Electronic Research Archive

https://www.aimspress.com/journal/ERA

ERA, 33(8): 4495–4528.
DOI: 10.3934/era.2025204
Received: 23 May 2025
Revised: 01 July 2025
Accepted: 28 July 2025
Published: 06 August 2025

Research article

GS_NeXt: Graph theory combining segment anything model for liver and tumor segmentation from CT

Qing Wang[1], Jinke Wang[1,2,*], Liang Guo[1] and Min Xu[3]

[1] School of Automation, Harbin University of Science and Technology, Harbin 150080, China
[2] Weihai Research Institute, Harbin University of Science and Technology, Weihai 264300, China
[3] Weihai Municipal Hospital, Affiliated to Shandong University, Weihai 264299, China

* **Correspondence:** Email: jkwang@hitwh.edu.cn; Tel: +86-13863132787.

Abstract: The static convolutional network is designed with a restricted sense field, but it limits the global feature extraction. While dynamic convolution addresses the issue of limited static receptive fields, it struggles to perform well in discontinuous liver regions, liver tumor border regions, and microtumor segmentation. To alleviate the above issues, we proposed a network, GS_NeXt, based on graph theory and the Segment Anything Model (SAM) for liver and tumor segmentation from CT scans. First, we employed the feature extraction module of ConvNeXt-v2 to learn features across channels, enabling the network to focus on critical liver and tumor regions. Second, we utilized the SAM with frozen weights to extract more comprehensive global information, thereby enhancing feature representation. Third, we applied graph reasoning to globally model unstructured local features, improving the network's understanding of CT images in discontinuous liver regions, liver-tumor boundaries, and microtumor areas. Finally, we incorporated a deep supervision mechanism to facilitate the learning of multi-scale features throughout the network. We evaluated the proposed segmentation method for two publicly available abdominal liver tumor CT datasets. On the LiTS17 dataset, GS_NeXt achieved 97.74% and 87.25% on the Dice scores, 1.01 and 2.23 mm on the average symmetric surface distance (ASD), and 3.68% and 22.60% on the volume overlap errors (VOE) for liver and tumor segmentation, respectively. On the 3DIRCADb dataset, it achieved 97.31% and 87.36% on the Dice scores, ASD values of 1.01 and 2.12 mm, and VOE scores of 3.56% and 21.56% for liver and tumor segmentation, respectively.

Keywords: segmentation; deep learning; liver; graph inference; dynamic convolution

1. Introduction

The liver is one of the most vital organs in the human body, owing to its crucial roles in detoxification and digestion. According to the Global Cancer Statistics Report 2022, liver cancer ranked sixth in global incidence and third in mortality [1]. Accurate segmentation of liver and tumor from CT scans plays a vital role in diagnosis and treatment planning. Therefore, the diagnosis and treatment of liver cancer are crucial. Since CT can generate high-resolution slice images of the liver and surrounding abdominal organs in a non-invasive way, it has become the primary diagnostic method for liver cancer. Multi-dimensional CT technology has been widely used because of its high cost-performance ratio. However, the large number of CT slices puts a considerable burden on the radiologist's diagnosis, which is not only time-consuming but also susceptible to the surgeon's subjective experience, thus affecting the efficiency and accuracy of the diagnosis. Therefore, automatic and accurate liver and tumor segmentation is urgently needed. Besides, segment liver and tumor from CT images also face some challenges: (i) In abdominal CT images, the liver and its malignant tumors are not traceable in terms of their appearance and morphology, how many they are, and their volume size; (ii) the differences in CT values between diseased and non-diseased tissues, and between the liver and adjacent organs are not significant, resulting in an inconspicuous boundary in CT images (as shown in Figure 1, the red and green lines are the gold-standard boundaries of liver and tumor, respectively). These challenges motivate us to develop advanced deep learning methods capable of handling discontinuities, poorly defined boundaries, and small tumor regions.

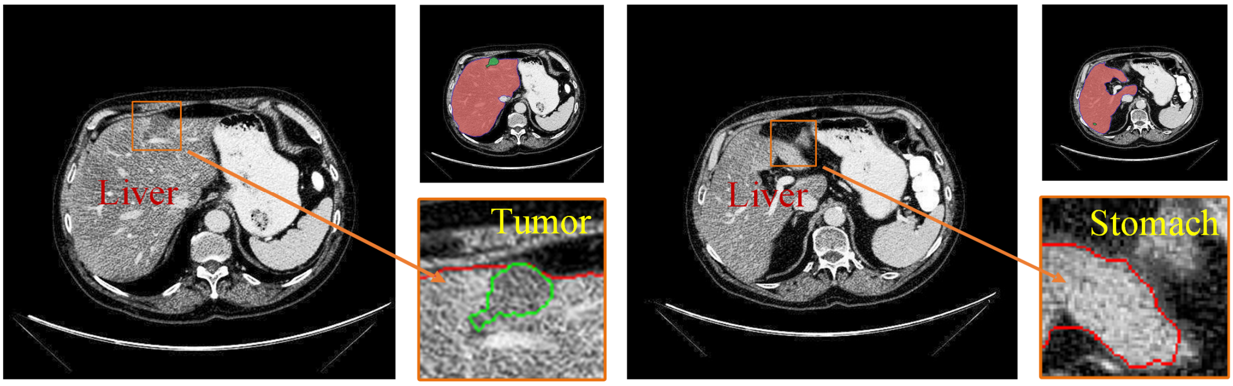


Figure 1. Challenges in liver and tumor segmentation. (a) Tumor in the liver boundary and (b) liver edge with adjacent organs.

Accompanied by the rapid development of computer technology, image segmentation techniques are rapidly evolving towards automation and high precision. Traditional machine learning methods, such as those based on level sets, edge detection, and region growing, can improve segmentation efficiency. However, they depend highly on manual feature design and complex parameter tuning, limiting their wide application.

Deep learning methods can automatically learn essential features from datasets, demonstrating significant potential for the development of medical image segmentation [2]. Long et al. [3] proposed a full convolutional network (FCN). Its main advantage is that the input and output images are the same size, and the input image can be of arbitrary resolution. Its outstanding feature extraction capability has attracted wide attention. Compared to natural image segmentation, the accuracy of medical image segmentation is typically affected by several factors, including the size, shape, and location of the lesion region, as well as the low contrast and blurring of boundaries between organs in CT images. Therefore, accurate segmentation is a challenging task. Researchers have proposed numerous advanced networks to mitigate the effects of the factors mentioned above. Furthermore, U-Net [4] is the most typical network whose encoder consists of convolutional and pooling layers. Still, this oversimplified encoder structure makes it easy to lose a lot of feature information during feature extraction. Thus, it is not suitable for high-precision segmentation requirements. Therefore, many researchers have improved the encoder. The encoder of the improved U-shaped network usually consists of the backbone network of other excellent classification networks, e.g., using the VGG [5] network as the encoder of the VGG-UNet [6], using ResNet [7] as the encoder of the ResUNet [8], and putting ConvNeXt [9] improved as the encoder's ConvUNeXt [10]. They replace or enhance the coding layer of the U-Net to improve the feature extraction capability of the network encoder. The networks above primarily improve segmentation performance by strengthening the encoder. However, in medical image segmentation, an efficient and effective decoding mechanism is equally crucial. Rahman et al. proposed EMCAD [11], an efficient multi-scale convolutional attention decoder for medical image segmentation. By combining Multi-Scale Convolutional Attention Modules (MSCAM) with Large-kernel Grouped Attention Gates (LGAG), EMCAD addresses the high computational cost of traditional decoders. This design enhances spatial feature representation while maintaining a lightweight architecture. These improvements to the U-shaped structural encoder have enhanced the network’s ability to represent image features, leading to improved performance in several medical image segmentation tasks. However, most of the above methods are based on the static convolution mechanism, which makes it difficult to perform effective global feature extraction, resulting in poor segmentation results. Therefore, it is necessary to introduce novel network structures capable of global semantic reasoning to address these challenges more comprehensively.

Nevertheless, the static convolutional network represented by FCN is designed with a restricted receptive field, which limits the global feature extraction ability and leads to unsatisfactory segmentation results. Full-dimensional dynamic convolution [12] is a dynamic convolution algorithm that enhances the performance of a convolutional neural network by introducing a learnable deformation module that dynamically adjusts the shape and size of the convolution kernel according to the input features during the convolution process. Mamba is a proposed selective state-space model designed for efficient long-range sequence modeling with linear time complexity. Xing et al. proposed SegMamba [13], a 3D segmentation network that utilizes the Mamba state-space model to model long-range dependencies efficiently. It incorporates tri-directional feature modeling and spatial attention to enhance global and local representation. However, the sheer volume of data in 3D CT images poses significant challenges for image processing and analysis. The boundary segmentation between the liver and tumors has long been a substantial challenge in liver and tumor segmentation. Due to the blurred boundaries and the influence of noise, over-segmentation and under-segmentation often occur. Liu et al. [14] proposed a lightweight and efficient network for polyp segmentation, TSCA-Net, aiming to achieve high segmentation accuracy with extremely low computational cost. To address challenges such as the diverse morphology of polyps, blurred boundaries, and real-time deployment requirements, the authors designed a novel Two-Stream Cross-scale Attention (TSCA) module, which captures texture details in shallow layers and semantic information in deeper layers, and effectively fuses multi-scale features through a cross-guided attention mechanism. A Feature Aggregation Module (FAM) is introduced to integrate multi-level features and enhance the representation of salient regions. Furthermore, the extensive use of depth-wise separable convolutions ensures computational efficiency. The design of TSCA-Net provides valuable insights for addressing the problem of blurred boundaries between the liver and tumors. Although dynamic convolution addresses the limitation of static convolutional receptive fields, it struggles to extract vertex information in graphs effectively when dealing with discontinuous, unstructured regions (e.g., liver and tumor in CT). This limitation prevents the neural network from effectively learning the key features of segmentation in rugged areas during the training process, resulting in poor segmentation accuracy in the liver and tumor boundaries, microtumors, and discontinuous regions.

To address the limitations of standard convolutional operations in capturing global features and handling discontinuous, unstructured regions, such as liver tumor boundaries, microtumors, and fragmented liver areas, we propose a liver and tumor segmentation method based on graph convolutional inference and SAM models, named GS_NeXt. A frozen SAM model is introduced to extract more comprehensive global information, while graph inference is applied to globally model unstructured local features. This enhances the network's understanding of CT images involving discontinuous liver regions, tumor boundaries, and small tumor areas. The major contributions of this work are as follows:

- Learn vertex relationships in the graph using boundary-aware input-dependent graph convolution layers and an improved Laplacian operator to extract feature information along liver-tumor boundaries.
- Embed a graph reasoning module between the encoder and decoder to capture the dependencies between global and local features, thereby enhancing the feature extraction capability in unstructured regions such as discontinuous liver areas and microtumor regions.
- Employing the feature extraction unit of the ConvNeXt-v2 network to suppress extraneous noise in the feature information, allowing the network to focus on important liver tumor features without increasing the network parameters.
- Construct a dual-encoder network by incorporating the SAM network with frozen weights to enhance the global modeling capability for liver and tumor regions.
- Applying a deep supervision mechanism in the decoder of the network to add supervisory signals to different layers of the decoder to optimize the network's ability to learn key features.

2. Related work

We categorized the related work into two aspects: (i) Different network dimensions and (ii) different feature extraction and feature fusion strategies.

2.1. Different network dimensions

3D-based segmentation networks can utilize the spatial information of slices in 3D CT data to enhance the network's spatial feature extraction capability. For example, Lyu et al. [15] proposed a liver tumor segmentation network (CouinaudNet), which utilizes the feature information labeled by the Couinaud algorithm to train the neural network. The Couinaud algorithm can subregionally label the data, generating an approximate tumor region mask that serves as a pixel-level supervisory signal to guide the network's training process. Meng et al. [16] proposed a dual-scale 3D convolutional neural network for liver tumor segmentation. This network utilizes a deep learning-based TDP-CNN approach, leveraging three dimensions of abdominal CT images. Additionally, they modified the training strategy to address the feature interference problem inherent in multi-stage image segmentation tasks. Segmentation models based on convolutional neural networks (CNNs) effectively extract local features; however, due to their limited receptive field, they struggle to capture global contextual information. The emergence of Transformer models has effectively addressed this limitation by enabling better global feature extraction. Li et al. [17] proposed DHT-Net in 3D liver and tumor segmentation using Dynamic Hierarchical Transformers (DHTrans) and Edge Aggregation Module (EAB). DHTrans can efficiently capture and model local textures globally, and EAB can generate detailed edge features. Li et al. [18] proposed the CC-DenseUNet, a segmentation network based on the U-Net framework that combines a dense connectivity structure with a cross-attention mechanism (CCA). By introducing dense connections to enhance the efficiency of gradient propagation between the encoder's layers, the encoder's feature extraction capability is improved. The CCA module is also utilized in the network's bottleneck structure to capture global contextual information in a spatial direction. Chi et al. [19] proposed the X-Net multi-branch network, which uses a simplified 3D convolutional kernel to extract inter-slice features from stacked slices filtered by slice scoring maps and combines intra-slice features and inter-slice features to generate liver and tumor in 3D CT data segmentation results. Although Transformer models are effective in capturing global features and addressing the challenge of long-range dependency, they also introduce significantly higher computational complexity. In recent years, state-space sequence models [20], primarily structured state-space sequence models (S4) [21], have achieved cutting-edge results in analyzing continuous, long sequence data as efficient building blocks or layers for constructing deep networks. Structured state-space sequence models (S4) have emerged as a powerful method for modeling long sequences with linear complexity in terms of input size, thereby revealing efficient modeling of both local and global dependencies. Zhu et al. [22] proposed a visually stabilized Mamba U-shaped network (ViS-Mamba) for 3D brain tumor segmentation. This method integrates Mamba modules with a U-Net backbone to capture long-range dependencies while preserving spatial resolution. By introducing a visual stability module and skip connections, the network enhances robustness in regions with fuzzy boundaries and complex morphologies, resulting in excellent segmentation performance on multimodal MRI data. Large-scale data not only increases storage requirements but also leads to longer processing times and increased consumption of computational resources.

