School of Computing
FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

UNIVERSITY OF LEEDS

LiTS Insight: Deep Learning Approaches to Tumor Segmentation

Jun Gao

Submitted in accordance with the requirements for the degree of
MSc Advanced Computer Science

2023/2024

The candidate confirms that the following have been submitted*:*

Items	**Format**	**Recipient(s) and Date**
Deliverables 1, 2, 3	*Report*	*SSO (12/08/24)*
Participant consent forms	*Signed forms in envelop*	*SSO (12/08/24)*
Deliverable 4	*https://github.com/junver10086/MSC_ComputerScienceProject.git*	*Supervisor, assessor (12/08/24)*

Type of Project: **Interdisciplinary Project**

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

(Signature of student) Jun Gao

Summary

This paper aims to address the major challenge of liver lesion segmentation in clinical diagnosis and treatment. Accurate segmentation of liver lesions is crucial to improving diagnostic accuracy and formulating effective treatment plans. However, traditional imaging diagnostic methods mainly rely on the experience and subjective judgment of radiologists, which is not only time-consuming but also easily affected by human factors, leading to misdiagnosis or missed diagnosis.[1][2]. With the advancement of medical imaging technic, the amount of image data has increased rapidly, further increasing the workload of doctors. As a result, developing an efficient and accurate automated method for liver lesion segmentation is highly significant.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Nishant Ravikumar, for his patient and dedicated one-on-one guidance throughout the course of this project. His invaluable insights and constant support have been instrumental in shaping the outcome of this work.

I am also deeply thankful to my inspector, Marc de Kamps, for his unwavering support and companionship during the project. His presence and encouragement have been greatly appreciated.

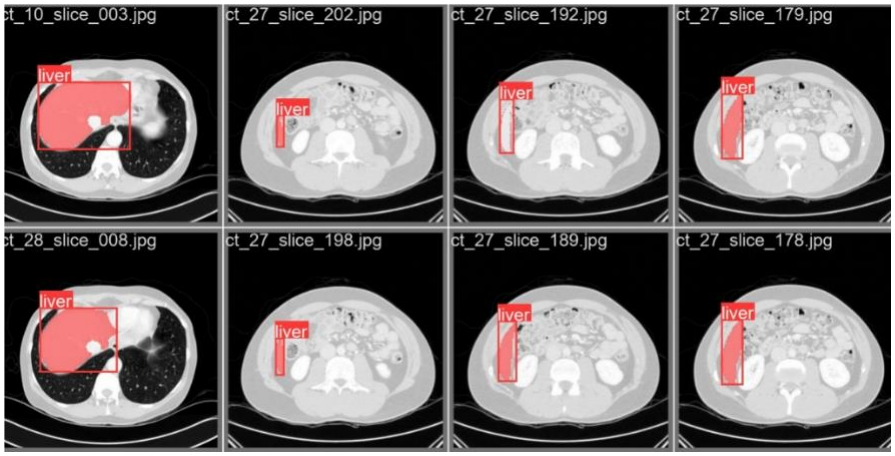
Table of Contents

Abstract

In recent years, deep learning technic, especially in The convolutional neural networks (CNNs have made significant progress in the field of medical image analysis}. UNet, a well-established CNN architecture, has gained widespread attention and application in medical image segmentation tasks due to its effective capability to capture multi-scale information[3][4][5]. Therefore, UNet was selected as a research tool for liver lesion segmentation to improve segmentation accuracy and provide a more reliable auxiliary tool for clinical diagnosis [6][7].

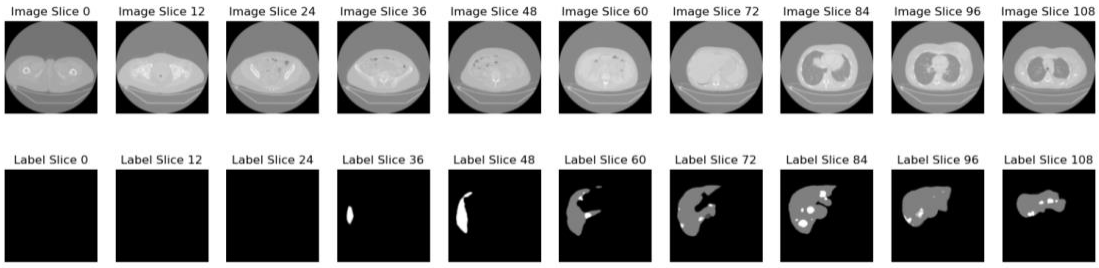
This study is based on The data from two challenges, LITS (Liver Tumor Segmentation) [8] and CHAOS (Combined (CT-MR) Healthy Abdominal Organ Segmentation) [9], were used to explore the optimal liver lesion segmentation method in CT scan images. The LITS challenge was jointly organized by the IEEE International Symposium on Biomedical Imaging (ISBI 2017) and the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017 and 2018), while the CHAOS challenge was organized by ISBI 2019 and focused on the segmentation of abdominal organs. The ultimate goal of this study is to improve the segmentation of liver lesions by The segmentation accuracy can provide a more effective auxiliary tool for clinical diagnosis.

Graph Abstract



Graph(1)

picture1Representative used in the healthy liver dataset CHAOSYOLOv7_segLiver part segmentation mask obtained by the algorithm.



Graph (2)

Figure (2) IImage Slicerepresents a CT slice directly obtained from the LITS dataset containing liver lesions, Lable SliceRepresents the corresponding segmentation mask, where 0 (black part) represents the background, 1 (gray part) represents the liver, and 2 (white part) represents the lesion area.

Keywords

Liver lesion, Liver Tumor,Cancer,UNet,UNet+++,AttentionUNet,RestUNet,YOLOV7seg,Semantic segmentation, IoU, Dice Coefficient

Chapter 1
Introduction

1.1 Project Aim

The main goal of this entire project is to find and evaluate one-stage and two-stage segmentation methods using YOLOv7_segAlgorithms andCycleGAN and 3D UNet were tested on the LITS dataset, and U-Net, U-Net++, and Attention U-Net were used for comparative experiments on the CHAOS and LITS datasets. The training results of each model were evaluated using evaluation parameters such as Dice score and mean IoU, and the optimal segmentation method was finally selected.,Used for automated segmentation of the liver and liver lesions. The specific goals are as follows:

1. Improve segmentation accuracy:
    - Screening for the most suitable segmentation algorithm: To design and train a deep learning model that can accurately segment the liver and its lesion areas, ensuring its excellent performance in different types of liver lesions.
    - Optimizing model parameters and architecture: By adjusting the model's hyperparameters, architecture design, and training strategies, the segmentation accuracy can be further optimized to minimize the occurrence of misdiagnosis and missed diagnosis.
    - Comprehensive evaluation of model performance: Use evaluation indicators such as Dice coefficient and IoU to comprehensively evaluate the model performance to ensure that the model has high accuracy and robustness in various segmentation tasks.
2. Comparing multiple deep learning models:
    - Experimental comparison of the Dice coefficient of different models: Experiments were conducted on the LITS dataset to compare the performance of different deep learning models (such as CycleGAN, YOLOv7-seg, and 3D UNet) in liver and lesion segmentation.
    - Comparison of one-stage and two-stage algorithms: Compare the performance differences in segmentation tasks between the one-stage (such as YOLOv7-seg) and the two-stage (after completing the semantic segmentation of the liver part, only retain the pixels under the liver part and then perform semantic segmentation of the liver lesion part) algorithms, and analyze the advantages and disadvantages of different methods and their applicable scenarios.
3. Comparison of the performance of various improved versions of UNet in semantic segmentation:
    - Detailed comparison of UNet and its improved versions: Experiments were conducted on the CHAOS dataset and the LITS dataset, respectively, to compare in detail the performance of UNet and its various improved versions (such as UNet+++, Attention-UNet, Res-UNet++) in liver and liver lesion segmentation.
    - Evaluating and analyzing model differences: Evaluate the performance differences of these models in segmentation tasks, find the most suitable segmentation method, and analyze the potential causes for the varying outcomes, such as network architecture, introduction of attention mechanism, etc.
4. Verify the generalization ability of the model:

- Cross-validation model performance: Use different datasets (such as LITS and CHAOS) for cross-validation to test the generalization ability of the model on the different datasets and ensure that the model can maintain high segmentation accuracy when processing data from different sources and of different quality.
- Analyze the outcomes of the model on different patient data: Detailed analysis of the model's performance on different patient data to ensure its reliability and stability in actual clinical applications.

