science Foundation of China No. 62272021 and Beijing Science and Technology Plan Project Z231100005923039.

Data availability

Data will be made available on request.

References

Aghajanyan, Armen, Zettlemoyer, Luke, Gupta, Sonal, 2020. Intrinsic dimensionality explains the effectiveness of language model fine-tuning. ArXiv abs/2012.13255.

Alom, Md Zahangir, Hasan, Mahmudul, Yakopcic, Chris, Taha, Tarek M, Asari, Vijayan K, 2018. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.

Arasu, Rohan, Arasu, Alexis, Muller, Juanita, 2021. Carotid artery stenosis: An approach to its diagnosis and management. Aust. J. Gen. Pr. 50 (11), 821–825.

Azad, Reza, Asadi-Aghbolaghi, Maryam, Fathy, Mahmood, Escalera, Sergio, 2019. Bi-directional convlstm U-net with densley connected convolutions. In: 2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019. IEEE, pp. 406–415. http://dx.doi.org/10.1109/ICCVW.2019.00052.

Chaudhry, Asmatullah, Hassan, Mehdi, Khan, Asifullah, Kim, Jin Young, 2013. Automatic Active Contour-Based Segmentation and Classification of Carotid Artery Ultrasound Images. 26 1071–1081.

Cheng, S.F., Brown, M.M., Simister, R.J., Richards, T., 2019. Contemporary prevalence of carotid stenosis in patients presenting with ischaemic stroke. 106.0. pp. 872–878.

Cheng, Junlong, Ye, Jin, Deng, Zhongying, Chen, Jianpin, Li, Tianbin, Wang, Haoyu, Su, Yanzhou, Huang, Ziyan, Chen, Jilong, Jiang, Lei, et al., 2023. Sam-med2d. arXiv preprint arXiv:2308.16184.

Dhamija, Tashvik, Gupta, Anunay, Gupta, Shreyansh, Anjum, Katarya, Rahul, Singh, Ghanshyam, 2022. Semantic segmentation in medical images through transfused convolution and transformer networks. pp. 1132–1148, 53.

Dolz, Jose, Gopinath, Karthik, Yuan, Jing, Lombaert, Herve, Desrosiers, Christian, Ayed, Ismail Ben, 2018. HyperDense-net: A hyper-densely connected CNN for multi-modal image segmentation. IEEE Trans. Med. Imaging 38 (5), 1116–1126.

Dong, Xiaoyi, Bao, Jianmin, Zhang, Ting, Chen, Dongdong, Gu, Shuyang, Zhang, Weiming, Yuan, Lu, Chen, Dong, Wen, Fang, Yu, Nenghai, 2022. CLIP itself is a strong fine-tuner: Achieving 85.7% and 88.0% top-1 accuracy with vit-b and vit-l on ImageNet. http://dx.doi.org/10.48550/ARXIV.2212.06138, CoRR abs/2212.06138.

Fu, Fan, Shan, Yi, Yang, Guang, Zheng, Chao, Zhang, Miao, Rong, Dongdong, Wang, Ximing, Lu, Jie, 2023. Deep Learning for Head and Neck CT Angiography: Stenosis and Plaque Classification. 307 220996.

Gibson, Eli, Giganti, Francesco, Hu, Yipeng, Bonmati, Ester, Bandula, Steve, Gurusamy, Kurinchi, Davidson, Brian, Pereira, Stephen P, Clarkson, Matthew J, Barratt, Dean C, 2018. Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans. Med. Imaging 37 (8), 1822–1834.

Houlsby, Neil, Giurgiu, Andrei, Jastrzebski, Stanislaw, Morrone, Bruna, De Laroussilhe, Quentin, Gesmundo, Andrea, Attariyan, Mona, Gelly, Sylvain, 2019. Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning. PMLR, pp. 2790–2799.

Hu, Edward J, Shen, Yelong, Wallis, Phillip, Allen-Zhu, Zeyuan, Li, Yuanzhi, Wang, Shean, Wang, Lu, Chen, Weizhu, 2021. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.

Ismail, Anas, Ravipati, Srinivasarao, Gonzalez-Hernandez, Diana, Mahmood, Humera, Imran, Abira, Muñoz, E. Cardona, Naeem, S., Abdin, Zain U, Siddiqui, Humza Faisal, 2023. Carotid artery stenosis: A look into the diagnostic and management strategies, and related complications.

Jensen, Troels Staehelin, 1991. MRC European carotid surgery trial: interim results for symptomatic patients with severe (70-99%) or with mild (0-29%) carotid stenosis. European carotid surgery trialists' collaborative group. Lancet 337 (8752), 1235–1243.

Jiale, Xu, Qinghui, Zhang, Junqiu, Li, Jian, Rong, Zhenping, Qiang, 2024. A method combining Kalman's algorithm and the YOLOv7 for estimating the motion of carotid plaques in B-mode ultrasound image sequences. In: 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering. ICSECE, IEEE, pp. 13–18.

Kirillov, Alexander, Mintun, Eric, Ravi, Nikhila, Mao, Hanzi, Rolland, Chloe, Gustafson, Laura, Xiao, Tete, Whitehead, Spencer, Berg, Alexander C, Lo, Wan-Yen, et al., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026.