2D-based segmentation networks typically handle 2D slice data from CT images, characterized by a small computational volume, fast processing speed, and suitability for real-time processing. To address the challenges in liver and tumor segmentation, especially the blurry boundaries and noise interference that often lead to over-segmentation or under-segmentation, many researchers have proposed enhanced 2D architectures. Tang et al. [23] designed a two-phase 2D segmentation network for segmenting livers and tumors. In the second stage, this network introduces an edge-enhanced network (E2-Net) to explicitly model the complementary relationship between the liver and tumor, as well as their edge information, to improve segmentation accuracy in the edge region. Seo et al. [24] proposed an improved U-shaped segmentation network (mU-Net) to enhance liver and tumor segmentation accuracy in CT images by combining the liver's and tumor's high-level features. The mU-Net improves on the U-Net by utilizing residual paths with inverse convolution and activation operations on jump connections, thereby increasing the efficiency of incorporating low-resolution features. Despite the success of CNNs in capturing local features, their limited receptive field restricts the modeling of global contextual information. To overcome this limitation, Transformer-based architectures have been explored in medical image segmentation [25]. These models are capable of modeling long-range dependencies, but also introduce greater computational complexity. Ni et al. [26] proposed DA-Tran, which uses an attentional fusion decoder (AFD) to generate supplementary information by fusing feature maps of liver CT images at each lesion stage, improving sensitivity to different lesion features and enabling the network to better understand and differentiate lesions at various stages. Hu et al. [27] developed Perspective+ Unet, which incorporates the Bi-Path Residual Block (BPRB) to combine standard and dilated convolutions for balanced local-global feature extraction, the Efficient Non-Local Transformer Block (ENLTB) to model long-range dependencies using a kernel-based approximation with linear complexity, and the Spatial Cross-Scale Integrator (SCSI) to unify multi-scale features across stages. In addition to improving feature modeling strategies, multi-stage cascaded frameworks have also proven effective. Christ et al. [28] proposed a cascaded full-convolution-based automatic segmentation network (CFCNs) for liver and tumor regions. The segmentation process is divided into two stages: First, a liver segmentation network is trained, and the segmentation result is used as input to another tumor segmentation network. This network then segments the tumor region only within the segmentation result of the liver segmentation network.

2.2. Different feature extraction and fusion strategies

Medical image segmentation, particularly in the context of liver and tumor segmentation, presents numerous challenges, including liver discontinuity, fuzzy boundaries between the liver and tumors, and the difficulty in detecting microtumors. To address these issues, researchers have proposed feature extraction and fusion strategies to enhance the model's localization accuracy and semantic representation capabilities. To address the deficiencies in liver and tumor boundary placement and small tumor segmentation, several researchers have explored dual-path networks that integrate different feature representations. Jiang et al. [29] designed a dual-coded branching network, MDCF-Net, which combines CNN and CnnFormer, enhancing the network's ability to sense 3D features by extracting information from both inside and outside the slices, as well as utilizing a novel feature map stacking method. To address the multi-scale nature of liver tumors and the difficulty in segmenting tumor edges, Wang et al. proposed a dual-branch liver tumor segmentation network (SBC-Net) [30]. SBC-Net introduces a context encoding module that utilizes a multi-scale adaptive kernel to more effectively identify the multi-scale features of tumors. A contour learning algorithm is combined with the Sobel operator to enhance boundary perception. Finally, a hybrid multitask loss function guides the network's learning for tumor multiscale and boundary features. Zhang et al. [31] proposed a dual-branch network for ultrasound image segmentation, which employs a parallel structure comprising a spatial detail branch and a semantic context branch. The spatial branch focuses on preserving fine-grained edge and boundary information, while the semantic branch captures high-level contextual features. A fusion module is then used to adaptively combine the features from both branches, enhancing the segmentation performance on fuzzy boundaries and small regions. This design demonstrates strong robustness to noise and anatomical variation, which provides valuable insights for liver and tumor segmentation in complex scenarios. Kaya et al. [32] proposed the Fusion-Brain-Net. This dual-encoder deep fusion network integrates low-level spatial features and high-level semantic representations through a hierarchical fusion strategy for accurate brain tumor classification from multi-sequence MRI scans. The model comprises two separate encoding pathways, designed to extract complementary information, which are later combined using a deep fusion module that preserves both spatial precision and contextual abstraction. Although their work focuses on classification, the dual-path fusion strategy offers valuable insights for designing segmentation networks that require fine-grained localization and contextual understanding. To overcome CNN's limitation in modeling long-range dependencies, researchers have incorporated Transformer-based architectures. For global feature extraction, Di et al. [33] proposed an end-to-end hybrid network (TD-Net). TD-Net embeds a Transformer module and an orientation guiding block in a convolutional network to achieve automatic segmentation of liver tumors from CT images. The Transformer module can create remote dependencies between tumor and non-tumor features to compensate for the shortcomings of CNNs in establishing remote dependencies between features. The Transformer module can correct the feature maps generated by multiple convolutional layers, thus improving the accuracy of tumor boundary recognition. The aforementioned networks, with their dual-branch encoder designs, can effectively address issues such as liver discontinuity, fuzzy boundaries between the liver and tumors, and the segmentation of small tumors. However, they face limitations in capturing global features.

Recent efforts have explored prompt-driven foundation models in medical image segmentation. The Segment Anything Model (SAM) [34] is a generalist segmentation model trained on natural images, capable of generating masks from user prompts in a zero-shot manner. However, its performance drops significantly when applied to medical data due to domain shifts, limited sensitivity to small structures, and weak boundary representations. Zhang et al. [35] systematically evaluated SAM across medical modalities to address these challenges. They categorized subsequent adaptations into fine-tuning strategies, automatic prompting, architecture modification, and 3D extensions. Among these, Qin et al. [36] proposed DB-SAM, a dual-branch network tailored for medical image segmentation. DB-SAM integrates a ViT branch enhanced by channel attention to model global semantics and a convolutional branch designed to capture fine-grained local features. A bilateral cross-attention mechanism facilitates feature exchange between the two branches, while an adaptive fusion module combines their outputs. This architecture enhances the segmentation of small, ambiguous, or irregular anatomical regions, achieving higher accuracy on both 2D and 3D medical image datasets compared to the baseline SAM and earlier adaptations. The SAM model achieves stronger global semantic modeling while preserving fine details, making it a powerful tool for segmenting complex structures in medical images.

Feature fusion is a crucial component of modern neural network architecture, enhancing the performance of convolutional neural networks. The InceptionNet network [37] utilizes multiple convolutional kernels of varying sizes within the same layer, and the feature fuses their outputs, thereby enhancing the network's ability to process feature information at different scales within the input image. This multi-kernel fusion strategy enhances the model's ability to handle tumors with large-scale variation, addressing the challenge of lesion heterogeneity. In addition, the residual network ResNet and its successor, iterative networks [38], address the deep network's gradient vanishing and explosion issues by combining features from the convolutional layer with residual learning features via short jump connections, followed by uniform output. The Feature Pyramid Network (FPN) and U-Net feature fusion, which combines low-resolution features with high-resolution features through jump connections, enables the network to access shallow features in the encoder and utilize them in the decoder. This approach mitigates the problem of feature loss and degradation of detailed information during the encoding-to-decoder information transfer, effectively improving the accuracy of the segmentation task. By preserving spatial details, these designs offer valuable insights to address the challenge of blurred boundaries in liver and tumor segmentation. Feature fusion techniques are also indispensable in liver and tumor segmentation networks. Kuang et al. [39] proposed an MP-LiTS segmentation network, which embeds a multi-stage channel attention (MCDA) module and a multi-scale supervised SWL function. MCDA dynamically assigns weights to each stage through a stage fusion operation, learning the interrelationships between feature information at different scales. Afterwards, the stage features are spliced along the channel dimension, and the attention features are fused through the channel attention mechanism, which effectively improves the network's learning ability for different levels of features of liver and tumor. This design enhances the network's ability to adapt to tumors at multiple scales and improves feature discrimination, particularly for small or irregular lesions. Liu et al. [40] proposed a liver segmentation network, HI-Net, based on hierarchical inter-scalar multiscale feature fusion. It includes a hierarchical multiscale feature fusion module and multiple inter-scalar dense connections, which integrate feature information at different levels, thus mitigating the potential loss of high-level semantic information. By preserving high-level semantics, HI-Net addresses the common issue of semantic degradation in deep networks, improving robustness for large and complex liver structures. Zhan et al. [41] proposed a three-direction feature fusion of vertical, sagittal, and coronal axes for the liver and tumor segmentation network (TFVS), which adequately fuses the feature information of the three dimensions to improve the accuracy and effectiveness of the segmentation process. This three-plane fusion approach significantly improves structural continuity across 3D space, which is critical in addressing liver discontinuity and spatial ambiguity. Therefore, advanced feature extraction and fusion techniques can effectively enhance the performance of segmentation networks by specifically addressing key challenges in liver and tumor segmentation, such as blurred liver-tumor boundaries and liver discontinuity.

3. Method

3.1. Structure of the proposed GS_NeXt

The structure of the proposed GS_NeXt[1] network is shown in Figure 2. It consists of an upper branch ConvNeXt encoder, a lower branch SAM encoder, a Graph Reasoning Module (GRM), a ResUNet decoder, and a deep supervision mechanism.

First, a 2D CT image of 448 × 448 size is input to the ConvNeXt encoder. It consists of the ConvNeXt-v2 feature extraction module, which uses a global response normalization layer on top of ConvNeXt-v1 to enhance the feature competition between channels. It thus makes the network more flexible in capturing feature information from different channels.

Second, a SAM encoder with frozen weights is employed, allowing the SAM encoder to capture long-distance dependencies in the input image through the self-attention mechanism and generate global semantic information, thereby enhancing the network's global inference ability for the liver and its tumors.

Then, the extracted features are processed by the GRM graph inference module for graph processing. GRM models the local features of the input features by constructing the nodes and edges of the graph, which enhances the network’s comprehension of the complex structured information and improves the network's understanding of the CT images with blurred liver boundaries, liver discontinuities, and microtumor regions.

Finally, using the decoder of the ResUNet network, the extracted features are upsampled, and the output of the intermediate layer is supervised using a depth supervision mechanism to optimize the feature representations across different levels. The final results of the liver and tumor segmentation are generated.

[1] The source code is publicly available at https://github.com/Coder-GSNeXt/GS_NeXt

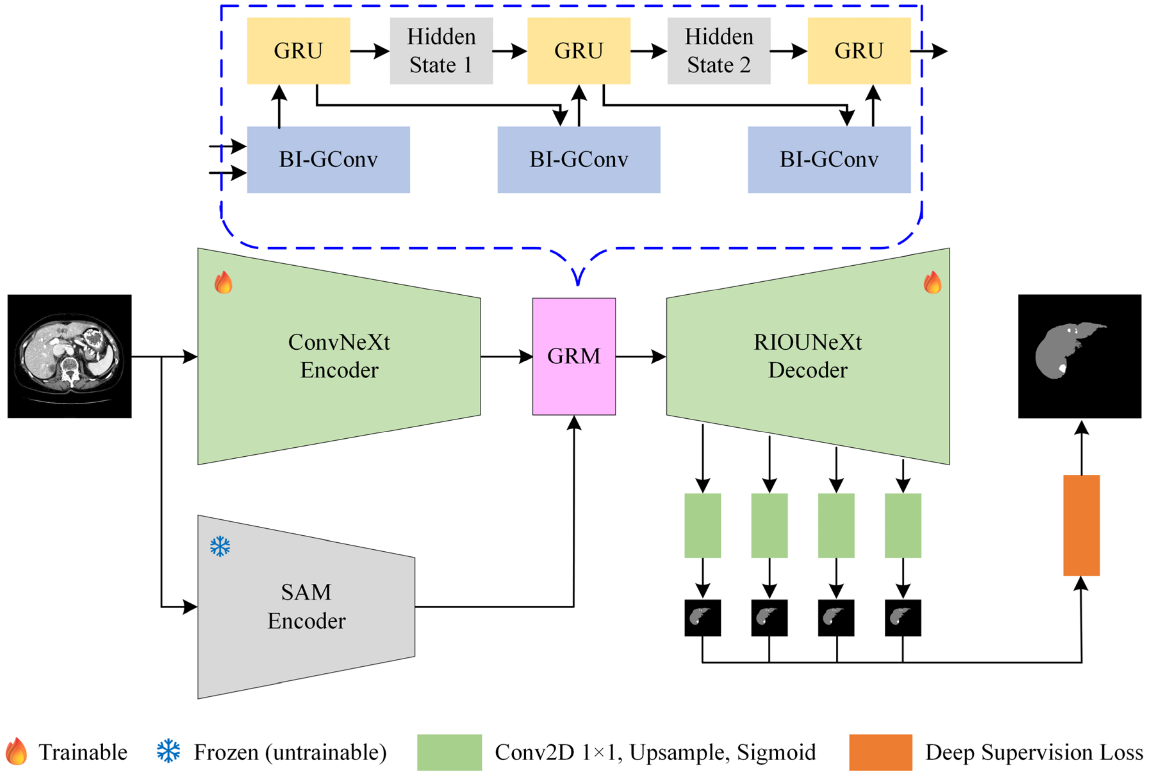


Figure 2. The structure of the proposed GS_NeXt.

3.2. Graph convolution with graph inference module

Given an input region feature $R_s \in \mathbb{R}^{N \times C}$, where C is the channel size, $N = H \times W$ and N is the number of vertices in the graph of the input feature. The input-associated adjacency matrix Boundary-aware input-dependent graph convolution (BI-GConv) [42] can perform channel attention on two diagonal matrices ($\tilde{\Lambda}^c$ and $\tilde{\Lambda}^s$) over the dot-product distances embedded in the input vertices, measuring the spatial weighting relationships between vertices separately. $\tilde{\Lambda}^c(R_s) \in \mathbb{R}^{C \times C}$ It is the diagonal matrix containing the channel attention, and its derivation formula is shown below:

$$\tilde{\Lambda}^c(R_s) = diag\left(MLP\left(Pool_c(R_s)\right)\right) \tag{1}$$

where $Pool_c(\cdot)$ denotes the global maximum pooling of each vertex embedding, and $MLP(\cdot)$ is a multi-layer perceptron containing a hidden layer.

$\tilde{\Lambda}^s(R_s) \in \mathbb{R}^{N \times N}$ is a spatially weighted diagonal matrix whose computational procedure is defined as follows:

$$\tilde{\Lambda}^s(R_s) = diag\left(Conv\left(Pool_s(R_s)\right)\right) \tag{2}$$

where $Pool_s(\cdot)$ denotes the global maximum pooling for each position where vertices are embedded along the channel axis; $Conv(\cdot)$ is a 1 × 1 convolutional layer.

In the implementation of BI-GConv, we initialize the convolutional and attention-related parameters using the Kaiming (He) initialization strategy, which is well-suited for networks with ReLU activations. This approach ensures stable variance propagation across layers, avoiding vanishing or exploding gradients. We empirically observe that compared to Xavier initialization, Kaiming initialization results in faster convergence during early training and more stable learning dynamics. These observations are consistent across multiple training runs, particularly on the LiTS17 dataset, which supports the robustness of our model design.