5. Provide ideas for improvement:
   - Identify and analyze model deficiencies: During the research process, the shortcomings and challenges of the current UNet model in liver lesion segmentation were identified and analyzed, such as handling complex backgrounds and blurred lesion edges.
   - Propose further improvements: Based on the research results, suggestions are put forward to further improve the effectiveness and adaptability of the model, providing guidance for future research and applications, such as introducing new loss functions, optimizing training strategies, etc.

By accomplishing the aforementioned research objectives, we aim to substantially enhance the accuracy and efficiency of liver lesion segmentation models, reduce the workload of doctors, and provide strong technical support for clinical diagnosis and treatment. The results of this study will not only help the automated segmentation of liver lesions, but will also provide useful references for other medical image segmentation tasks.

1.2 Objective

In order to achieve the research objectives of this project, the following specific goals must be met:

1. Dataset collection and preprocessing: Collect and preprocess LITS and CHAOS datasets suitable for deep learning model training to ensure data quality and diversity.

2. Model implementation and optimization: Implement and optimize the U-Net-based deep learning model, solve the overfitting and underfitting problems by adjusting hyperparameters and training strategies, and enhance the segmentation accuracy of the model.

3. Comparison and selection of model frameworks: Compare different modeling frameworks (such as CycleGAN, YOLOv7-seg, 3D UNet) and choose the model combination that demonstrates the optimal performance.

4. Verification of the generalization performance of the model: Through the cross-validation method, the generalization ability of the model on different data sets is tested to ensure its stability and reliability in practical applications.

5. Suggestions for improvement and future research directions: Identify and analyze the shortcomings of the current model, propose suggestions for further improvement, and provide guidance for future research.

1.3 Deliverables

Based on the above research objectives, the following are the expected deliverables of this project:

1. Trained and validated deep learning model: Provides a deep learning model that can accurately identify and segment the liver and its lesions, which has been fully trained and validated.

2. Detailed technical report: Write a detailed technical report, including model design, training process, dataset used, parameter adjustment and verification results, etc., recording each step of the project and the theoretical basis behind it.

3. Performance evaluation report: Complete an evaluation report to show the performance of the model in different data sets, including accuracy and dice coefficient, evaluate the practical application value of the model and explore possible improvement directions.

By completing these deliverables, the project will provide important technical support for clinical diagnosis and treatment, while laying the foundation for subsequent research.

Chapter 2
Literature review

2.1 Research Motivation

Liver lesions refer to abnormal tissues or masses that appear in the liver, including both benign lesions such as hepatic hemangiomas and liver cysts, and malignant lesions such as hepatocellular carcinoma (HCC) and liver metastasis [10]. As the largest solid organ in the human body, the liver undertakes important functions such as metabolism and detoxification, and its health directly affects overall health [11]. However, due to factors such as viral hepatitis and alcohol abuse, the incidence of liver lesions has increased year by year, especially the early detection of HCC and metastatic liver cancer is crucial [12]. Although traditional imaging examinations such as ultrasound, CT and MRI can provide high-resolution images, their interpretation depends on the experience of radiologists, which is prone to misdiagnosis or missed diagnosis due to subjective judgment, and the burden on doctors is increased when faced with a large amount of image data [13]. Consequently, the development of efficient and accurate automated liver lesion segmentation methods has emerged as a crucial approach to addressing this issue [14].

With the development of computer technology and artificial intelligence, automated medical image analysis has become a feasible solution to the above problems. Deep learning, especially for convolutional neural networks (CNNs), has shown powerful capabilities in image processing and analysis [15]. In recent years, UNet, a deep learning network structure designed specifically for biomedical image segmentation, has attracted great attention and application from researchers due to its excellent performance and wide applicability [16]. UNet, through its symmetrical encoder-decoder structure and jump connections, can not only effectively capture multi-scale image features, but also accurately locate the target area while maintaining high resolution [17].

The choice of using UNet for instance segmentation of liver lesions is based on the following main motivations:

1. **Improve segmentation accuracy**：UNet has performed well in many medical image segmentation tasks. Its structure can effectively handle complex image backgrounds and accurately segment the target area [18]. In liver lesion segmentation, high-precision segmentation results can significantly improve the detection rate of lesions and reduce misdiagnosis and missed diagnosis [19].
2. **Reducing the burden on doctors**: Automated segmentation approach can significantly reduce the time and energy doctors spend on image interpretation and lesion labeling, allowing them to focus on more complex diagnostic tasks that require more professional judgment [20].

3. **Promoting personalized medicine**: Accurate lesion segmentation results are not only important for diagnosis, but also provide a reliable basis for the formulation of personalized treatment plans [21]. For example, in the treatment of liver cancer, accurate tumor localization and volume measurement are important bases for surgical planning and efficacy evaluation [22]. Through automated segmentation methods, these goals can be better achieved and the development of personalized medicine can be promoted [23].
4. **Driving research progress**: Although UNet has achieved some success in medical image segmentation, there are still many challenges and room for improvement in specific applications [24]. Through application research in liver lesion segmentation, the UNet model can be further optimized and improved, its potential in processing complex medical image problems can be explored, and reference and reference can be provided for other related research [25].

In summary, using UNet for instance segmentation of liver lesions not only helps to improve the efficiency and accuracy of diagnosis, but also provides strong support for personalized medicine and future research [26]. These motivations drove us to explore and apply the UNet model in this study in order to provide new solutions and ideas for the automated segmentation menthod of liver lesions [27].

2.2 Related Work on Computer Vision

Traditional methods for diagnosing liver lesions mainly rely on imaging techniques such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) [28]. Although these imaging techniques can provide detailed liver images to help doctors identify and evaluate lesions, they still have certain limitations in clinical practice. For example, the diagnostic results are highly dependent on the doctor's experience and subjective judgment, which can easily lead to misdiagnosis or missed diagnosis [29]. In addition, with the rapid growth of medical imaging data, doctors face huge challenges in processing and interpreting this data [30].

In medical image processing, CT images are usually stored in the NIfTI format (.nii), while MRI images are often stored in the DICOM format (.dcm) [31]. The processing and analysis of CT and MRI images are important components of automated diagnostic systems. However, traditional image processing methods have certain limitations in medical image segmentation. Common traditional image processing methods include:

1. **Threshold segmentation**: For example, Otsu's method calculates the grayscale histogram of the image to determine the optimal threshold for segmentation. While this method is straightforward and easy to implement, its simplicity makes it challenging to effectively handle high noise levels or complex backgrounds [32].
2. **Region Growing**: Starting from one or more seed points, the region is gradually expanded according to pixel similarity. This method is very sensitive to the selection of initial seed points and the measurement of pixel similarity, and is easily affected by image noise [33].

3. **Edge Detection**: For example, Canny edge detection achieves segmentation by identifying edges in an image. This method works well when processing objects with clear edges, but performs poorly when faced with images with blurred edges or complex backgrounds [34].
4. **Morphological processing**: Includes erosion and dilation, which are used to remove noise and disconnected areas in images, but also have limitations when facing complex morphology and texture changes [35].

Although these traditional methods can assist in the segmentation of liver lesions to a certain extent, the application of traditional methods in accurate segmentation is limited because the liver and its lesions usually show complex morphology and texture changes in images. The low contrast between the liver and surrounding tissues, coupled with the variable shape of the lesions, makes it difficult for traditional methods to provide accurate segmentation results [36].

2.3 Related Work on Image segmentation

Instance segmentation is a computer vision task focused on identifying and segmenting each individual target object within an image. Traditional instance segmentation methods include methods based on thresholding, region growing, edge detection, and graph theory. However, these methods face challenges when addressing the complex morphology and background associated with the liver and its lesions. Consequently, with advancements in deep learning technology, instance segmentation approaches based on convolutional neural networks (CNNs) have gained widespread adoption.