Li, Shengze, Cao, Jianjian, Ye, Peng, Ding, Yuhan, Tu, Chongjun, Chen, Tao, 2024. Clipsam: CLIP and SAM collaboration for zero-shot anomaly segmentation. arXiv preprint arXiv:2401.12665.

Lialin, Vladislav, Deshpande, Vijeta, Rumshisky, Anna, 2023. Scaling down to scale up: A guide to parameter-efficient fine-tuning. arXiv preprint arXiv:2303.15647.

Liang, Ming, Hu, Xiaolin, 2015. Recurrent convolutional neural network for object recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3367–3375.

Liu, Haokun, Tam, Derek, Muqeeth, Mohammed, Mohta, Jay, Huang, Tenghao, Bansal, Mohit, Raffel, Colin A, 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. Adv. Neural Inf. Process. Syst. 35, 1950–1965.

Long, Jonathan, Shelhamer, Evan, Darrell, Trevor, 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Ma, Jun, He, Yuting, Li, Feifei, Han, Lin, You, Chenyu, Wang, Bo, 2024. Segment anything in medical images. Nat. Commun. 15 (1), 654.

Mohd, Ahmed B, Alabdallat, Yasmeen, Mohd, Omar B, Ghannam, Reem A, Sawaqed, Seri, Hasan, Hanan, Ellebedy, Mohamed, Turkmani, Khaled, Al-Ezzi, Shakir, 2023. Medical and surgical management of symptomatic and asymptomatic carotid artery stenosis: A comprehensive literature review. p. 15.

Perslev, Mathias, Dam, Erik Bjørnager, Pai, Akshay, Igel, Christian, 2019. One network to segment them all: A general, lightweight system for accurate 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. Springer, pp. 30–38.

Redmon, J., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Ronneberger, Olaf, Fischer, Philipp, Brox, Thomas, 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Shoeybi, Mohammad, Patwary, Mostofa, Puri, Raul, LeGresley, Patrick, Casper, Jared, Catanzaro, Bryan, 2019. Megatron-lm: Training multi-billion parameter language models using model parallelism. arXiv preprint arXiv:1909.08053.

effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis, Beneficial, 1991. North American symptomatic carotid endarterectomy trial collaborators. N. Engl. J. Med. 325 (7), 445–453.

Sung, Yi-Lin, Nair, Varun, Raffel, Colin A., 2021. Training neural networks with fixed sparse masks. Adv. Neural Inf. Process. Syst. 34, 24193–24205.

Tetteh, Giles, Efremov, Velizar, Forkert, Nils D., Schneider, Matthias, Kirschke, Jan, Weber, Bruno, Zimmer, Claus, Piraud, Marie, Menze, Bjoern H., 2018. DeepVessel-Net: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes. CoRR abs/1803.09340.

Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N, Kaiser, Łukasz, Polosukhin, Illia, 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Wang, Chien-Yao, Bochkovskiy, Alexey, Liao, Hong-Yuan Mark, 2023a. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7464–7475.

Wang, Tao, Su, Xiaojun, Liang, Yuchao, Luo, Xu, Hu, Xiao, Xia, Ting, Ma, Xuebin, Zuo, Yongchun, Xia, Huilin, Yang, Lei, 2024. Integrated deep learning model for automatic detection and classification of stenosis in coronary angiography. Comput. Biol. Chem. 112, 108184.

Wang, Haoxiang, Vasu, Pavan Kumar Anasosalu, Faghri, Fartash, Vemulapalli, Raviteja, Farajtabar, Mehrdad, Mehta, Sachin, Rastegari, Mohammad, Tuzel, Oncel, Pouransari, Hadi, 2023b. Sam-clip: Merging vision foundation models towards semantic and spatial understanding. arXiv preprint arXiv:2310.15308.

Wu, Junde, Fu, Rao, Fang, Huihui, Liu, Yuanpei, Wang, Zhaowei, Xu, Yanwu, Jin, Yueming, Arbel, Tal, 2023. Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620.

Wu, Junke, Wei, Guoliang, Fan, Yi, Yu, Liang, Chen, Bo, 2024. B-ultrasound guided venipuncture vascular recognition system based on deep learning. Biomed. Signal Process. Control. 87, 105495.

Yang, Jinzhu, Huang, Mingxu, Fu, Jie, Lou, Chunhui, Feng, Chaolu, 2020. Frangi based multi-scale level sets for retinal vascular segmentation. Comput. Methods Programs Biomed. 197, 105752. http://dx.doi.org/10.1016/J.CMPB.2020.105752.

Zaken, Elad Ben, Ravfogel, Shauli, Goldberg, Yoav, 2021. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. arXiv preprint arXiv:2106.10199.

Zhang, Kaidong, Liu, Dong, 2023. Customized segment anything model for medical image segmentation. arXiv preprint arXiv:2304.13785.

Zhou, Zongwei, Siddiquee, Md Mahfuzur Rahman, Tajbakhsh, Nima, Liang, Jianming, 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39 (6), 1856–1867.