The spatial and channel-attention augmented input vertex embedding gives the input correlation adjacency matrix. The input correlation adjacency matrix $\overline{A}$ is initialized as shown in Eq (3):

$$\begin{aligned}\overline{A} &= \psi\left(R_S, W_\psi\right) \cdot \tilde{\Lambda}^c(R_S) \cdot \psi\left(R_s, W_\psi\right)^T \\ &+ \phi(R_S, W_\phi) \cdot \phi\left(R_S, W_\phi\right)^T \odot \tilde{\Lambda}^s(R_S)\end{aligned} \tag{3}$$

where $\cdot$ denotes the matrix product; $\psi(R_s, W_\psi) \in \mathbb{R}^{N\times C}$ and $\phi(R_s, W_\phi) \in \mathbb{R}^{N\times C}$ both denote the linear embedding (1 × 1 convolution); $\odot$ denotes *the Hadamard* product; W_ψ and W_ϕ are learnable parameters.

The predicted boundary function mapping $B_s \in \mathbb{R}^{N\times 1}$ is integrated into the Laplace matrix $\tilde{L}$ to fuse the boundary information and fused into the spatial weighting matrix $\tilde{\Lambda}^s(R_s)$ to obtain the boundary-aware spatial weighting matrix $\tilde{\Lambda}_b^s(R_s, B_s)$, and the derivation formula for $\tilde{\Lambda}_b^s(R_s, B_s)$ is shown below:

$$\tilde{\Lambda}_b^s(R_s, B_s) = Conv\left(Pool_s(R_s)\right) \cdot \left(Conv\left(Pool_s(R_s \odot B_s)\right)\right)^T \tag{4}$$

where $\odot$ denotes the broadcast *Hadamard* product across channels; $Conv(\cdot)$ is a 1 × 1 convolutional layer.

In this way, larger weights can be assigned to emphasize the features of the boundary pixels. Thus, the boundary-aware map convolution can learn the spatial features of the boundary while reasoning about the correlation between regions. The input correlation neighbor matrix for boundary perception, $\tilde{A}$, is calculated as shown below:

$$\begin{aligned}\tilde{A} &= \psi\left(R_S, W_\psi\right) \cdot \tilde{\Lambda}^c(R_S) \cdot \psi\left(R_s, W_\psi\right)^T \\ &+ \zeta(R_s, W_\zeta) \cdot \zeta(R_s, W_\zeta)^T \odot \tilde{\Lambda}_b^s(R_s, B_s)\end{aligned} \tag{5}$$

where $\zeta(R_s, W_\zeta) \in \mathbb{R}^{N\times C}$ is a 1 × 1 convolution; W_ζ is a learnable parameter.

Using the constructed $\tilde{A}$, the normalized Laplace matrix $\tilde{L}$ is defined as shown below:

$$\tilde{L} = I - \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \tag{6}$$

where I is a unit matrix, $\tilde{D}$ is a diagonal matrix, and $\tilde{D}_{ii} = \widetilde{\Sigma \imath_j} \tilde{A}_{i,j}$ denotes the degree of each vertex.

The derivation of the computational degree matrix $\tilde{D}$ is shown in Eq (7):

$$\tilde{D} = diag\left(\boldsymbol{\psi}\left(\tilde{\Lambda}^c(\boldsymbol{\psi}^T \cdot \vec{1})\right)\right) + diag\left(\boldsymbol{\zeta}(\boldsymbol{\zeta}^T \cdot \vec{1})\right) \odot \tilde{\Lambda}_b^s \tag{7}$$

where $\vec{1}$ denotes the all-one vector in $\mathbb{R}^N$ and the calculations in the inner brackets are performed first.

Based on the normalized Laplace matrix $\tilde{L}$, with R_s as the input vertex embedding, the single-layer BI-GConv structure (shown in Figure 3) is calculated as follows:

$$Y = \sigma\left(\tilde{L} \cdot R_s \cdot W_G\right) + R_s \tag{8}$$

where $W_G \in \mathbb{R}^{C \times C}$ denotes the trainable weights of BI-GConv; σ is the ReLU activation function; and Y is the output vertex features. Additionally, BI-GConv incorporates residual concatenation to preserve the features of the input vertices.

The Graph Reasoning Module (GRM, shown on the upper side of Figure 2) is constructed by combining the BI-GConv layer with Gated Recurrent Units (GRUs) [43] in a chained fashion, enabling GRM to capture and process both long-term and short-term dependencies between chained layers. Among them, the BI-GConv layer is used to extract features, while the GRU helps to model the dependencies of these features. Introducing GRM into the segmentation network can enhance the network's global modeling of local features of the liver and tumor, improve the network's learning ability for boundary features, and improve the network's segmentation accuracy in regions with segmentation difficulties, such as liver discontinuities and the boundaries of the liver and its tumors.

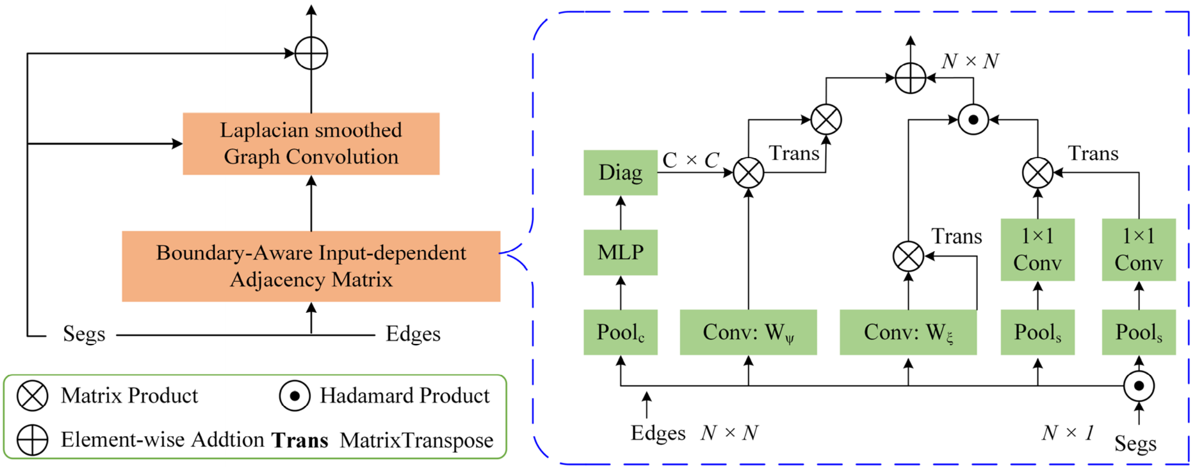


Figure 3. Structure of Boundary-aware input-dependent graph convolution (BI-GConv).

3.3. *Upper-branching ConvNeXt encoder*

The up-branching ConvNeXt encoder proposed in this section consists of a 3 × 3 convolutional layer, four feature extraction modules (ConvBlocks), and four down-sampling layers, whose structure is shown in Figure 4. Among them, ConvBlock-1, ConvBlock-2, and ConvBlock-4 consist of one ConvNeXt-v2 Block [44], and ConvBlock-3 consists of three ConvNeXt-v2 Blocks sequentially. The global response normalization (GRN) layer in the ConvNeXt-v2 Block is a convolutional neural network layer. To achieve feature competition among different channels, GRN takes a normalization operation on the feature maps of each channel, respectively. The GRN layer performs three operations: Global feature aggregation, feature normalization, and feature calibration across channels. In the global feature aggregation operation, the L2 paradigm is introduced to perform the aggregation operation on feature maps across channels, resulting in an aggregated feature vector.

In the feature normalization operation, the aggregated vector is feature-normalized using the standard division normalization function. In the feature calibration operation, the original feature maps are calibrated using the normalized feature vectors. The CT images usually have varying degrees of noise, especially in the case of liver tumors with blurred boundaries. Using an encoder composed of a feature extraction module with a GRN layer, the over-response caused by some irrelevant noise can be suppressed by normalizing the feature maps during the feature extraction process, which enables the encoder to focus more on the critical liver tissues and tumor regions.

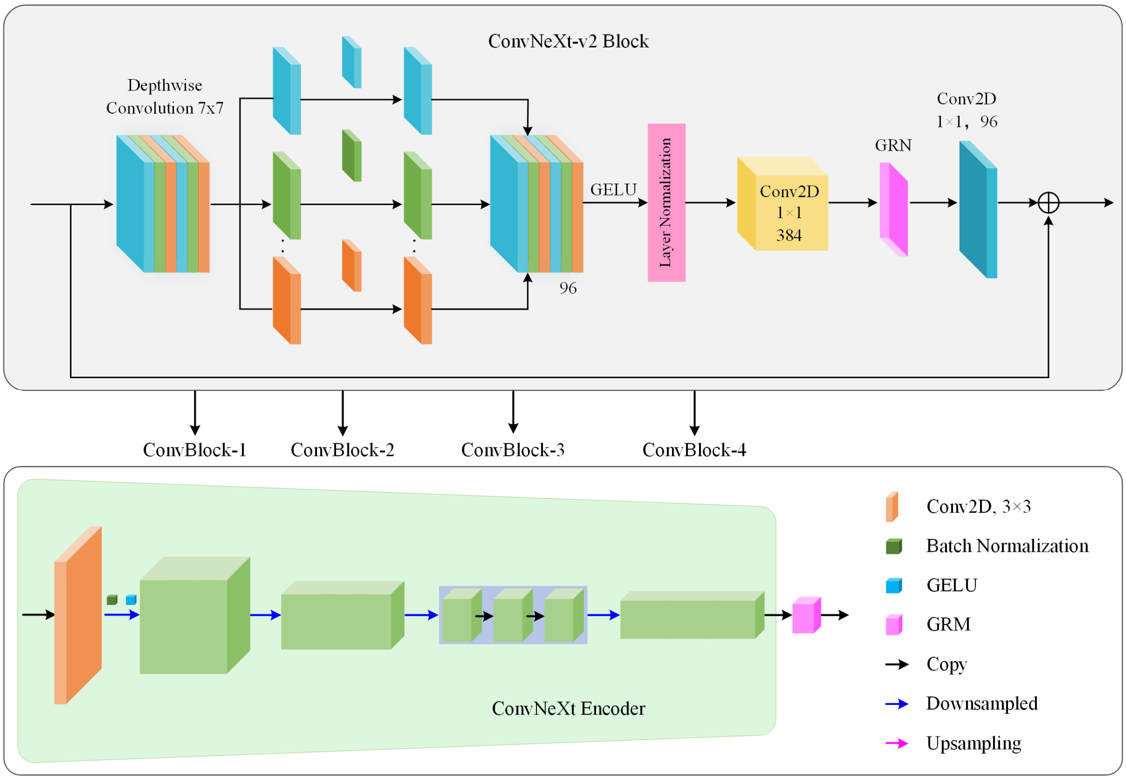


Figure 4. Structure diagram of the ConvNeXt encoder on the upper branch.

3.4. Lower branch SAM encoder

In this section, the down-branching SAM encoder with frozen weights is designed to guide the liver and tumor segmentation tasks, thereby enhancing the network's ability to extract global feature information and improve the segmentation of discontinuous liver regions. SAM is a network micromodel for image segmentation, trained on 11 million images with 1 billion masks. Several medical image segmentation studies have attempted to utilize SAM's zero-sample capability, but SAM struggles to achieve high-performance results on medical images. This is due to the domain gap between the training images, as the images used to train SAM are natural and differ significantly from medical images. However, with its global feature extraction and generalization capabilities, the network encoder retains high-order global contextual information during the feature extraction process.

Figure 5 shows the SAM encoder structure: SAM is based on the ViT architecture but with the classification header part removed. The image encoder of the SAM network consists of a standard ViT. The CT image is processed as follows when it is fed into the encoder: First, the input image is divided into multiple image blocks of equal size. Next, the image blocks are spread using vectorization coding and positional coding. Then, the flattened image blocks are adjusted into feature vectors that match the input dimensions of the Transformer encoder using the linear projection module. Finally, feature extraction is performed on the feature vectors containing the position-encoded information using the Transformer encoder. The convolutional layer in the SAM encoder performs a block embedding operation on the input image, where the size of each block is set to 16 × 16 with a step size of 16, and the convolution operation is performed on the input image. This process reduces the size of the feature map to 1/16 of its original size while increasing the number of channels from the initial 3 to 768. A tunable positional embedding matrix is initialized with all elements set to zero. After completing the addition of the positional embedding, the feature map is sequentially passed through a feature extraction unit consisting of 16 Transformer encoder blocks. Among these 16 blocks, 12 employ a local attention mechanism based on window division, i.e., the feature map is subdivided into 14 × 14 windows for processing, while the remaining 4 employ a global attention mechanism, and these global attention modules are skillfully arranged between the window attention modules to achieve more comprehensive feature extraction.

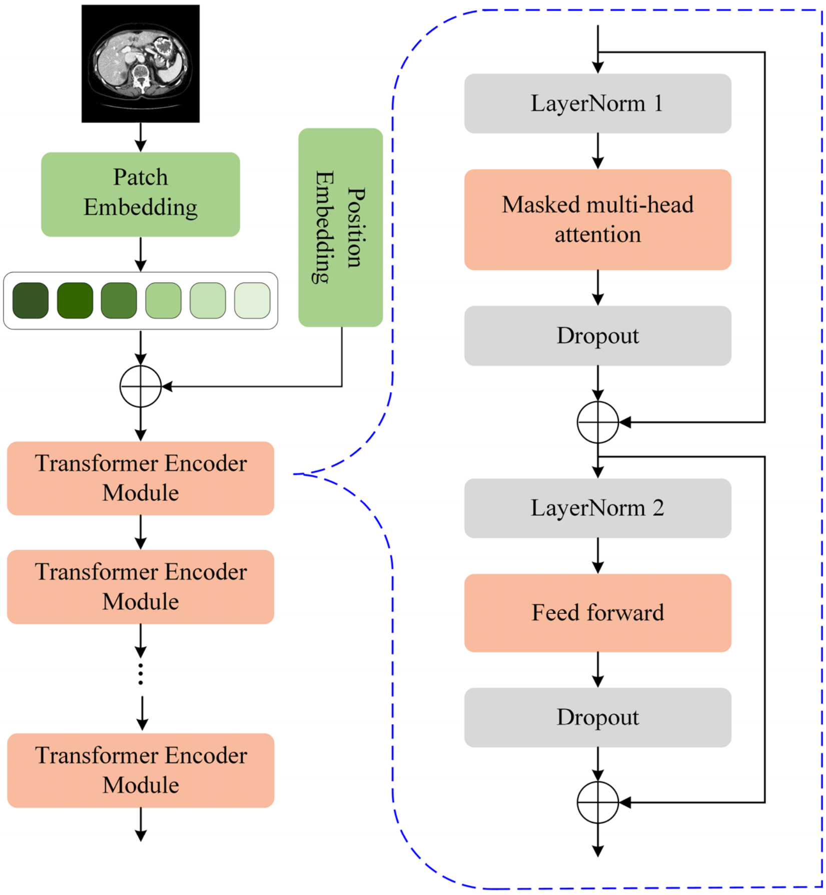


Figure 5. SAM encoder architecture.