Common deep learning instance segmentation methods include:

1. **Mask R-CNN**: A method based on the Region Convolutional Neural Network (R-CNN) that achieves high-precision instance segmentation by simultaneously predicting the bounding box and pixel-level segmentation mask of the object [37].
2. **Fully Convolutional Networks (FCN)**: Replace the traditional convolutional network's fully connected layer with a convolutional layer to enable end-to-end pixel-level segmentation.FCN performs well in semantic segmentation, but in instance segmentation, it needs to be combined with other techniques such as conditional random field (CRF) to improve accuracy [38].
3. **UNet**: An encoder-decoder network that fuses feature information of different scales through skip connections to achieve high-precision segmentation. UNet performs well in medical image segmentation and has become a widely used benchmark model [39].
4. **YOLOv7_seg**YOLOv7_seg is an object detection and segmentation method based on the YOLO (You Only Look Once) series, which completes object detection and instance segmentation tasks simultaneously through a single forward propagation [40]. Its advantage lies in its efficient real-time performance, which is suitable for application scenarios that require fast processing and response.
5. **CycleGAN**: CycleGAN is mainly used for image-to-image translation tasks, converting one image style to another through unsupervised learning [41]. In medical image processing, CycleGAN can be used for data augmentation and cross-modal image synthesis, thereby improving the robustness of segmentation models.

When selecting a specific instance segmentation method, UNet, due to its symmetrical encoder-decoder structure and jump connection, allows high-level features and low-level features to be fully integrated, thereby achieving accurate segmentation. UNet performs well in processing multi-scale information and complex backgrounds in medical images, which gives it a clear advantage in the liver lesion segmentation task [42]. Compared with other methods, UNet has a simple structure, is easy to implement, and performs well in a variety of medical image segmentation tasks, making it the main model selected in this study.

2.4 Related Work on UNet Model

UNet was proposed by Ronneberger et al. in 2015 and is designed specifically for biomedical image segmentation [20]. UNet's key features include its symmetrical encoder-decoder architecture and skip connections, which allow for the seamless integration of high-level and low-level features, resulting in precise segmentation.

1. Applications in Medical Image Segmentation：UNet performs well in different types of medical image segmentation tasks, including tumor segmentation, organ segmentation, and lesion detection. For example, in brain tumor segmentation, lung nodule detection, and liver lesion segmentation, UNet shows high accuracy and robustness [21][22].
2. Improved UNet model:
    - U-Net++: Based on UNet, dense skip connections are introduced to further improve the segmentation accuracy [23].
    - Attention U-Net: By introducing the attention mechanism, the network can pay more attention to important feature areas and improve the segmentation effect [24].
    - 3D U-Net: Extending UNet to three-dimensional convolution for three-dimensional medical image segmentation, the effect is significantly better than the two-dimensional model [25].
    - ResNet-UNet: Combining the residual connection of ResNet and the symmetric encoder-decoder structure of UNet, the model's feature extraction capability and segmentation accuracy are improved [26].
3. Application of UNet in cross-modal segmentation：UNet not only performs well in single-modality image segmentation, but also has achieved certain success in cross-modality segmentation tasks. For example, combining CT and MR image information for joint segmentation can provide more comprehensive diagnostic information [27].

In summary, UNet and its improved models have broad application prospects and research value in medical image segmentation. Through in-depth research in liver lesion segmentation, we can further optimize and improve the performance of the UNet model and provide more effective technical support for clinical diagnosis and treatment [28][29].

Chapter 3
Methodology

3.1 Dataset selection and preprocessing

In this study, we selected two main datasets: the LITS (Liver Tumor Segmentation Challenge) dataset and the CHAOS (Combined (CT-MR) Healthy Abdominal Organ Segmentation) dataset. These two datasets have significant advantages in terms of data volume, annotation quality, and community support.

Mainstream datasets involving the liver include LITS (131 CT images, 2017), CHAOS (70 CT images, 2019), Dircadb (30 CT images, 2007), and Sliver07 (20 CT images, 2007). In comparison, the LITS and CHAOS datasets have large data volumes, accurate annotations, and comprehensive community support, making them suitable as the main datasets for this study [1][2][3][4].

3.2 Data Standardization

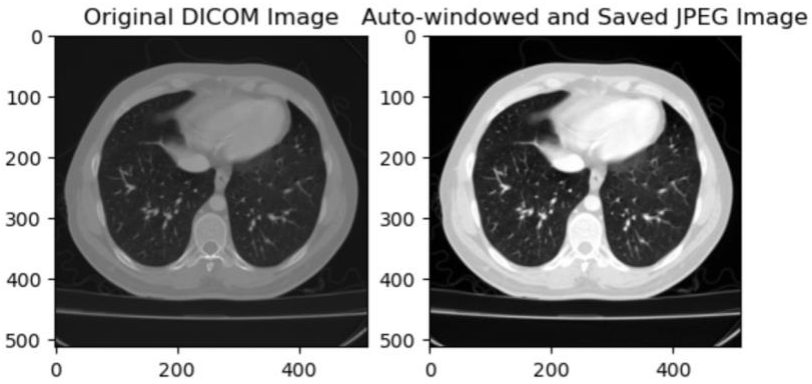
First, the CT and MR images, the LITS and CHAOS datasets were normalized by mapping the pixel value of each slice to the range of 0-255 and saving them in PNG format for subsequent processing and analysis.

1. Data Standardization:
    - For NIfTI format images in the LITS dataset, the Nibabel library in Python is used to read the nii files and normalize them into PNG format images.
    - For DICOM format images in the CHAOS dataset, the pydicom library is used to read the dcm files and normalize them into PNG format images.
2. Segmentation mask processing:
    - For the segmentation masks in the LITS dataset, whose label values are floating point numbers, we use a double threshold (0.5 and 1.5) method to process the pixel values in the segmentation masks into a format suitable for model training, retaining the pixel values between 0.5 and 1.5 and setting the rest to 0.
    - For the segmentation masks in the LITS dataset, 0 (background), 1 (liver), and 2 (lesion) are mapped to 0, 127, and 255, respectively, so that the segmentation label image appears in black, gray, and white.
    - For the segmentation mask in the CHAOS dataset, 0 (background) and 1 (liver) are mapped to 0 and 255, respectively, so that the segmentation label image appears in black and white.
3. Dataset division:

    The CHAOS dataset and the LITS dataset are unified according toRandomly divided into training set (80%) and validation set (20%). Similarly, the training set is used for model training, and the validation set is used for model performance evaluation [3]
    The processed LITS dataset contains 6532 PNG slice images, 5431 images in the training set, and 1101 images in the test set. The processed CHAOS dataset contains 1084 PNG slice images, 934 images forme the training set, and 150 images forme the test set.

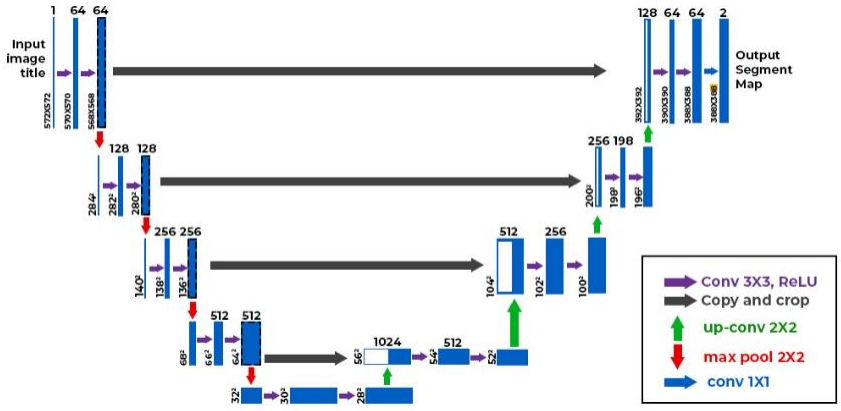
In order to meet the needs of the YOLOv7-seg algorithm, we use the Canny edge detection algorithm to process the segmentation mask part to obtain the edge coordinates of the image. The Canny algorithm can effectively extract the edge information of the image, which is convenient for subsequent segmentation tasks.