3.5. Deep supervision mechanism

The structures of the liver and tumor typically exhibit multi-scale and multi-layer characteristics, with tissue boundaries often blurred, making it difficult for traditional single-output supervisory mechanisms to capture detailed information adequately. For this reason, we introduce the deep supervision mechanism in the network decoder, as shown in Figure 6. The core idea of deep supervision as an improved neural network training paradigm is to introduce auxiliary supervisory signals in multiple implicit layers of the network architecture in parallel, thus optimizing the gradient propagation path and enhancing the network's ability to characterize features [45].

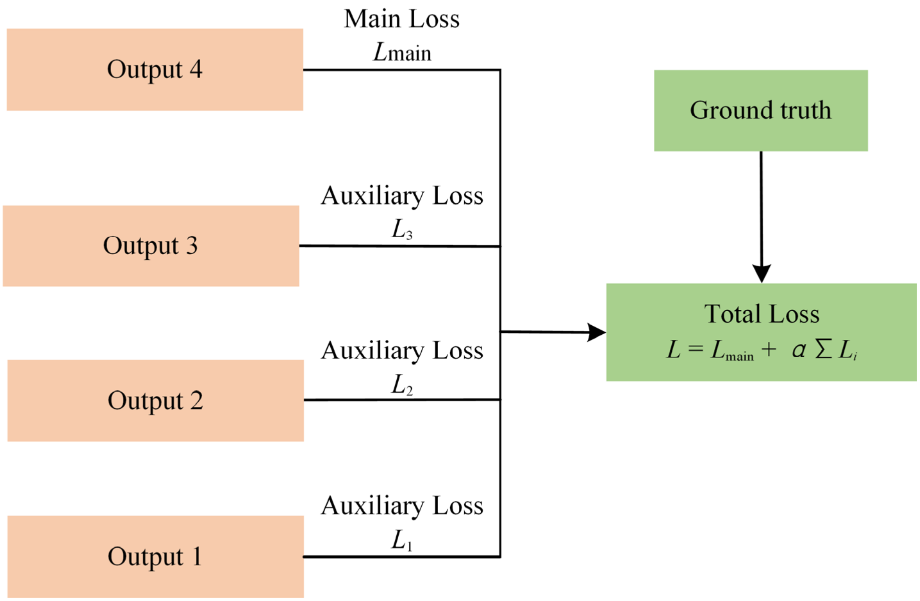


Figure 6. Deep supervision mechanism.

As shown in Figure 6, the deep supervision mechanism employed in this section is designed to produce four probabilistic outputs. The output pixel-level probability maps are then compared with their corresponding gold standard of the same size. The loss values of the different layers are obtained separately, and these loss values are weighted and summed to obtain the total loss of the network. The total loss L after each iteration is defined as follows:

$$L = L_{\text{main}} + \alpha \sum_{i=1}^{N} L_i (i = 1, 2, 3) \tag{9}$$

where L_{main} is the main loss, which represents the loss value of the output layer of the network decoder (the layer whose output probability map has the same size as the network input), and L_1, L_2, and L_3 are the auxiliary loss functions, which represent the loss value of the middle layer of the decoder. Parameter α is the weight coefficient; the output layer contains more complex features than the intermediate layer, so the weights of the different losses of the intermediate layer and the output layer are assigned to increase the percentage of the loss in the output layer. We update every 20 rounds (Epochs) according to the formula.

3.6. Evaluation metrics

In this paper, six metrics, namely, Dice coefficient (Dice), volume overlap error (VOE), relative volume error (RVD), average symmetric surface distance (ASD), maximum average surface distance (MSD), and root-mean-square symmetric surface distance (RMSD), are used to evaluate the segmentation results. In the definition of these six evaluation metrics, A and B represent the data of gold standard segmentation results and network segmentation results, respectively.

✧ Dice: Expressed as the ratio of the overlapping area of the gold standard and the segmentation result to the total area; the larger the Dice value is, the better the segmentation effect. The formula is as follows:

$$Dice(A,B)=\frac{2|A\bigcap B|}{|A|+|B|} \tag{10}$$

- ✧ VOE: Denotes the volume overlap error between the two sets of voxels, defined as follows:

$$VOE(A,B)=1-\frac{|A\bigcap B|}{|A\bigcup B|}\times 100(\%) \tag{11}$$

- ✧ RVD: Indicates the difference in volume or area of the segmentation result relative to the gold standard, which is an asymmetric indicator and cannot be used as the only measure of segmentation quality, and is defined by the formula:

$$RVD(A,B)=\frac{|B|-|A|}{|A|}\times 100(\%) \tag{12}$$

- ✧ ASD: Is a metric that evaluates the difference between the segmentation result and the true boundary. It calculates the average distance from *S(A)* to *S(B)* and *S(B)* to *S(A),* respectively. The voxels *S(A)* and *S(B)* denote *A*'s and *B*'s real and predicted boundaries, respectively. *p* and *q* represent the surface voxels in *S(A)* and *S(B)*. The shortest distance from any voxel *v* to *S(A)* is defined as:

$$d\left(v,S(A)\right)=\min_{p\in S(A)}\|v-s_A\| \tag{13}$$

where || || denotes the Euclidean distance. The average symmetric surface distance (ASD) formula is defined as follows:

$$ASD(A,B)=\frac{1}{|S(A)|+|S(B)|}\left(\sum_{p\in S(A)}d^2\left(p,S(B)\right)+\sum_{q\in S(B)}d^2\left(p,S(A)\right)\right) \tag{14}$$

- ✧ MSD: Similar to ASD, the operation of calculating the average value is changed to the operation of calculating the maximum value. Its formula is defined as:

$$MSD(A,B)=max\left\{max_{p\in S(A)}d(p,S(B),max_{q\in S(B)}d(p,S(A)\right\} \tag{15}$$

- ✧ RMSD: Indicates root mean square symmetric surface distance, which is one of the key criteria for assessing segmentation accuracy. The closer the value is to 0, the better the segmentation is. Its formula is defined as:

$$RMSD(A,B)=\sqrt{\frac{1}{|S(A)|+|S(B)|}}\times\sqrt{\sum_{p\in S(A)}d(p,S(B)+\sum_{p\in S(B)}d\left(p,S(A)\right)} \tag{16}$$

4. Experimental setup and analysis of results

4.1. Dataset source and division

The open abdominal CT image datasets for liver and tumor segmentation used in this paper are

LiTS17[2] and 3DIRCADb[3].

The LiTS17 dataset consists of 131 training sets and 70 test sets. Each CT has an axial plane resolution of 512 × 512 pixels, with the number of slices ranging from 42 to 1026 and slice spacing ranging from 0.45 to 6.0 mm. Medical image segmentation is typically performed randomly based on the training set of the LiTS17 dataset, as the test set of the LiTS17 dataset does not have publicly available image labels.

The 3DIRCADb-01 dataset provides more complex data about the liver and tumors, with a resolution of 512 × 512 pixels. It includes 10 men and 10 women with enhanced CT, where three-quarters of the women have liver tumors. Additionally, 3DIRCADb-02 consists of two anonymized enhanced 3D CT scans of the chest and abdomen. The specific parameters of these two datasets are shown in Table 1.

Table 1. Two public data set parameters ("-" means none).

Dataset	Training set	Test set	In-slice resolution (mm)	Inter-slice resolution (mm)	Number of slices	Slice size
LiTS17	131	70	0.55–1.0	0.45–6.0	42–1026	512 × 512
3DIRCADb	20	-	0.56–0.81	1.0–4.0	74–225	512 × 512

Table 2. Experimental environment and parameter configuration.

Item	Version	Parameter	Configuration
Operating System	Ubuntu 18.04	Learning Rate	0.001
GPU	RTX3080Ti	Batch Size	6
Python Version	Python 3.8	Optimizer	Adam
CUDA Version	CUDA11.4	Momentum	0.9
Framework	Pytorch 1.9	Number of Epochs	60

For the LiTS17 dataset, a randomized division strategy is adopted to divide the 131 subsets into a training validation set (116 cases) and a test set (15 cases); for the 3DIRCADb dataset, the 20 case samples are randomly divided into a training validation subset (12 cases) and a test subset (8 cases). The hardware resources and environment configurations used in this section are shown in Table 2. The key hyperparameters in our method are determined based on standard practices in prior literature and empirical tuning on the validation set.

4.2. Dataset conversion

Our research object of this paper is to segment the liver and tumor in a 2D dataset automatically. The publicly available liver and tumor datasets, LiTS17 and 3DIRCADb, are 3D CT datasets; therefore, converting them to 2D datasets is necessary. The process involves several steps: Dataset loading and preprocessing, conversion of the 3D CT dataset to a 2D image dataset, removal of invalid slices that do not contain the liver, and division of the training set, validation set, and testing set. The dataset conversion process is shown in Figure 7.

[2] The LiTS17 dataset is publicly available at: http://www.lits-challenge.com

[3] The 3DIRCADb dataset can be accessed at: https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/

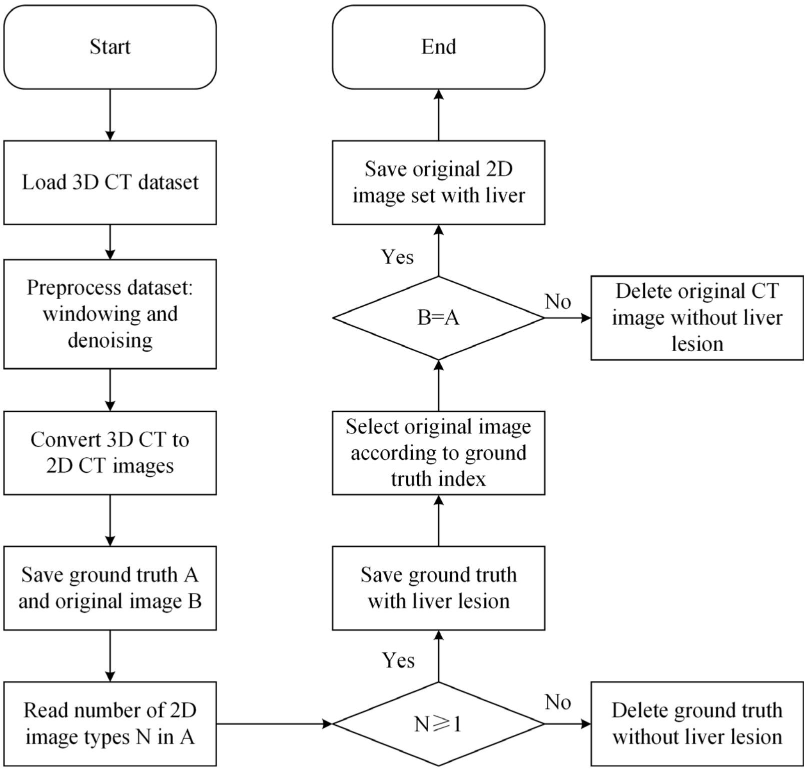

Start
Load 3D CT dataset
Preprocess dataset: windowing and denoising
Convert 3D CT to 2D CT images
Save ground truth A and original image B
Read number of 2D image types N in A
N≥1
No
Delete ground truth without liver lesion
Yes
Save ground truth with liver lesion
Select original image according to ground truth index
B=A
No
Delete original CT image without liver lesion
Yes
Save original 2D image set with liver
End

Figure 7. Flowchart of dataset conversion.

i) *Dataset loading:* The 3D dataset is first loaded to highlight the liver region of interest using the windowing technique, and then the Hu values of the CT are converted to grayscale values for image preprocessing.

ii) *Conversion of 3D CT dataset to 2D images:* CT scanning is performed by scanning an object layer by layer and obtaining a series of tomographic images (i.e., 2D images), which reflect the internal structure of the body from different perspectives (cross-sectional, sagittal, and coronal planes). 3D CT dataset consists of several 2D slices of images, which are arranged at a specific spacing, and reconstruction can be performed to obtain the 3D data; thus, after receiving the number of slices and slice spacing of a CT dataset, 3D CT can be converted into 2D slices.

iii) *Removal of invalid slices without liver:* Ensure that only valid slices containing liver tissue are retained for subsequent analysis, processing, and experiments. This process not only enhances the accuracy and reliability of the data but also effectively mitigates the interference of noisy data and eliminates bias in results due to irrelevant samples, thereby improving the efficiency and effectiveness of network training. By screening effective slices, the network can focus more precisely on liver-related regions.

4.3. Dataset enhancement

In this section, data enhancement techniques are employed to expand 2D slices containing tumors, thereby addressing the imbalance between liver and tumor categories in medical images. Specific methods include adopting elastic deformation to simulate tissue deformation characteristics, utilizing central magnification to enhance the characteristics of the tumor region, and employing random rotation to increase data diversity. The data enhancement technique can effectively alleviate the problem of imbalance between the liver and liver tumor categories. The comparative visualization results before and after data enhancement are shown in Figure 8.

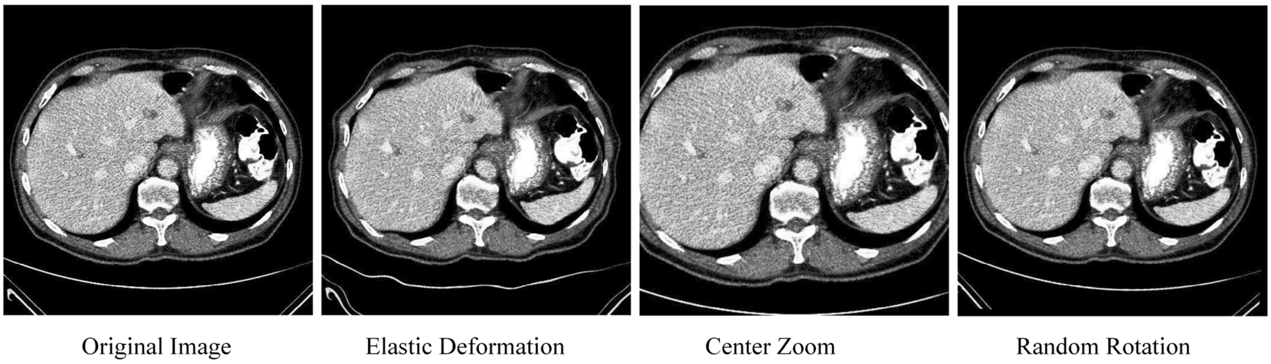


Figure 8. Illustration of data augmentation.

4.4. Ablation experiments

In this section, we utilize the ConvNeXt-v2-based encoder and ResUNet-based decoder as the Base network, and superimpose graph inference (GI), SAM encoder, and deep supervision (DS) mechanisms in the decoder to validate the effectiveness of the proposed GS_Next framework.

We perform the ablation experiments on the LiTS17 dataset. By comparing the segmentation metrics and visualization results, we observe that the network's segmentation accuracy gradually improves with the superposition of the three models, which proves the rationality and effectiveness of the proposed combined model.

Table 3. Liver segmentation results in the ablation experiment.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
Base	95.18 ± 0.39	5.42 ± 0.63	0.42 ± 1.54	1.34 ± 0.41	15.74 ± 2.17	4.51 ± 1.47
+DS	95.46 ± 0.67	5.13 ± 0.42	0.39 ± 1.63	1.35 ± 0.24	14.37 ± 5.30	4.18 ± 1.49
+DS+SAM	96.83 ± 0.34	4.92 ± 1.72	0.35 ± 1.31	1.21 ± 0.28	13.42 ± 4.27	3.73 ± 1.37
+DS+SAM+GI (Ours)	**97.74 ± 0.31**	**3.68 ± 0.83**	**0.23 ± 1.01**	**1.01 ± 0.17**	**10.86 ± 1.65**	**2.04 ± 1.62**

Table 4. Tumor segmentation results in the ablation experiment.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
Base	83.83 ± 10.12	29.26 ± 9.18	0.39 ± 0.64	5.26 ± 2.49	16.31 ± 5.26	9.54 ± 4.43
+DS	84.58 ± 2.69	27.74 ± 2.36	0.31 ± 1.54	4.84 ± 0.73	15.25 ± 4.86	7.18 ± 2.94
+DS+SAM	86.62 ± 0.86	24.33 ± 1.81	0.26 ± 1.67	3.36 ± 0.59	13.32 ± 4.49	6.76 ± 1.12
+DS+SAM +GI (Ours)	**87.25 ± 0.23**	**22.60 ± 1.57**	**0.11 ± 1.28**	**2.23 ± 0.64**	**12.97 ± 2.35**	**5.34 ± 1.27**

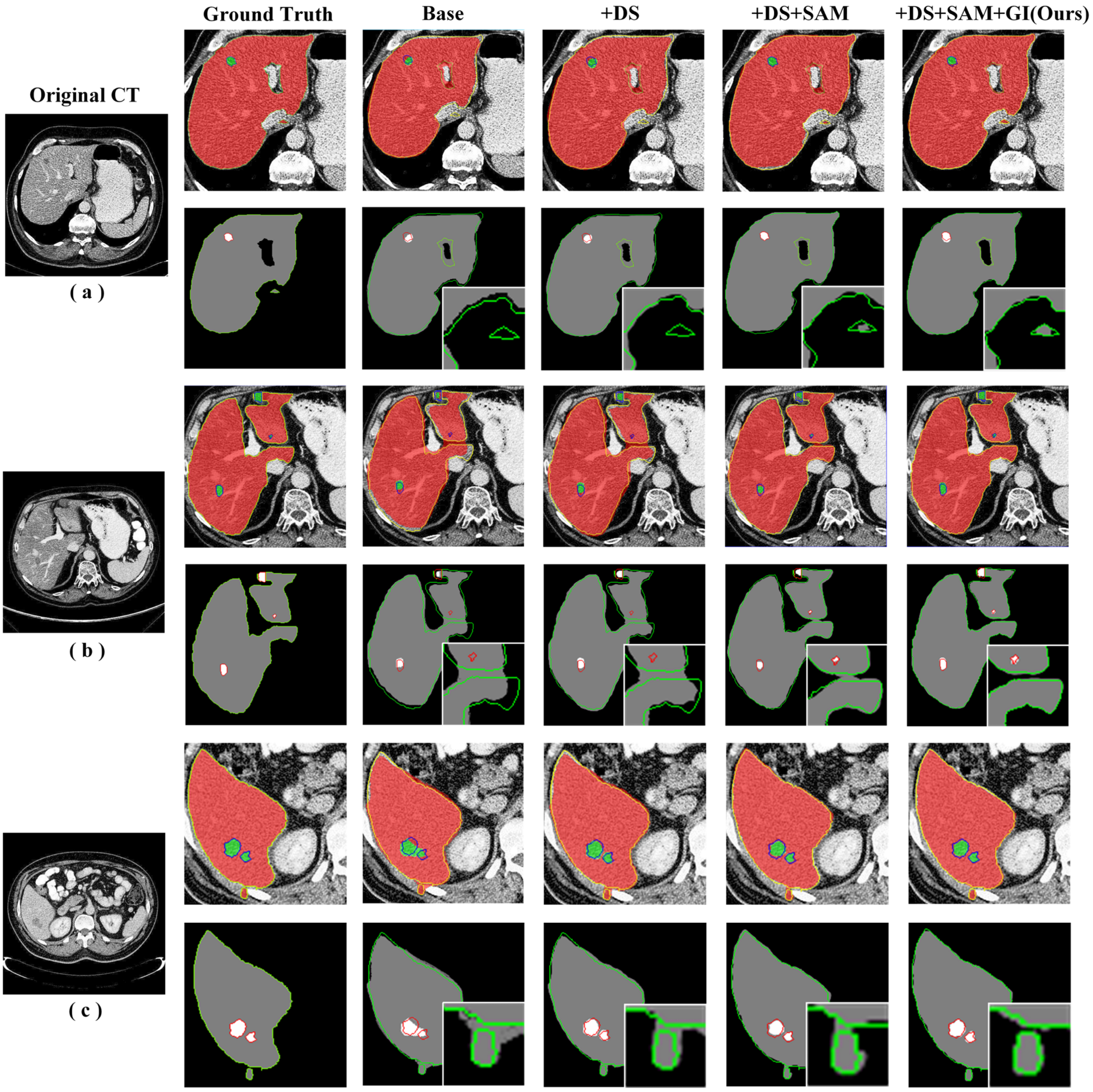


Figure 9. Visualization of liver and tumor segmentation results in ablation experiments.

Tables 3 and 4 present the base network segmentation results with the addition of the Deep Supervision Mechanism, SAM Encoder, and Graph Convolutional Graph Inference modules. The Dice value gradually improves, and the remaining five evaluation results also show gradual improvement. This indicates that the network's segmentation performance is steadily improving. Figure 9 shows three CT cases: Figure 9(a) contains a discontinuous region of the liver, Figure 9(b) includes a discontinuous microtumor region, and Figure 9(c) contains a discontinuous tumor region. From the segmentation visualization of Figure 9(a)–(c), it can be seen that the addition of the deep supervision mechanism improves the segmentation of the network for the liver boundaries, and the SAM encoder, graph convolution, and graph inference modules enhance the segmentation of the network for the liver discontinuous and microtumor regions.

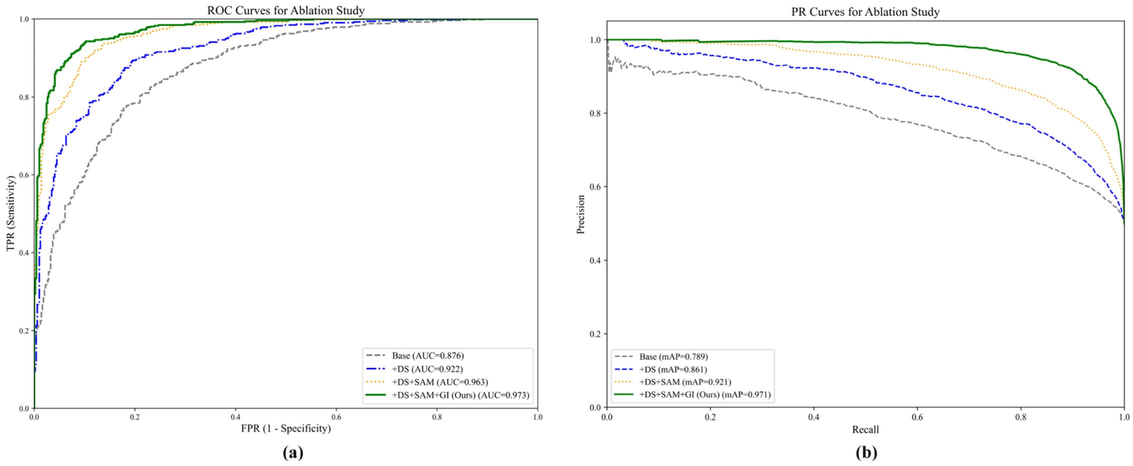


Figure 10. Evaluate ablation experiments on the LiTS17 dataset by ROC and PR curves.

To more intuitively evaluate the results of the ablation experiments, we plot the ROC and PR curves for different module combinations, as shown in Figure 10. From Figure 10(a), it can be observed that the AUC of the Base model (gray dashed line) is only 0.876. When the DS (blue dashed line), SAM (yellow dashed line), and GI (green solid line) modules are added to the Base model sequentially, the performance improves progressively. In particular, after introducing the GI module, the AUC reaches the highest value of 0.973, thereby validating the effectiveness of the proposed method. From Figure 10(b), the Base model (gray dashed line) achieves the lowest mAP of 0.789, indicating poor precision and recall performance in liver tumor segmentation. With the addition of the DS module (blue dashed line), the mAP increases to 0.861. When the SAM module (yellow dashed line) is further incorporated, the mAP rises to 0.921. Finally, the introduction of the GI module (green solid line) leads to the highest mAP of 0.971, significantly outperforming the other combinations.

4.5. Comparison of different methods on the LiTS17 dataset

The network proposed in this section is compared with three common segmentation networks for medical images, U-Net, TransUNet, and DHT-Net, to evaluate its performance on the LiTS17 dataset. U-Net is an encoder-decoder architecture network that conveys information by skipping connections, and it has been widely used in segmentation tasks for medical images. TransUNet [46] proposes a hybrid architecture network that achieves collaborative modeling of local features and global context by merging the Transformer network with the U-Net framework. The network utilizes the Transformer encoder to reconstruct the feature maps extracted by the CNN, thereby enhancing the network's global modeling capability and improving its extraction of global contextual features. The DHT-Net network comprises a dynamic hierarchical transformer (DHTrans) and an edge aggregation block (EAB). DHTrans automatically detects the tumor location through dynamic adaptive convolution and utilizes a hierarchical algorithm with varying receptive field sizes to learn tumor features of different scales. Furthermore, EAB can effectively enhance the segmentation effect of the liver and tumor boundary area by extracting fine-grained detail information from the network's shallow features.

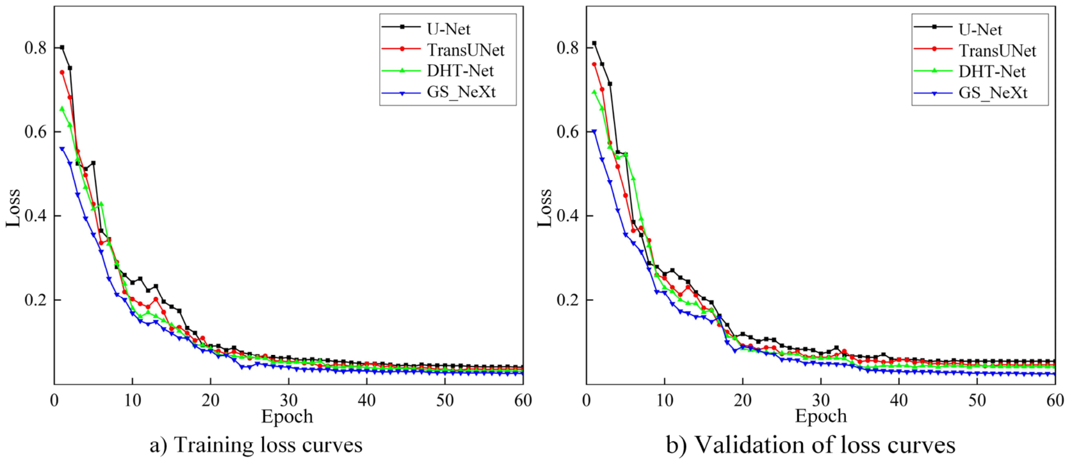
Table 5. Liver segmentation results using different networks on the LiTS17 dataset.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
U-Net	91.47 ± 1.59	9.56 ± 2.69	0.78 ± 1.45	3.81 ± 1.26	26.84 ± 8.96	8.45 ± 3.57
TransUNet	94.95 ± 1.55	7.34 ± 0.23	0.51 ± 1.79	2.62 ± 0.78	18.28 ± 6.55	5.76 ± 1.34
DHT-Net	95.23 ± 0.76	5.23 ± 1.86	0.46 ± 1.63	2.45 ± 0.29	14.55 ± 5.79	3.95 ± 1.37
GS_NeXt	**97.74 ± 0.31**	**3.68 ± 0.83**	**0.23 ± 1.01**	**1.01 ± 0.17**	**10.86 ± 1.65**	**2.04 ± 1.62**

Table 6. Tumor segmentation results using different networks on the LiTS17 dataset.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
U-Net	66.35 ± 1.86	43.95 ± 2.70	1.59 ± 1.76	8.56 ± 1.34	31.67 ± 8.45	14.34 ± 4.36
TransUNet	84.49 ± 1.38	35.48 ± 0.62	0.96 ± 1.34	6.56 ± 0.72	23.36 ± 6.87	11.55 ± 1.74
DHT-Net	86.56 ± 0.82	27.33 ± 1.43	0.68 ± 1.28	3.79 ± 0.32	17.85 ± 5.43	8.64 ± 1.38
GS_NeXt	**87.25 ± 0.23**	**22.60 ± 1.57**	**0.11 ± 1.28**	**2.23 ± 0.64**	**12.97 ± 2.35**	**5.34 ± 1.27**

Tables 5 and 6 demonstrate the segmentation results of the GS_NeXt network proposed in this section and the other three common networks on the LiTS17 dataset. As seen from the tables, the segmentation results of GS_NeXt outperform those of the other three common networks in six evaluation metrics, and the network in this section achieves a leading position in segmentation accuracy. The Dice coefficients of the proposed network reaches 97.74% and 87.25% in liver and tumor, respectively. The experimental results demonstrate that the network designed in this section performs exceptionally well on the LiTS17 dataset. A high segmentation accuracy is achieved.



a) Training loss curves

b) Validation of loss curves

Figure 11. Loss curve in networks comparison experiment for the LiTS17 dataset.

To further validate the stability of GS_NeXt, the training and validation loss curves of different networks on the LiTS17 dataset are shown in Figure 11. The loss values of the GS_NeXt network during training and validation are lower than those of the other three networks, and GS_NeXt can converge more smoothly and efficiently during training. This performance is attributed to the advantages of the graph-convolutional graph inference module in local feature transfer and information fusion, the SAM encoder's ability to understand global contextual information, and the deep supervision mechanism, which further optimizes the network's training process.