In summary, through the standardization, segmentation mask processing, image enhancement and dataset partitioning of LITS and CHAOS datasets, we have laid a solid foundation for subsequent model training and evaluation. These processing steps not only help to improve the segmentation accuracy of the model, but also enhance the generalization ability and robustness of the model, providing more reliable technical support for clinical applications.



Graph(3)

3.3 UNet Basic network structure



Graph(4)

UNet is a CNN for image segmentation, which extracts and reconstructs multi-scale features through encoder and decoder structures. Its main features are the symmetrical encoder-decoder structure and jump connection, which allows high-level features and low-level features to be fully integrated, thus achieving accurate segmentation.

Encoder

The encoder consists of four downsampling modules, each of which contains a maximum pooling layer and two convolutional layers (DoubleConv). As downsampling progresses, the spatial dimensions of the feature map decrease gradually, while the number of channels increases incrementally. The detailed process is as follows:

1. Input Layer: Input size is (1, 570, 570).
2. After in_conv (DoubleConv): The output size is (base_c, 570, 570).
3. After down1 (Down): The output size is (base_c * 2, 285, 285).
4. After down2 (Down): The output size is (base_c * 4, 142, 142).
5. After down3 (Down): The output size is (base_c * 8, 71, 71).
6. After down4 (Down): The output size is (base_c * 16 // factor, 35, 35).

Decoder

The decoder consists of four upsampling modules, each of which contains an upsampling layer and two convolutional layers (DoubleConv). The decoder reconstructs the image by gradually restoring the spatial size of the feature map while reducing the number of channels. Each upsampling module is concatenated (skip connection) with the corresponding encoder layer output to utilize the high-resolution information of the early encoder layers. The specific process is as follows:

1. After up1 (Up): The output size is (base_c * 8 // factor, 71, 71).
2. After up2 (Up): The output size is (base_c * 4 // factor, 142, 142).
3. After up3 (Up): The output size is (base_c * 2 // factor, 285, 285).
4. After up4 (Up): The output size is (base_c, 570, 570).

The role of skip connections

Skip connections are used in UNet to combine the features of the encoder and decoder, allowing the decoder to utilize the high-resolution information of the early layers of the encoder, thereby better restoring the details of the input image and improving segmentation performance. This connection method provides fine-grained features during the reconstruction process, helping to preserve the edges and details of the image.

Output Layer

1. After out_conv (OutConv): The final output size is (num_classes, 570, 570).

Through this encoder-decoder structure and jump connection, UNet can effectively extract and reconstruct multi-scale features and achieve high-precision image segmentation. This network structure performs well in tasks such as medical image segmentation and is widely used in various segmentation tasks.

3.4. Loss Function and Model Evaluation Metrics

In order to optimize the segmentation performance of the UNet model, we selected the cross-entropy loss function as the main loss function. The cross-entropy loss function can

effectively measure the difference between the predicted segmentation result and the actual label, and is suitable for pixel-level classification tasks. Its calculation formula is as follows:

$$\text{Loss} = -\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

in:

y_i is the actual label.

$\hat{y}_i$ is the predicted value.

N is the total number of pixels.

In the liver lesion segmentation task, in order to evaluate the performance of the UNet model, we used common evaluation indicators, including the Dice coefficient and the intersection over union (IoU). These indicators can effectively measure the accuracy and robustness of the model in the segmentation task.

1. Dice Coefficient

The Dice coefficient is a commonly used indicator for evaluating the performance of segmentation models. It is mainly used to measure the similarity between two samples. It evaluates the segmentation accuracy of the model by calculating the overlap area between the predicted segmentation result and the actual label. The Dice coefficient ranges from 0 to 1, where 1 indicates complete overlap and 0 indicates no overlap .

The formula for the Dice coefficient is as follows:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}$$

in:

$|A|$ is the predicted segmentation region.

$|B|$ is the actual segmentation area.

$|A \cap B|$ is the intersection of the predicted area and the actual area, meaning the number of pixels in the overlapping part.

2. Intersection over Union (IoU)

Intersection over Union (IoU) is another commonly used segmentation performance evaluation metric, which is used to measure the overlap between the predicted segmentation region and the actual label. IoU evaluates the segmentation accuracy of the model by calculating the intersection and union of the predicted segmentation result and the actual label.

The formula for IoU is as follows:

$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

in:

$|A|$ is the predicted segmentation region.

$|B|$ is the actual segmentation area.

$|A \cap B|$ is the intersection of the predicted area and the actual area, meaning the number of pixels in the overlapping part.

$|A \cup B|$ is the union of the predicted area and the actual area.

The IoU value ranges from 0 to 1, where 1 indicates complete overlap and 0 indicates no overlap. Compared to the Dice coefficient, IoU is more sensitive to the difference between the predicted area and the actual area, and therefore can more rigorously evaluate the performance of the segmentation model.

3.5 Experimental Design

The main purpose of this project is to improve the accuracy of the segmentation of liver lesions. Therefore, the evaluation will use the IoU of 2 (lesions) in the LiTS data as the main evaluation indicator, and the Dice coefficient as the secondary evaluation indicator. For the first-stage algorithm, I will compare the segmentation effects of three different frameworks: YOLOv7_seg, CycleGAN, and 3D_UNet. For the second-stage algorithm, I mainly implement it by stacking two neural networks. By comparing the four backbones of UNet, ResUNet, UNet++, and Attention UNet, the best semantic segmentation model in the healthy liver (Chaos dataset) is selected, and then the LiTS dataset is inferred through the saved weights, retaining only the mask of the liver part. Finally, this processed dataset is used as input, and different backbones are used for training, and the Dice coefficients are compared to select the best segmentation framework.
In order to verify that stacking neural networks can indeed improve segmentation accuracy, the following is an experiment using the Lits dataset containing only labels 0 (background) and 1 (lesion). The Dice coefficients of the 2stage_UNet method are compared between using UNet for semantic segmentation and first using UNet to segment the liver and then using UNet for lesion segmentation.

The main purpose of this project is to improve the accuracy of liver lesion segmentation. The evaluation will use the IoU of label 2 (lesion) in the LiTS dataset as the primary evaluation metric and the Dice coefficient as the secondary evaluation metric.

One-stage algorithm evaluation

For the one-stage algorithm, we will compare the segmentation effects under the following three different frameworks:

- YOLOv7_seg
- CycleGAN
- 3D_UNet

Two-stage algorithm evaluation

For the two-stage algorithm, we will implement it by stacking two neural networks. The specific process is as follows:

1. Healthy liver (CHAOS dataset) segmentation:
    - Compare the segmentation effects under four different backbones: UNet, ResUNet, UNet++, and Attention UNet.
    - Select the semantic segmentation model with the best performance in the CHAOS dataset.
2. Liver lesion (LiTS dataset) segmentation:
    - The weights of the optimal model are used to perform inference on the LiTS dataset, keeping only the mask of the liver part.
    - The processed data set is used as input, and different backbones are used for training. The Dice coefficient is compared to select the best segmentation framework.

3.5.1 Verifying the effectiveness of stacked neural networks

To verify that the stacked neural network approach can indeed improve segmentation accuracy, we will conduct the following experiments:

1. Directly use the LiTS dataset (labeled as 0: background, 1: lesion):
    - Directly use UNet for lesion segmentation.
    - Calculate and record the Dice coefficient.
2. Two-stage segmentation (2Stage_UNet):
    - First use UNet to segment the liver, and then use UNet to segment the lesion.
    - Calculate and record the Dice coefficient.

Through the above experiments, we directly compare the performance of using UNet and the two-stage segmentation method to evaluate the effectiveness of the two-stage segmentation method in improving segmentation accuracy. According to the experimental results in Table (a) below, it can be seen that using 2Stage_UNet can improve the segmentation accuracy of liver lesions.