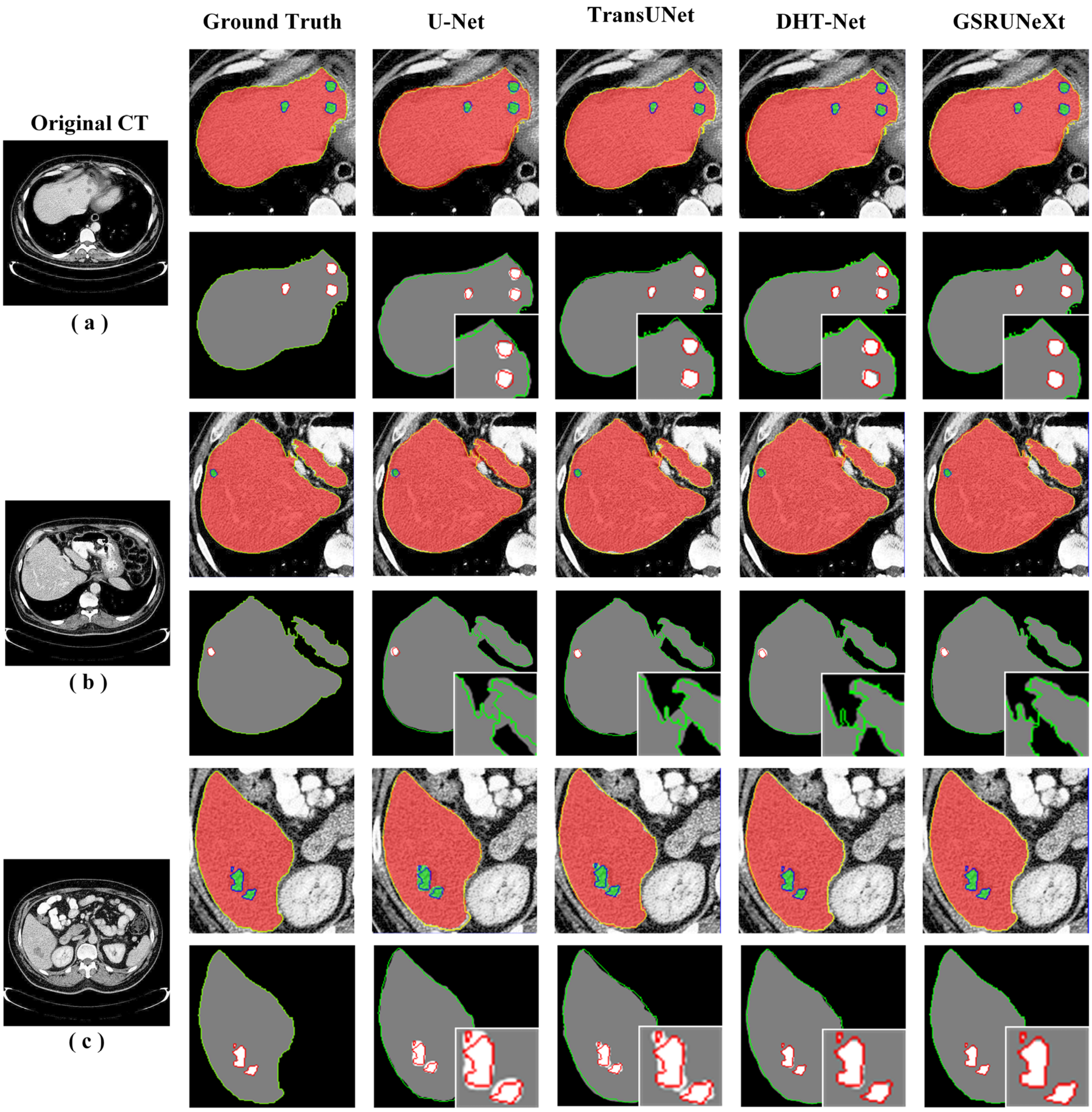
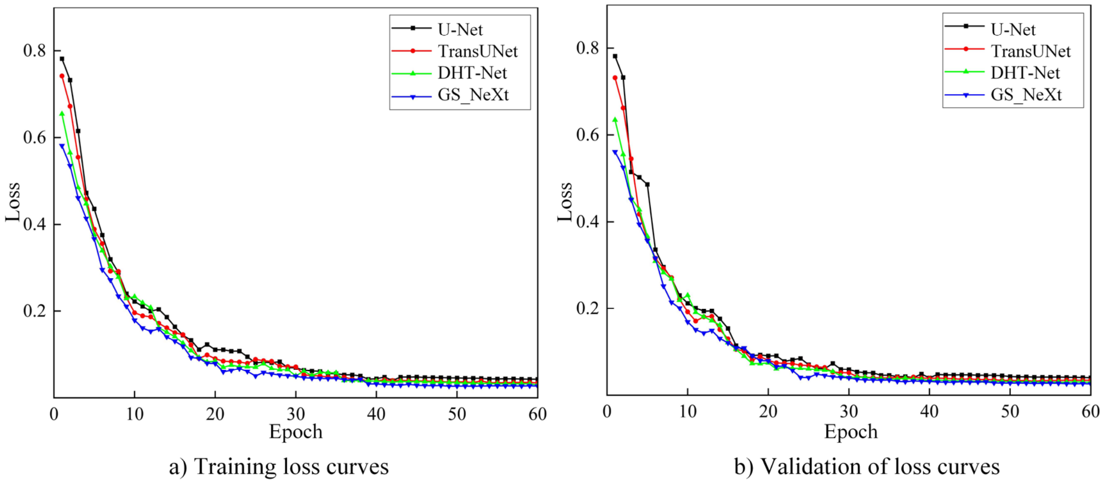


Figure 12. Visualization of liver and tumor segmentation results on the LiTS17 dataset.

Figure 12 provides the visualization results of different networks in the liver and tumor segmentation task. The figure shows three CT cases, Figure 12(a)–(c), containing liver and tumor: Case Figure 12(a) contains multiple discontinuous liver tumors, case Figure 12(b) contains microtumors and liver discontinuous regions, and case Figure 12(c) contains microtumors and tumor discontinuous regions. The visualization results indicate that U-Net exhibits low segmentation accuracy for microtumors and liver discontinuities. Although TransUNet and DHT-Net can identify the areas of microtumors and liver discontinuities, their segmentation results are not satisfactory. The segmentation is too smooth at the boundary. The segmentation results of GS_NeXt proposed in this section are closer to the gold standard in liver and tumor segmentation, particularly in segmenting complex regions containing multiple microtumors and liver discontinuities in CT images. The network in this section demonstrates higher segmentation accuracy.

4.6. Comparison of different methods for the 3DIRCADb dataset

The training and validation loss curves on the 3DIRCADb dataset are shown in Figure 13. From the figure, it can be seen that with the increase of training rounds, the training loss and validation loss of both the common networks and the network in this section show a gradual decrease, which indicates that the network gradually learns the features of the dataset and stabilizes during the training process, and that the loss of the network proposed in this section is much flatter and has the lowest loss value.



a) Training loss curves

b) Validation of loss curves

Figure 13. Loss curve in networks comparison experiment for the 3DIRCADb dataset.

Table 7. Liver segmentation results using different networks for the 3DIRCADb dataset.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
U-Net	93.25 ± 1.78	9.43 ± 1.97	0.73 ± 1.63	3.63 ± 1.21	25.34 ± 6.23	8.47 ± 2.14
TransUNet	95.27 ± 1.26	7.21 ± 0.43	0.47 ± 1.79	2.54 ± 0.73	17.36 ± 5.73	5.34 ± 1.67
DHT-Net	96.23 ± 0.76	4.83 ± 1.25	0.43 ± 1.46	2.32 ± 0.22	12.56 ± 4.75	3.87 ± 1.48
GS_NeXt	**97.31 ± 0.27**	**3.56 ± 0.67**	**0.22 ± 1.11**	**1.01 ± 0.15**	**9.76 ± 2.25**	**2.02 ± 0.81**

The segmentation results of specific performance metrics are listed in Tables 7 and 8, showing that the network in this section outperforms the other three common networks on the 3DIRCADb dataset, especially in the two key metrics of Dice coefficient and Volume Overlap Error (VOE), in which the network in this section shows a leading edge. The Dice coefficients in the liver region and liver tumor region reach 97.31% and 87.36%, respectively, which indicates that the network proposed in this section can effectively extract the details of the liver and its tumors, especially when dealing with the tiny lesions in the liver region, and the GS_NeXt proposed in this section possesses stronger robustness and higher segmentation accuracy.

Table 8. Tumor segmentation results using different networks for the 3DIRCADb dataset.

Model	Dice (%)	VOE (%)	RVD (%)	ASD (mm)	MSD (mm)	RMSD (mm)
U-Net	67.26 ± 1.84	41.23 ± 1.97	1.23 ± 1.57	7.97 ± 1.85	30.67 ± 7.24	13.52 ± 3.78
TransUNet	85.45 ± 1.37	34.56 ± 0.87	0.75 ± 1.26	6.24 ± 0.46	22.13 ± 5.27	11.23 ± 3.98
DHT-Net	86.51 ± 0.46	26.24 ± 0.45	0.61 ± 1.35	4.25 ± 0.21	**11.28 ± 3.59**	7.94 ± 1.63
GS_NeXt	**87.36 ± 0.33**	**21.56 ± 0.66**	**0.11 ± 0.48**	**2.12 ± 0.39**	12.34 ± 2.84	**5.27 ± 1.84**

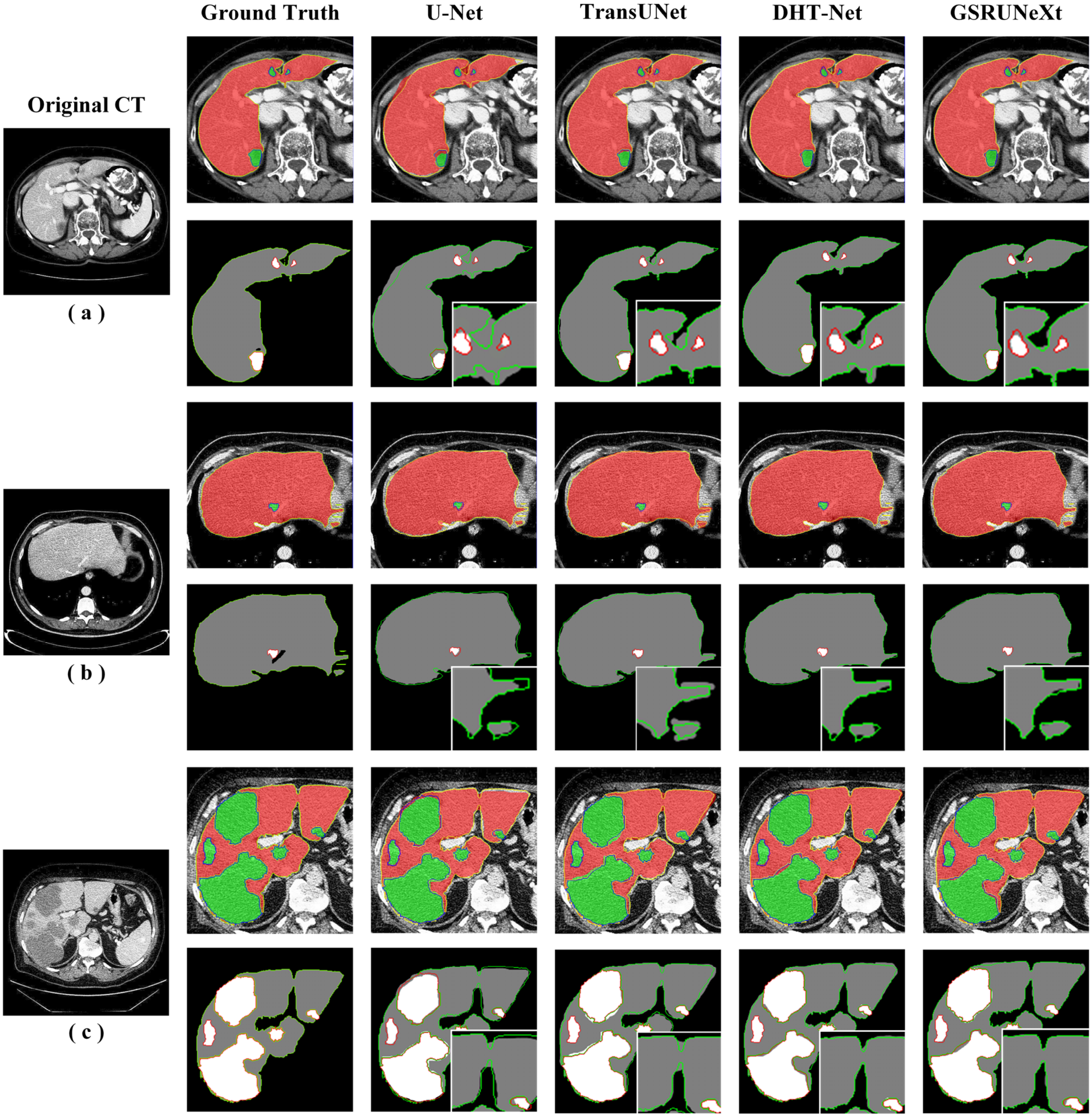


Figure14. Visualization of liver and tumor segmentation results for the 3DIRCAD dataset.

To more intuitively show the segmentation effect of the network in this section in the 3DIRCADb dataset, the visualization comparison results are shown in Figure 14, which illustrates three typical CT cases, a, b, and c, in the 3DIRCADb dataset with different characteristics: Case a contains multiple microscopic tumors, case b contains both microscopic tumors and discontinuous areas in the liver, and case c has a larger and non-complex case with continuous liver tumors. These three types of cases exhibit a high degree of complexity in the distribution of livers and tumors, making it often difficult for conventional networks to accurately segment small livers and microtumors, especially when the liver is discontinuous or the tumor boundaries are ambiguous. A comparison of these complex cases reveals that the network proposed in this section yields better segmentation results when handling discontinuities and fuzzy boundaries in liver and liver tumors. It has been proven that the network in this section also exhibits better segmentation performance across datasets.

4.7. Complexity and runtime analysis

The results of comparing the training and testing times of the GS_NeXt network proposed in this section with the Section 3 network and the other three common networks on the LiTS17 and 3DIRCADb datasets are shown in Table 9, from which it can be seen that: The training time sums of the GS_NeXt on the two datasets are 50 hr, 23 min, and 51 sec and 4 hr, 16 min, and 45 sec, respectively; and the testing times are 341 and 237 seconds for the LiTS17 and 3DIRCADb datasets, respectively, which are higher than U-Net, ResUNet and TransUNet networks, but lower than DHT-Net networks. The table compares the parameters and the number of floating-point operations (FLOPs) of the different networks, which increase the network parameters and floating-point computation due to the introduction of the SAM encoder to extract global feature information for the networks in this section. Our method, GS_NeXt, achieves a competitive FPS of 8.80, which is only slightly lower than U-Net (9.20) and higher than DHT-Net (8.55) and ResUNet (8.85) despite integrating the SAM and graph reasoning modules. This result suggests that although SAM contributes to global feature extraction, freezing its weights during training and inference significantly reduces its computational burden. Therefore, the overall inference efficiency remains acceptable.

Table 9. Parameters, FPS, FLOPs, training, and testing times for different networks.

Network	Parameters	FPS	FLOPs	LiTS17 Dataset		3DIRCADb Dataset	
				Training time	Testing time	Training time	Testing time
U-Net	**7.8 M**	**9.20**	**15.8 G**	**43h45m32s**	**326 s**	**2h41m14s**	**220 s**
TransUNet	79.9M	9.01	230.9G	56h11m26s	333 s	4h34m58s	227 s
DHT-Net	62.5M	8.55	81.3G	52h14m06s	351 s	4h37m36s	241 s
ResUNet	38.7M	8.85	67.4G	49h23m33s	339 s	3h55m11s	229 s
GS_NeXt	53.6M	8.80	71.8G	50h23m51s	341 s	4h16m45s	237 s

4.8. Comparison with advanced methods

In this section, we present a quantitative comparison between the proposed network and state-of-the-art (SOTA) methods from the past five years (2020–2025) to validate the superiority of the proposed approach. Table 10 presents the segmentation results of this section's network, comparing them with those of other state-of-the-art networks on two datasets. For the LiTS17 dataset, the network in this section achieves the optimal result in liver segmentation with a Dice value of 97.74%. In tumor segmentation, it ranks fourth, with a Dice score of 87.25%, which is 0.56% lower than the best-performing SADSNet that incorporates a spatial attention mechanism. It is also 0.27% lower than

DefED-Net, which comprises a Ladder-Atrous Spatial Pyramid Pooling (Ladder-ASPP) module, and 0.23% lower than LGMA-Net, which employs Channel-Attentive Skip Connections (CASC). SADSNet enhances multi-scale feature extraction for fine-grained structures, such as liver tumors, by introducing spatial attention mechanisms into the encoder and decoder. DefED-Net improves contextual information learning through deformable convolution and the Ladder-ASPP module. LGMA-Net incorporates an SNP-inspired convolutional Transformer block and CASC to improve segmentation of small and fine-grained structures. For the 3DIRCADb dataset, the network in this section achieves Dice values of 97.31% and 87.36%, respectively. The best performance on this dataset is obtained by FUF-TransUNet, which achieves Dice scores of 98.38% (liver) and 89.77% (tumor), outperforming the proposed method by 1.07% and 2.41%, respectively. FUF-TransUNet enhances feature interaction across distances using the Displaceable Module (DS) and improves multi-scale detail integration through the Skip-connection Feature Fusion Module (SFF). The second-best liver segmentation result is achieved by AD-DUNet, which adopts a dual-branch encoder and a residual decoder. Its Transformer encoder includes an Axial Transformer (AT) block capable of capturing long-range dependencies in medical images. GS_NeXt enhances global feature extraction by leveraging the SAM model with frozen weights, thereby providing a more comprehensive global representation. Additionally, the integration of the graph reasoning module enhances the understanding of liver-tumor boundaries and discontinuous liver regions, thereby improving segmentation accuracy. It is worth noting that while the incorporation of the SAM and graph reasoning modules contributes to higher segmentation precision, it also leads to an increase in network parameters and floating-point operations. Reducing model complexity and computational cost while maintaining segmentation accuracy will be a key focus of our future research. By comprehensively analyzing the segmentation results of the liver and its tumors, the proposed network in this section demonstrates several advantages over existing state-of-the-art networks, and its segmentation effect achieves a level comparable to that of advanced networks.

Table 10. Comparison of Dice score (%) with the SOTA networks.

Network	Year	LiTS17 Dataset		3DIRCADb Dataset	
		Liver	Tumor	Liver	Tumor
RA-UNet [47]	2020	94.77	71.38	95.32	72.28
ASU-Net [48]	2021	95.13	72.65	95.42	70.26
Swin-Unet [49]	2022	94.59	81.50	95.21	82.07
DefED-Net [50]	2022	96.30	87.52	96.60	66.25
MDCF_Net [29]	2023	96.60	77.10	96.90	72.70
RMAU-Net [51]	2023	95.52	76.16	96.97	83.07
AD-DUNet [52]	2024	97.18	80.75	97.43	86.89
SADSNet [53]	2024	97.03	**87.81**	96.11	87.50
FUF-TransUNet [54]	2025	97.01	83.41	**98.38**	**89.77**
LGMA-Net [55]	2025	97.72	87.48	97.20	83.24
GS_NeXt	2025	**97.74**	87.25	97.31	87.36

To demonstrate the performance of our designed GS_NeXt model in the segmentation task more intuitively, we present the results of 3D error visualization on two different datasets in Figure 15. The red and blue regions indicate significant over-segmentation and under-segmentation errors, respectively. In contrast, the green regions represent areas consistent with the ground truth. As observed from the visualizations, in the segmentation of liver and tumor regions, most of the areas of the 3D error map generated by the proposed method are covered by a green surface, indicating that it is highly consistent with the real segmentation (i.e., the ground truth), thus further validating the validity and accuracy of the model.

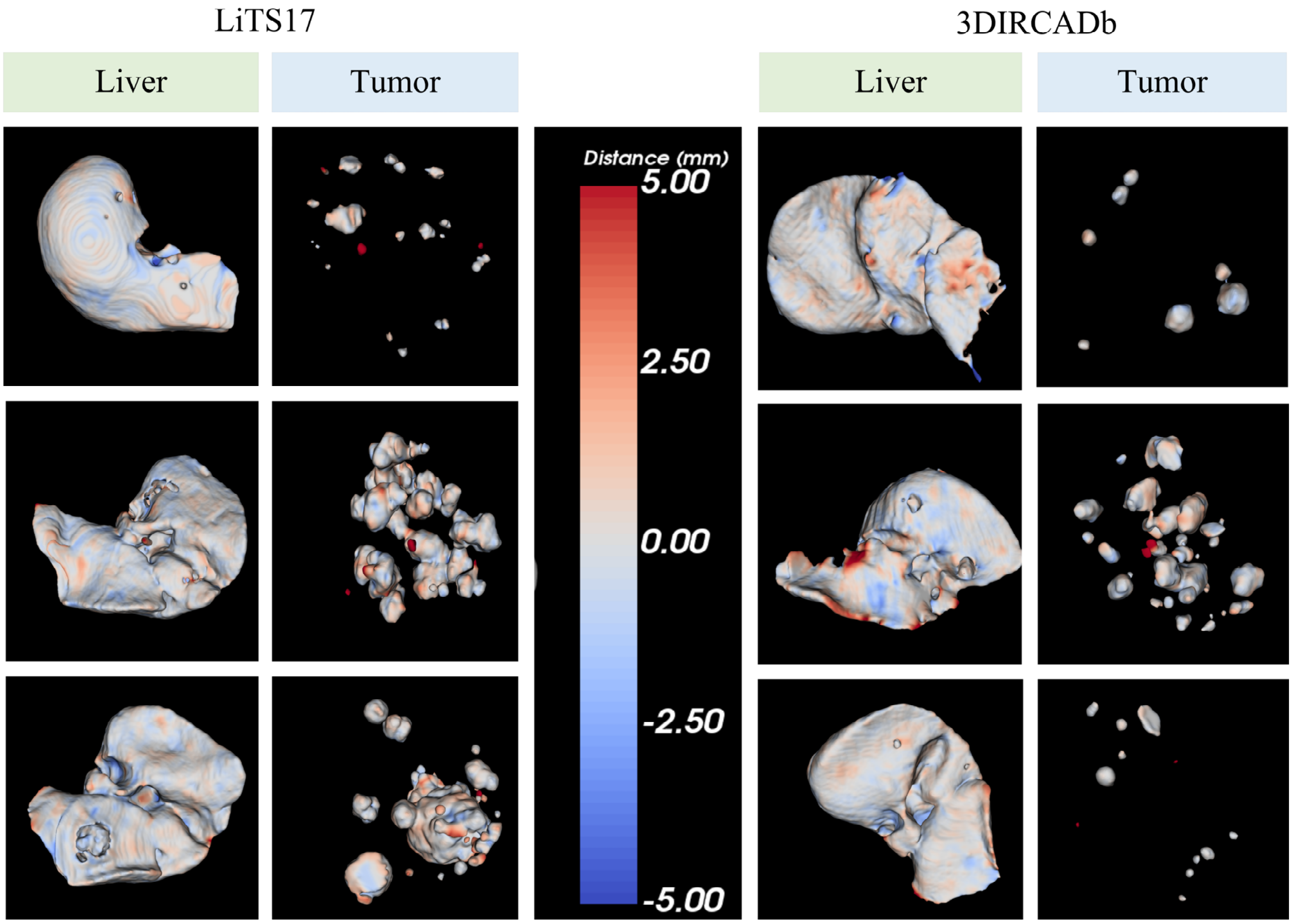


Figure 15. 3D errors of the proposed GS_NeXt on the LiTS17 and 3DIRCADb datasets.

4.9. Noise robustness experiment

To evaluate the robustness of our proposed GS-NeXt network under clinically challenging scenarios, we conduct additional experiments for the LiTS17 dataset by introducing varying levels of Poisson noise to the input CT images. This simulates the noise commonly encountered in low-dose imaging, motion artifacts, or equipment-induced degradation.

We define three levels of Poisson noise (Levels 1–3), with increasing intensity corresponding to more severe image degradation. The segmentation model was trained on the clean (original) dataset and directly applied to the noisy versions without any fine-tuning, thereby testing the model's inherent generalization ability.

As shown in Figure 16, the first column displays the ground truth (GT), and the second column shows the segmentation result from GS-NeXt on clean input. The remaining columns illustrate the segmentation performance under Poisson noise levels 1 to 3. Even under visible degradation in image quality (e.g., noise artifacts in soft tissue regions), the segmentation boundaries of liver and tumor remain largely consistent, especially for Levels 1 and 2. At Level 3, although the background noise increases significantly, GS-NeXt manages to localize the target region effectively, showing its strong resistance to noise perturbations.

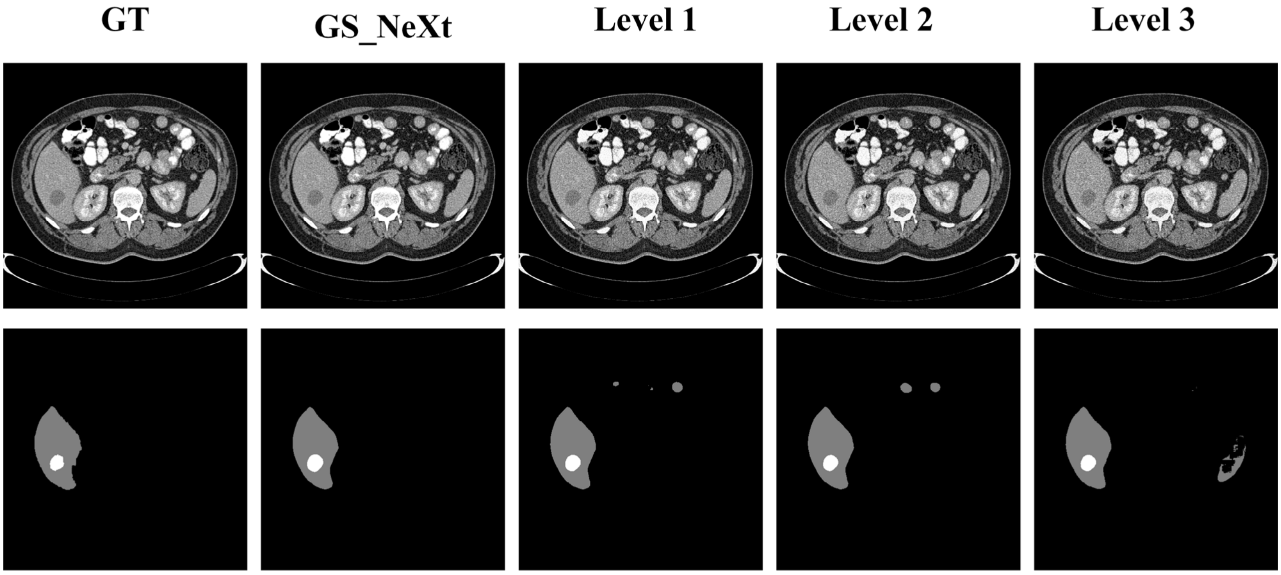


Figure 16. Robustness experiments under different levels of Poisson noise for the LiTS17 dataset.

As shown in Tables 11 and 12, the Dice similarity coefficient experiences only a minor decline as noise intensity increases. On average, the liver Dice score decreases from 0.9774 (clean data) to 0.9260 (Level 3 noise), while the tumor Dice score decreases from 0.8725 (clean data) to 0.7983 (Level 3 noise). This indicates that the model maintains high accuracy under high noise conditions, and the incorporation of global-local feature modeling and graph reasoning structures effectively enhances the model's robustness to image noise.

Table 11. Liver Dice scores (%) based on the LiTS17 dataset.

Dice score	Without noise	Level 1	Level 2	Level 3
Poisson	97.74	95.03	94.26	92.60

Table 12. Tumor Dice scores (%) based on the LiTS17 dataset.

Dice score	Without noise	Level 1	Level 2	Level 3
Poisson	87.25	83.03	81.26	79.83

5. Conclusions

An automatic liver and tumor segmentation network based on graph convolution, the graph inference algorithm, and the SAM network is studied (GS_NeXt). The contribution of this paper are mostly manifested in four aspects: First, we use graph convolution in the middle of the encoder and decoder of the network to perform graph inference on the features extracted by the encoder as a way to capture the dependency relationship between global and local features, and to improve the ability of the network to extract unstructured features; second, we design a two-branch encoder structure, using the encoder of the SAM network as the lower branch of the GS_NeXt network encoder, retaining its ability to extract global feature information by freezing the weights of the SAM, thus extracting more comprehensive feature information of the liver and tumor; and third, we use the feature extraction module of ConvNeXt-v2 to construct the upper-branch ConvNeXt encoder within the network. The global response normalization layer within this encoder suppresses extraneous noises in the feature information, enabling the network to focus on important liver and tumor information. Fourth, a deep supervision mechanism is employed in the network's decoder to facilitate faster convergence.

The proposed network is experimentally validated for LiTS17 and 3DIRCADb datasets. The Dice values in the liver segmentation results are 97.74% and 97.31%, which are 2.79% and 2.04% higher than those of TransUNet, respectively. The Dice values in the tumor segmentation results are 87.25% and 87.36%, which are 2.76% and 1.91% higher than those of TransUNet, respectively. The best scores are also achieved in the four metrics of VOE, RVD, ASD, and RMSD. In terms of the segmentation effect of liver discontinuity and microtumor regions, the network demonstrates a strong segmentation ability, effectively improving the segmentation accuracy of these regions.