Model	Dice	IoU(class0)	IoU(class1)	MiO
UNet	**0.871**	**99.2**	**87.6**	**92.4**
2Stage_UNet	**0.934**	**98.9**	**89.7**	**94.3**

Table(a)

Summary of the evaluation process

1. Data preprocessing: Standardizing and processing LiTS and CHAOS datasets.
2. Model Training:
    - One-stage algorithms: YOLOv7_seg, CycleGAN, 3D_UNet.
    - Two-stage algorithm: Select the best backbone and perform two-stage segmentation of the liver and lesions.
3. Performance Evaluation:
    - Main evaluation metric: IoU of lesions in the LiTS dataset.
    - Secondary evaluation indicator: Dice coefficient.
4. Experimental verification: Verify the effectiveness of the two-stage segmentation method and compare the Dice coefficients of different methods.

Through the above steps, this study hopes to find the optimal liver lesion segmentation method, improve segmentation accuracy, and provide more reliable technical support for clinical applications.

Chapter 4
Experiment

4,1 Experiment Overview

The experimental part of this project mainly consists of two parts. The first stage algorithm is used to directly perform instance segmentation to obtain the segmentation mask. In this part, the focus is on comparing CycleGAN, 3D_UNet, YOLOv7_segThree different frameworks with diceThe coefficient is the segmentation effect of the evaluation index; using the two-stage algorithm, the liver part of the image is first semantically segmented, only the liver part of the image is selected to reduce the impact of the background on the subsequent liver lesion segmentation task, and then the semantic segmentation of the liver part is performed. In this part, the comparison will be made successively.UNet, Attentio_UNet,ResUNet, UNet+++In ChaosThe segmentation results of the liver part in the data set, select the b with the highest segmentation accuracyackbone,The output of this network is then used for the subsequent segmentation of liver lesions. The following section will briefly describe the network structure of the one-stage method and the two-stage method as well as the characteristics of each method.

The experimental part of this project mainly consists of two parts: direct instance segmentation using a one-stage algorithm, and segmentation using a two-stage algorithm. The following is a brief description of the experimental settings and methods.

4.1.1 First-stage algorithm experiment

In the one-stage algorithm experiment, we directly perform instance segmentation on the image to obtain the segmentation mask.The performance of three different frameworks, CycleGAN, 3D_UNet and YOLOv7_seg, in segmentation tasks, with the main evaluation indicator being the Dice coefficient.

CycleGAN: Used to generate synthetic image data to augment the training set, thereby improving segmentation performance.

3D_UNet: Processes 3D medical images through 3D convolution to capture the features of 3D structures such as tumors.

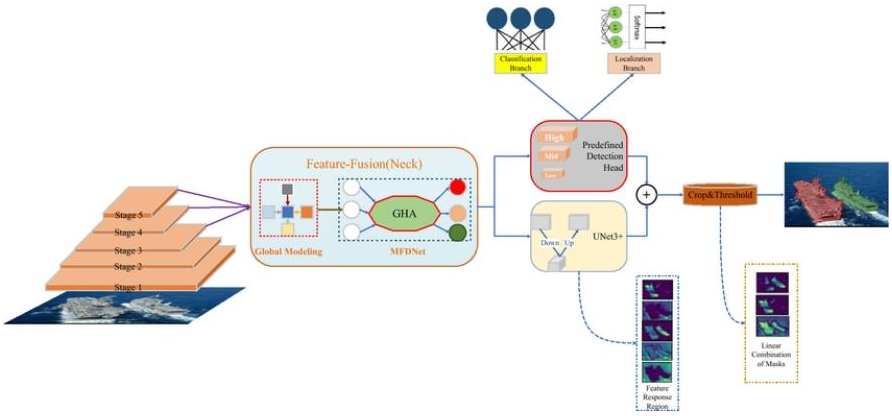
YOLOv7_seg: Combines object detection and instance segmentation with efficient real-time performance.

4.1.2 Two-stage algorithm experiment

In the two-stage algorithm experiment, we first perform semantic segmentation on the liver part of the image and retain only the liver part of the image to reduce the impact of the background on the subsequent liver lesion segmentation task. Then, we further perform liver lesion segmentation on the image containing only the liver part.The segmentation results of the liver part of UNet, Attention_UNet, ResUNet and UNet+++ in the CHAOS dataset. The backbone with the highest segmentation accuracy is selected and the output of the backbone is used as the input for subsequent liver lesion segmentation., and compare each backboneThe performance of segmentation is still poor, and the main evaluation indicator is Dicecoefficient.

Through the above two parts of experiments, we aim to find the optimal liver lesion segmentation method and verify its performance on different datasets. Specifically, we will evaluate various network structures and methods under the framework of the one-stage algorithm and the two-stage algorithm, and finally select the segmentation model with the best effect.

4.2YOLOv7_SEG



Graph(5)

YOLOv7-SEG extends YOLOv7 to handle image segmentation tasks, combining object detection and segmentation within a single model. This approach leverages multi-task learning to improve both tasks' performance by sharing the feature extraction layers. YOLOv7-SEG retains YOLO's efficient end-to -end training and inference capabilities, making it suitable for real-time applications even on resource-constrained devices.

The model architecture includes:

Backbone: CSPDarknet53 for multi-scale feature extraction.

Neck: PANet and FPN to enhance feature pyramid structure and fuse features from different scales.

Detection Head: Outputs bounding boxes, class probabilities, and confidence scores, similar to YOLOv7.

Segmentation Head: Added to generate pixel-level segmentation masks from the feature pyramid.

Advantages:

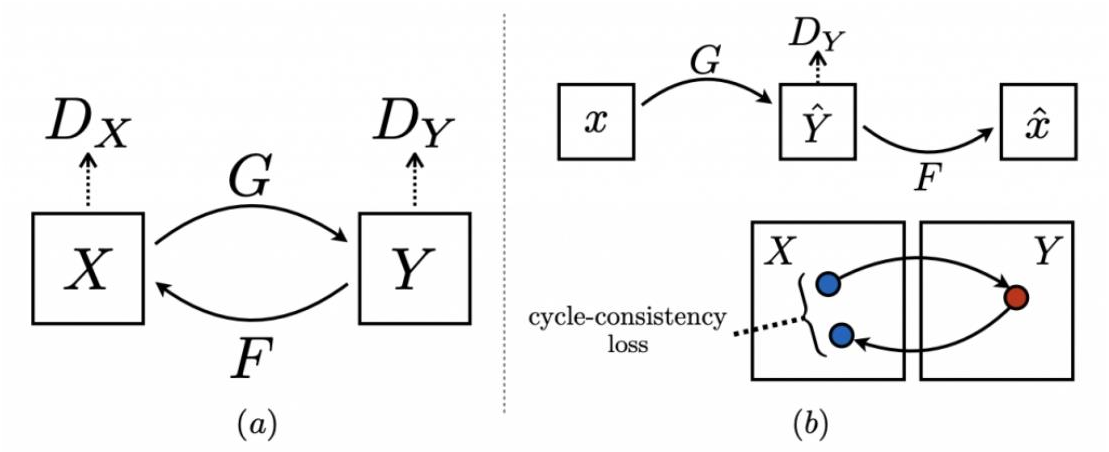
Unified Design: Combines detection and segmentation, reducing parameters and computation.

High Performance: Improved feature pyramid network enhances segmentation and detection accuracy.

Fast Inference: Maintains real-time performance on edge devices and embedded systems.

YOLOv7-SEG provides a powerful solution for tasks requiring both detection and segmentation, such as autonomous driving, video surveillance, and medical image analysis

4.3 Cycle GAN



Graph(6)

CycleGAN (Cycle-Consistent Generative Adversarial Networks) is a generative adversarial network for image-to-image translation that can achieve mutual conversion of two image styles without paired samples.

Network structure

- Generator G: Convert an image in domain X to an image in domain Y.
- Generator F: Convert an image in domain Y to an image in domain X.
- Discriminator DXD XDX: Distinguish between real domain X images and generated domain X images.
- Discriminator DYD YDY: Distinguish between real domain Y images and generated domain Y images.