Here, we employ a deep learning approach to introduce modern techniques, including full-dimensional dynamic convolution, iterative attentional feature fusion, graph convolution, a graph inference algorithm, and a SAM network. The segmentation network has been modernized and improved to effectively enhance the segmentation of liver and tumor boundaries, as well as to reduce blurring, discontinuity, and micro-tumor regions in abdominal CT images. However, the proposed method segments the 2D CT image as input. Although it has achieved a relatively ideal segmentation effect, the irrelevant areas outside the liver organ also occupy a significant amount of system resources during training. In the future, an auxiliary network will be designed to locate the liver in the input CT image and perform image cropping to improve the utilization of system resources. Moreover, we will explore lightweight variants of SAM or adopt more compact state-space models (e.g., FlashMamba) to reduce latency further while maintaining segmentation accuracy.

Data availability

The LiTS17 and 3DIRCADb datasets used in this paper are publicly available as follow s: http://www.lits-challenge.com and https://www.ircad.fr/research/data-sets/liver-segmentation-3d-ircadb-01/.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Natural Science Foundation of Heilongjiang Province of China under Grant PL2024F010.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. Siegel, I. Soerjomataram, et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: Cancer J. Clin.*, **74** (2024), 229–263. https://doi.org/10.3322/caac.21834
2. R. S. Wang, T. Lei, R. X. Cui, B. T. Zhang, H. Y. Meng, A. K. Nandi, Medical image segmentation using deep learning: A survey, *IET Image Proc.*, **16** (2022), 1243–1267. https://doi.org/10.1049/ipr2.12419
3. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
4. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2015*, Springer, **9351** (2015), 234–241. https://doi.org/10.1109/CVPR.2015.7298965
5. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
6. V. Rajinikanth, S. Kadry, R. Damaševičius, D. Sankaran, M. A. Mohammed, S. Chander, et al., Skin melanoma segmentation using VGG-UNet with Adam/SGD optimizer: a study, in *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*. IEEE, (2022), 982–986. https://doi.org/10.1109/ICICICT54557.2022.9917848
7. K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2016), 770–778. https://doi.org/10.1109/CVPR.2016.90
8. F. I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.*, **162** (2020), 94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013
9. Z. Liu, H. Z. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, S. N. Xie, A convnet for the 2020s, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2022), 11976–11986. https://doi.org/10.1109/CVPR52688.2022.01167
10. Z. Z. Han, M. W. Jian, G. G. Wang, ConvUNeXt: An efficient convolution neural network for medical image segmentation, *Knowl.-Based Syst.*, **253** (2022), 109512. https://doi.org/10.1016/j.knosys.2022.109512
11. M. M. Rahman, M. Munir, R. Marculescu, Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2024), 11769–11779. https://doi.org/10.1109/CVPR52733.2024.01118
12. C. Li, A. J. Zhou, A. B. Yao, Omni-dimensional dynamic convolution, preprint, arXiv:2209.07947. https://doi.org/10.48550/arXiv:2209.07947

13. Z. H. Xing, T. Ye, Y. J. Yang, G. L, L. Zhu, Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2024*, Springer, (2024), 578–588. https://doi.org/10.1007/978-3-031-72111-3_54
14. Y. H. Fu, J. F. Liu, J. Shi, TSCA-Net: Transformer based spatial-channel attention segmentation network for medical images, *Comput. Biol. Med.*, **170** (2024), 107938. https://doi.org/10.1016/j.compbiomed.2024.107938
15. F. Lyu, A. J. Ma, T. C. F. Yip, G. L. H. Wang, P. C. Yuen, Weakly supervised liver tumor segmentation using couinaud segment annotation, *IEEE Trans. Med. Imaging*, **41** (2021), 1138–1149. https://doi.org/10.1109/TMI.2021.3132905
16. L. Meng, Y. Y. Tian, S. H. Bu, Liver tumor segmentation based on 3D convolutional neural network with dual scale, *J. Appl. Clin. Med. Phys.*, **21** (2020), 144–157. https://doi.org/10.1002/acm2.12784
17. R. Y. Li, L. Ch. Xu, K. Xie, J. F. Song, X. W. Ma, L. A. Chang, Dht-net: Dynamic hierarchical transformer network for liver and tumor segmentation, *IEEE J. Biomed. Health. Inf.*, **27** (2023), 3443–3454. https://doi.org/10.1109/JBHI.2023.3268218
18. Q. Li, H. Song, Z. H. Wei, F. B. Yang, J. F. Fan, D. N. Ai, Densely connected u-net with criss-cross attention for automatic liver tumor segmentation in ct images, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **20** (2022), 3399–3410. https://doi.org/10.1109/TCBB.2022.3198425
19. J. N. Chi, X. Y. Han, C. D. Wu, H. Wang, P. Ji, X-Net: Multi-branch UNet-like network for liver and tumor segmentation from 3D abdominal CT scans, *Neurocomputing*, **459** (2021), 81–96. https://doi.org/10.1016/j.neucom.2021.06.021
20. A. Gu, T Dao, Mamba: Linear-time sequence modeling with selective state spaces, preprint, arXiv:2312.00752.https://doi.org/10.48550/arXiv:2312.00752
21. A. Gu, K. Goel, C. Ré, Efficiently modeling long sequences with structured state spaces, preprint, arXiv:2111.00396.https://doi.org/10.48550/arXiv:2111.00396
22. Z. Zhu, Z. Wang, G. Qi, Y. X. Zhao, Y. Liu, Visually Stabilized Mamba U-shaped network with strong inductive bias for 3D brain tumor segmentation, *IEEE Trans. Instrum. Meas.*, **74** (2025), 2518511. https://doi.org/10.1109/TIM.2025.3551581
23. Y. B. Tang, Y. X. Tang, Y. Y. Zhu, J. Xiao, R. M. Summers, E 2 Net: An edge enhanced network for accurate liver and tumor segmentation on CT scans, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2020*, Springer, **12264** (2020), 512–522. https://doi.org/10.1007/978-3-030-59719-1_50
24. H. Seo, C. Huang, M. Bassenne, R. X. Xiao, L. Xing, Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images, *IEEE Trans. Med. Imaging*, **39** (2019), 1316–1325. https://doi.org/10.1109/TMI.2019.2948320
25. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, Curran Associates, Inc., **30** (2017), 1–11. https://doi.org/10.5555/3295222.3295349
26. Y. F. Ni, G. Chen, Z. Feng , H. Cui, D. Metaxas, S. T. Zhang, et al., DA-Tran: Multiphase liver tumor segmentation with a domain-adaptive transformer network, *Pattern Recognit.*, **149** (2024), 110233. https://doi.org/10.1016/j.patcog.2023.110233

27. J. T. Hu, S. Y. Chen, Z. Y. Pan, S. Zen, W. M. Yang, Perspective+ unet: Enhancing segmentation with bi-path fusion and efficient non-local attention for superior receptive fields, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2024*, Springer, **15009** (2024), 499–509. https://doi.org/10.1007/978-3-031-72114-4_48
28. P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, M. Armbruster, et al., Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2016*, Springer, **9901** (2016), 415–423. https://doi.org/10.1007/978-3-319-46723-8_48
29. J. Jiang, Y. J. Peng, Q. F. Hou, J. Wang, MDCF_Net: A Multi-dimensional hybrid network for liver and tumor segmentation from CT, *Biocybern. Biomed. Eng.*, **43** (2023), 494–506. https://doi.org/10.1016/j.bbe.2023.04.004
30. K. N. Wang, S. X. Li, Z. Y. Bu, F. X. Zhao, G. Q. Zhou, S. J. Zhou, SBCNet: Scale and boundary context attention dual-branch network for liver tumor segmentation, *IEEE J. Biomed. Health. Inf.*, **28** (2024), 2854–2865. https://doi.org/10.1109/JBHI.2024.3370864
31. Z. Q. Zhu, Z. M. Zhang, G. Q. Qi, Y. Y. Li, Y. Z. Li, L. Mu, A dual-branch network for ultrasound image segmentation, *Biomed. Signal Process. Control*, **103** (2025), 107368. https://doi.org/10.1016/j.bspc.2024.107368
32. Y. Kaya, E. Akat, S. Yıldırım, Fusion-Brain-Net: A Novel Deep Fusion Model for Brain Tumor Classification, *Brain Behav.*, 15(2025), e70520. https://doi.org/10.1002/brb3.70520
33. S. H. Di, Y. Q. Zhao, M. Liao, F. Zhang, X. Li, TD-Net: A hybrid end-to-end network for automatic liver tumor segmentation from CT images, *IEEE J. Biomed. Health. Inf.*, **27** (2022), 1163–1172. https://doi.org/10.1109/JBHI.2022.3181974
34. A. Kirillov, E. Mintun, N. Ravi, H. Z. Mao, C. Rolland, L. Gustafson, et al., Segment anything, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2023), 4015–4026. https://doi.org/10.1109/ICCV51070.2023.00371
35. Y. C. Zhang, Z. R. Shen, R. S. Jiao, Segment anything model for medical image segmentation: Current applications and future directions, *Comput. Biol. Med.*, **171** (2024), 108238. https://doi.org/10.1016/j.compbiomed.2024.108238
36. C. Qin, J. L. Cao, H. Z. Fu, F. S. Khan, R. M. Anwer, Db-sam: Delving into high quality universal medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2024*, Springer, **15012** (2024), 498–508. https://doi.org/10.1007/978-3-031-72390-2_47
37. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2016), 2818–2826. https://doi.org/10.1109/CVPR.2016.308
38. N. Ibtehaz, M. S. Rahman, MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation, *Neural Networks*, **121** (2020), 74–87. https://doi.org/10.1016/j.neunet.2019.08.025
39. H. P. Kuang, X. Yang, H. J. Li, J. W. Wei, L. H. Zhang, Adaptive multiphase liver tumor segmentation with multiscale supervision, *IEEE Signal Process Lett.*, **31** (2024), 426–430. https://doi.org/10.1109/LSP.2024.3356414
40. Z. Liu, Q. Y. Teng, Y. Q. Song, W. Hao, Y. Liu, Y. Zhu, et al., HI-Net: Liver vessel segmentation with hierarchical inter-scale multi-scale feature fusion, *Biomed. Signal Process. Control*, **96** (2024), 106604. https://doi.org/10.1016/j.bspc.2024.106604

41. F. Zhan, W. W. Wang, Q. Chen, Y. N. Guo, L. D. He, L. L. Wang, Three-direction fusion for accurate volumetric liver and tumor segmentation, *IEEE J. Biomed. Health. Inf.*, **28** (2023), 2175–2186. https://doi.org/10.1109/JBHI.2023.3344392
42. Y. D. Meng, H. R. Zhang, D. X. Gao, Y. T. Zhao, X. Y. Yang, X. S. Qian, et al., BI-GCN: Boundary-aware input-dependent graph convolution network for biomedical image segmentation, preprint, arXiv:2110.14775.https://doi.org/10.48550/arXiv:2110.14775
43. P. B. Weerakody, K. W. Wong, G. J. Wang, W. Ela, A review of irregular time series data handling with gated recurrent neural networks, *Neurocomputing*, **441** (2021), 161–178. https://doi.org/10.1016/j.neucom.2021.02.046
44. S. Woo, S. Debnath, R. H. Hu, X. L. Chen, Z. Liu, I. S. Kweon, Convnext v2: Co-designing and scaling convnets with masked autoencoders, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, (2023), 16133–16142. https://doi.org/10.1109/CVPR52729.2023.01548
45. Q. Dou, L. Q. Yu, H. Chen, Y. M. Jin, X. Yang, J. Qin, et al., 3D deeply supervised network for automated segmentation of volumetric medical images, *Med. Image Anal.*, **41** (2017), 40–54. https://doi.org/10.1016/j.media.2017.05.001
46. J. N. Chen, Y. Y. Lu, Q. H. Yu, X. D. Luo, E. Adeli, Y. Wang, et al., Transunet: Transformers make strong encoders for medical image segmentation, preprint, arXiv:2102.04306. https://doi.org/10.48550/arXiv:2102.04306
47. Q. G. Jin, Z. P. Meng, C. M. Sun, H. Cui, R. Su, RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans, *Front. Bioeng. Biotechnol.*, **8** (2020), 605132. https://doi.org/10.3389/fbioe.2020.605132
48. Q. H. Gao, M. Almekkawy, ASU-Net++: A nested U-Net with adaptive feature extractions for liver tumor segmentation, *Comput. Biol. Med.*, **136** (2021), 104688. https://doi.org/10.1016/j.compbiomed.2021.104688
49. H. Cao, Y. Y. Wang, J. Chen, D. S. Jiang, X. P. Zhang, Q. Tian, et al., Swin-unet: Unet-like pure transformer for medical image segmentation, in *European Conference on Computer Vision*, Springer, Cham, **13803** (2022), 205–218. https://doi.org/10.1007/978-3-031-25066-8_9
50. T. Lei, R. S. Wang, Y. X. Zhang, Y. Wan, C. Liu, A. K. Nandi, et al., DefED-Net: Deformable encoder-decoder network for liver and liver tumor segmentation, *IEEE Trans. Radiat. Plasma Med. Sci.*, **6** (2021), 68–78. https://doi.org/10.1109/TRPMS.2021.3059780
51. L. F. Jiang, J. J. Ou, R. H. Liu, Y. Y. Zou, T. Xie, H. G. Xiao, et al., Rmau-net: Residual multi-scale attention u-net for liver and tumor segmentation in ct images, *Comput. Biol. Med.*, **158** (2023), 106838. https://doi.org/10.1016/j.compbiomed.2023.106838
52. H. Qi, W. J. Wang, Y. T. Shi, X. H. Wang, AD-DUNet: A dual-branch encoder approach by combining axial Transformer with cascaded dilated convolutions for liver and hepatic tumor segmentation, *Biomed. Signal Process.*, **95** (2024), 106397. https://doi.org/10.1016/j.bspc.2024.106397
53. S. J. Yang, Y. B. Liang, S. Wu, P. Sun, Z. C. Chen, SADSNet: A robust 3D synchronous segmentation network for liver and liver tumors based on spatial attention mechanism and deep supervision, *J. X-Ray Sci. Technol.*, **32** (2024), 707–723. https://doi.org/10.3233/XST-230312

54. H. Y. Ma, M. Maimaiti, FUF-TransUNet: A Transformer-Based U-Net with fully utilize of features for liver and liver-tumor segmentation in CT images, in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Springer Nature, Singapore, (2024), 34–47. https://doi.org/10.1007/978-981-97-8499-8_3
55. W. J. Ren, B. Li, H. Peng, J. Wang, Lgma-net: liver and tumor segmentation methods based on local–global feature mergence and attention mechanisms, *Signal, Image Video Process.*, **19** (2025), 43. https://doi.org/10.1007/s11760-024-03731-y