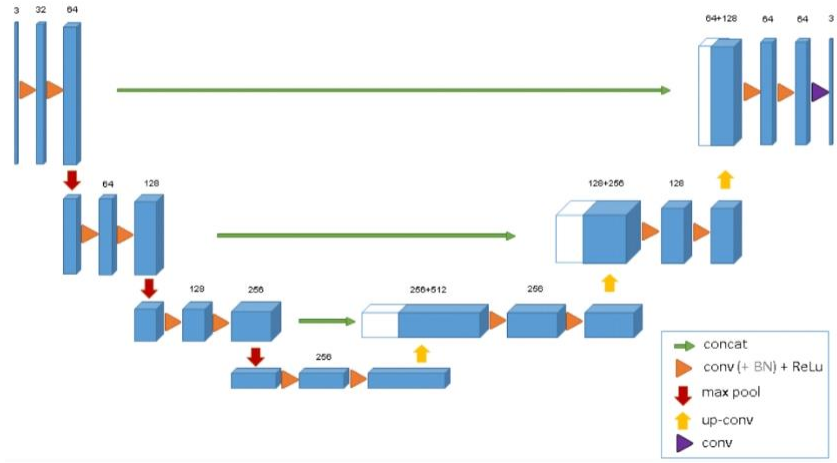
Loss Function

- Fighting Losses: Ensure that the generated images are as realistic as possible.
- Cycle consistency loss: Ensure that the image can be restored to its original state after two conversions.

The connection between CycleGAN and LITS liver instance segmentation

- Image Domain Adaptation: Convert images from different imaging devices or hospitals to a unified domain to reduce the impact of inter-domain differences on the segmentation model.
- Data Augmentation: Generate liver images in different styles, expand the training dataset, and improve the generalization ability of the segmentation model.
- Improve robustness: By converting training data, the model's adaptability to different image changes is improved and its robustness in practical applications is enhanced.

4.4 3D UNet



Graph(7)

3D U-Net is a deep learning model suitable for 3D medical image segmentation. Its architecture includes an encoder, intermediate layers, and a decoder, and processes 3D image data through 3D convolutional layers and 3D pooling layers.

Network structure

- **Encoder:** The encoder is composed of multiple convolutional blocks, each featuring two 3D convolutional layers, a ReLU activation function, and a 3D max-pooling layer, all working together to extract the spatial features of the image.

    - **Middle Layer:** This layer includes two 3D convolutional layers with ReLU activation, designed to extract deeper features from the image.

    - **Decoder:** The decoder comprises several upsampling blocks that utilize 3D deconvolution layers to restore the image's spatial resolution. Each of these upsampling blocks is followed by two 3D convolutional layers and a ReLU activation function.

    - **Output Layer:** A 128x128x128 3D convolutional layer is employed in the output layer to convert the feature map into the desired number of categories.

-

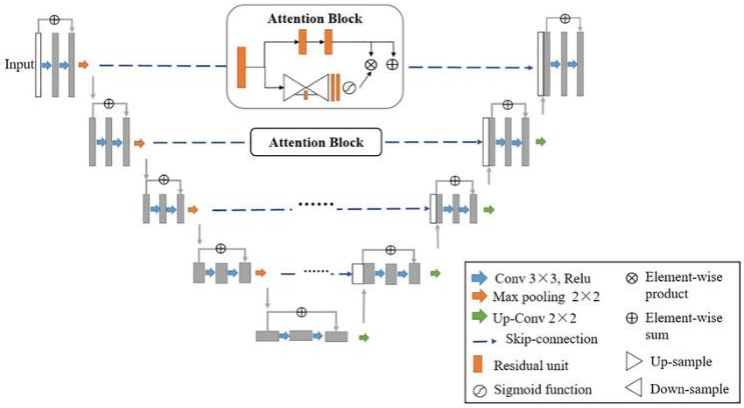
Implementation of 3 class instance segmentation

In order to achieve instance segmentation of the three classes, 3D U-Net is designed as follows:

1. Multi-channel output:
    - The last layer of the model uses a 1x1x1 3D convolutional layer to convert the feature map to the required number of categories. For instance segmentation of 3 classes, the number of output channels is set to 3, that is, num_classes=3. Each channel corresponds to a probability map of a category.
2. Loss Function:
    - Use CrossEntropyLoss as the loss function. This loss function is suitable for multi-class segmentation tasks and can handle multi-channel outputs and corresponding multi-channel labels.
    - CrossEntropyLossEach channel of the model output is compared with the corresponding label, and the classification error of each pixel is calculated to guide model optimization.
3. Label preprocessing:
    - In the data loader, make sure the labels also have the same dimensions as the model output. Convert the labels to long integers (torch.long) to accommodate the requirements of CrossEntropyLoss.
    - The label needs to correspond to the input image and contain the classification information of each pixel so that the model can learn the characteristics of each category during training.

Through the above design, 3D U-Net can effectively perform instance segmentation of three classes. During the training process, the model learns the characteristics of each class and outputs a probability map corresponding to each class in the inference stage to achieve accurate instance segmentation.

4.5 **AtAttention UNet**



Graph(8)

Attention U-Net is an improved version of the traditional U-Net, which enhances feature fusion and positioning capabilities by introducing an attention mechanism in the upsampling process.

Encoder (downsampling path)

Conv1 - Conv5: Each convolution block (conv_block) contains two convolutional layers, batch normalization and ReLU activation, and the number of channels is 64, 128, 256, 512, and 1024 respectively.

Maxpool1 - Maxpool4: Each max pooling layer (MaxPool2d) halves the spatial size.

Decoder (upsampling path)

Up2 - Up5: Each upsampling block (up_conv) contains an upsampling layer and a convolutional layer, and the number of channels is 512, 256, 128, and 64 respectively.

Att2 - Att5: Attention_block introduces the attention mechanism in the upsampling process to enhance feature fusion.

Up_conv2 - Up_conv5: Each upsampled convolution block (conv_block) processes the upsampled feature map.

Output Layer

Conv: 1x1 convolutional layer, converting the number of channels to the number of target categories.

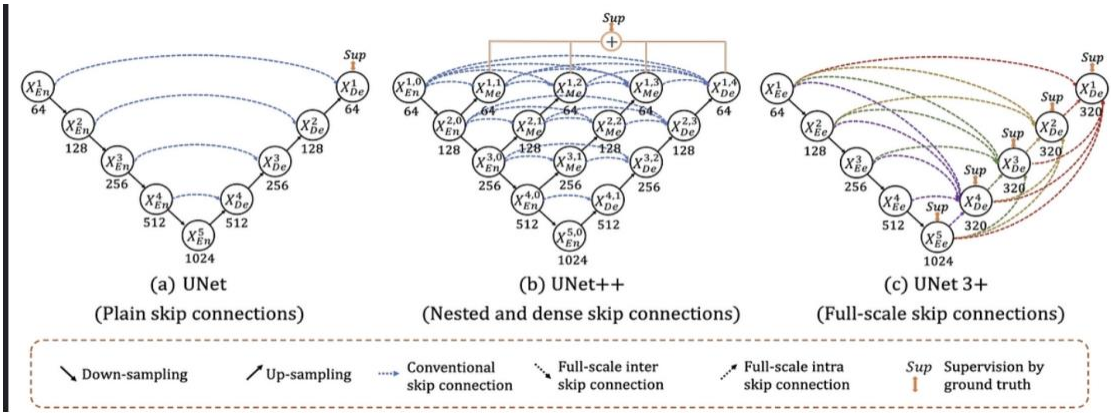
TraditionDifferences between U-Net and Attention U-Net

Attention Mechanism:Attention U-Net introduces attention blocks in the upsampling process to enhance feature fusion and localization capabilities.

Feature stitching:Attention U-Net weights the features before concatenation, while traditional U-Net directly concatenates the features.

Network complexity:The Attention U-Net structure is more complex, but it can better retain important features and improve segmentation accuracy.

4.6UNet++



Graph(9)

UNet 3+ is an extension of traditional UNet with a more complex feature fusion mechanism and improved segmentation performance.

Encoder (downsampling path)

conv1 - conv5: five convolutional blocks, each containing two convolutional layers, batch normalization and ReLU activation, with the number of channels being 64, 128, 256, 512, and 1024, respectively.

maxpool1 - maxpool4: Four max pooling layers that halve the spatial size.

Decoder (upsampling path)

At each decoding stage, the feature maps of each layer are adjusted to the same size through maximum pooling and upsampling and then fused:

Stage 4d: fuses five feature maps with 320 channels.

Stage 3d: fuses five feature maps with 320 channels.

Stage 2d: fuses five feature maps with 320 channels.

Stage 1d: fuses five feature maps with 320 channels.

Output Layer

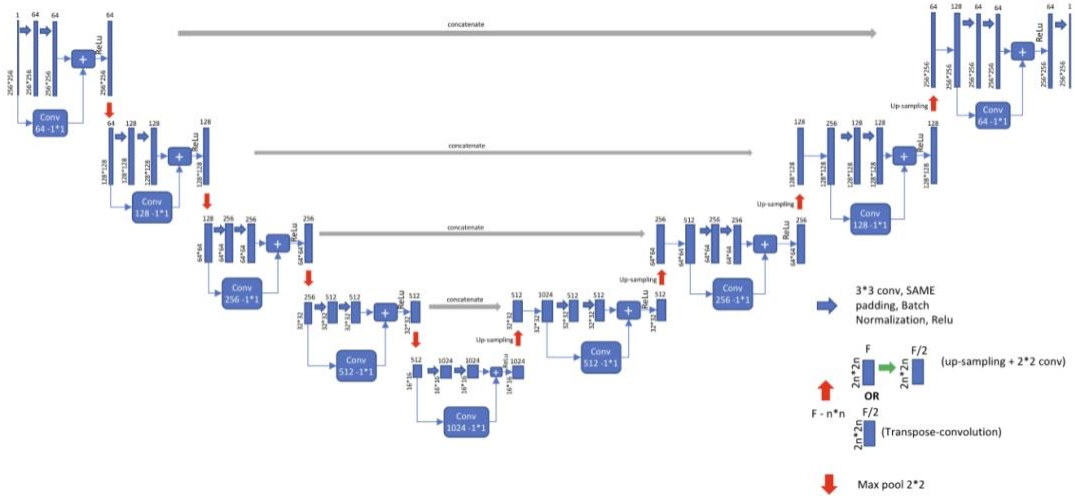
outconv1: 1x1 convolutional layer that converts the number of channels to the number of target categories.

The main differences between UNet 3+ and traditional UNet

Feature Fusion:UNet 3+ fuses feature maps from different levels at each decoding stage, while traditional UNet only performs simple concatenation.

Multi-scale information utilization: Through complex feature fusion mechanism,UNet 3+ makes better use of multi-scale information.

4.7 ResUNet++



Graph(10)

ResUNet++ Network Structure Overview

ResUNet++ is an improved U-Net network that combines the residual module (ResNet), attention mechanism and multi-scale feature extraction, suitable for image segmentation tasks. The following is a brief overview of its network structure:

Encoder (downsampling path)

Stem Block (c1): The initial convolutional block, which contains two convolutional layers, batch normalization and ReLU activation, and the number of output channels is 16.

ResNet Block (c2 - c4): Three residual blocks, each containing two convolutional layers, batch normalization and ReLU activation, with the number of channels being 32, 64, and 128 respectively.

Encoder bottleneck layer

ASPP (b1): Spatial pyramid pooling module, which uses different dilated convolution rates to capture multi-scale information and has 256 output channels.

Decoder (upsampling path)

Decoder Block (d1 - d3): Three decoding blocks, each containing an attention mechanism, an upsampling and a residual block, with the number of channels being 128, 64, and 32, respectively.

Decoder Block: includes attention blocks, upsampling and residual blocks, gradually restoring the spatial resolution.

Output Layer

ASPP (aspp): multi-scale feature extraction, the number of output channels is 16.

Output: 1x1 convolutional layer, converting the number of channels to the number of target categories.

The main differences between ResUNet++ and traditional UNet

Residual Module:ResUNet++ uses residual blocks in the encoder and decoder to enhance feature extraction capabilities and gradient flow.

Attention Mechanism:ResUNet++ uses attention blocks in the decoder to improve feature fusion and localization capabilities.

Multi-scale feature extraction: UseThe ASPP module captures multi-scale information and improves the model's ability to perceive features of different scales.

Chapter 4
Discussion

5.1 Discussion Summary

In this study, we conducted a detailed evaluation of the segmentation effect of liver lesions. The experimental results show that the performance of the one-stage algorithm in the liver lesion segmentation task is obviously insufficient. In particular, in the prediction results of 3D_UNet and CycleGAN, the untrained test set showed a problem of weak generalization. Among them, the best performing YOLOv7_seg model, with a Dice coefficient of 0.792, still did not achieve the ideal segmentation effect.

In contrast, in the two-stage algorithm, the Attention UNet is first used to perform semantic segmentation of the liver to reduce the impact of the background on lesion segmentation, and then the standard UNet is used to refine the lesion segmentation. This method performs best in the lesion segmentation task, achieving a Dice coefficient of 0.934, which is significantly better than the results of the one-stage algorithm.

These experimental results show that the two-stage algorithm effectively improves the accuracy and robustness of liver lesion segmentation through step-by-step processing and key area segmentation. In particular, the combination of Attention UNet and standard UNet shows great potential in handling complex medical image segmentation tasks.

5.2 Disadvantages of the one-stage algorithm

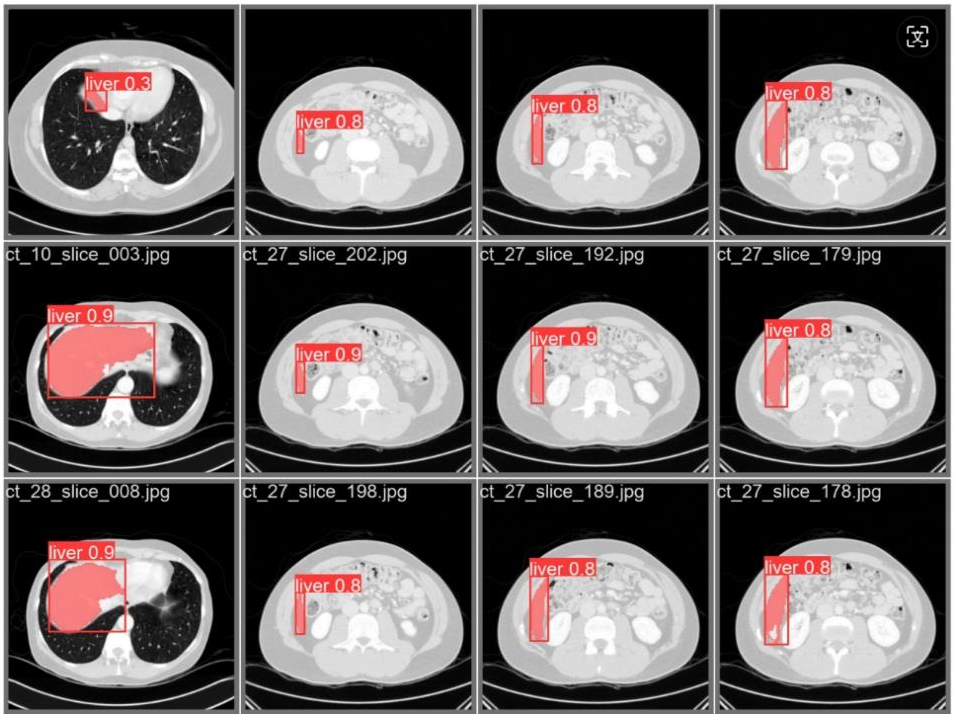
Table (Table b) shows the Dice coefficients of YOLOv7_seg, CycleGAN and 3D_UNet in the liver lesion segmentation task:

Model	YOLOv7_seg	CycleGAN	3D_UNet
Dice	**0.792**	**0.192**	**0.702**

Table(b)

Although YOLOv7_seg performs well in segmenting the liver, the overall segmentation effect is not ideal. The following will analyze the segmentation effect of YOLOv7_seg in detail.

5.2.1 Segmentation effect of YOLOv7_seg



Graph(11)

By directly using the training weights for prediction, Graph (11) shows the segmentation results of YOLOv7_seg on the test set. Although YOLOv7_seg performs well in segmenting the liver (class 1), it performs poorly in segmenting liver lesions (class 2). Table (Table c) shows the proportion of segmentation masks of each category in CT images

Class	0(background)	1(liver)	2(lesion)
Masks rate	**97.96%**	**2.02%**	**0.02%**

Table(c)

As can be seen from Table c, the segmentation mask area of the liver lesion accounts for only 0.02% of the total area of the CT image, while the segmentation mask area of the liver accounts for 2.02%. This imbalance significantly affects the performance of the model in segmenting lesions.

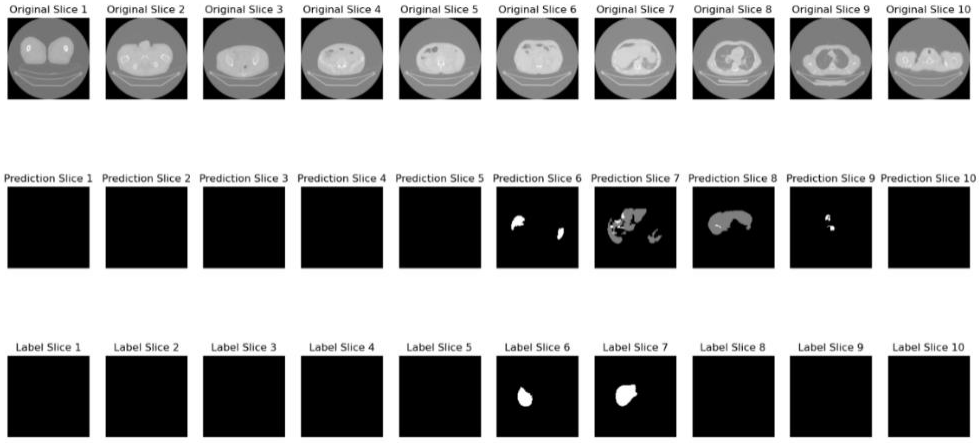
The main reason for the poor segmentation results of YOLOv7_seg is the data set partitioning strategy of the model. YOLOv7_seg relies on the edge information of the segmentation mask for segmentation, while the edge of the segmentation mask of the liver is relatively regular. The lesion part has a complex shape due to different degrees of lesions,

resulting in unsatisfactory segmentation results. The lesions have diverse shapes and occupy a very small proportion in the image, making it difficult for the model to fully learn the characteristics of the lesions during training, thus affecting the overall segmentation accuracy.

5.2.2 Segmentation Effect of CycleGAN

As shown in Graph (11), the results of using CycleGAN to predict the test set show that although CycleGAN does generate images that are very similar to the segmentation mask, its recognition accuracy is very low, with a Dice coefficient of only 0.192. This is because CycleGAN's cycle consistency loss is not for the one-to-one correspondence between images and segmentation masks. CycleGAN achieves instance segmentation by translating images in the X domain into images in the Y domain. This method is not suitable for instance segmentation, a task where the input and output have a strong correspondence. Therefore, the application of CycleGAN in this task is limited and cannot meet the needs of high-precision segmentation.

The following Graph (11) has incorrect segmentation.



Graph(11)

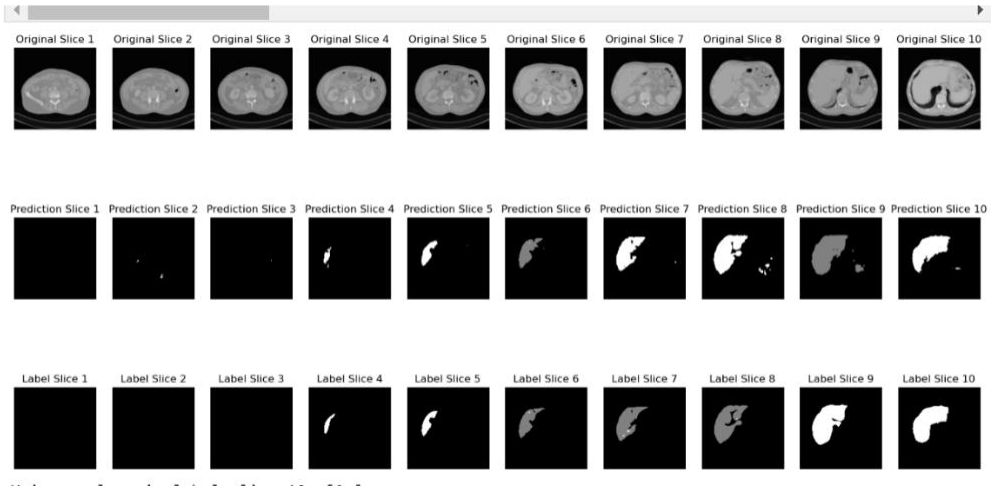
5.2.3 3D_UNet segmentation effect

In Graph (12), we can see that there are several obvious classification errors in the prediction results of 3D_UNet. This problem is mainly caused by the following two aspects:

- **Model complexity and data volume**: Due to the use of 3D_UNet, the network is directly input with a complete CT image in nii format, and the output is also a CT image in nii format. Compared with the standard 2D UNet, the trainable weights of 3D_UNet have increased significantly, but the amount of data available for training has decreased relatively, resulting in poor generalization and robustness of the model. When the model processes unseen data, it is easy to overfit, which affects the segmentation accuracy.
- **Similarity of dataset annotations**: The liver part labels and liver lesion part labels in the dataset itself have high similarity, which increases the difficulty of the

segmentation task. The contrast difference between the liver and lesion areas in CT images is not large, which can easily lead to model confusion and misclassification.

The following Graph(12)There was a problem with misclassification



Graph(12)

5.3 Advantages and Comparison of Two-Phase Algorithm

Compared with the one-stage algorithm, the two-stage algorithm first semantically segments the liver part of the image, and then further segments the lesion area based on this, effectively reducing the impact of the background on the segmentation results. We first selected the best backbone in the CHAOS dataset, and then further trained and segmented it in the LiTS dataset. The experimental results show that the two-stage algorithm significantly improves the segmentation accuracy and robustness of the model.

As shown in Table d, we compared the segmentation effects of four backbones, UNet, UNet+++, Attention_UNet, and ResUNet, in the CHAOS dataset. Experiments show that Attention_UNet performs best in segmentation accuracy, with a Dice coefficient of 0.919 and a MioU of 96.7. Based on this result, we selected Attention_UNet as the initial backbone for the two-stage segmentation.

Table d: Segmentation results of the CHAOS dataset representing healthy liver using different backbones

Model	Dice	IoU(class0)	IoU(class1)	MiO
U-Net	**0.849**	**99.8**	**88.6**	**93.7**

U-Net+++	**0.906**	**99.3**	**93.1**	**96.2**
AttUNet	**0.919**	**99.4**	**94.0**	**96.7**
ResUNet	**0.910**	**99.3**	**92.6**	**95.9**

Table(d)

In the two-stage segmentation experiment, we used different backbones for training and evaluated them on the LiTS dataset. The results are shown in Table e. Finally, we chose a combination of using Attention_UNet for liver segmentation and then using standard UNet for lesion segmentation. The Dice coefficient of this method reached 0.936 and MioU was 96.7, showing its superiority in handling liver lesion segmentation tasks.

Table e:Contains diseased liverTwo-stage segmentation results in LiTS dataset

Model	Dice	IoU(class0)	IoU(class1)	MiO
2stage_UNet	**0.934**	**98.8**	**89.7**	**94.3**
2stage_UNet+++	**0.925**	**99.8**	**87.3**	**92.1**
2stage_AttUNet	**0.843**	**88.6**	**48.0**	**68.3**
2stage_ResUNet	**0.930**	**98.9**	**89.3**	**94.0**

Table(e)

5.4 Backbone Feature Analysis

By analyzing the performance of different backbones in the two-stage algorithm, we can draw the following conclusions:

1. UNet：UNet's symmetric encoder-decoder structure and jump connections enable it to effectively capture the multi-scale features of images. However, its basic structure is relatively simple, which makes it perform slightly worse when processing complex images.
2. UNet+++: Compared with the basic UNet, UNet+++ further enhances the ability of feature fusion by introducing dense skip connections. This improvement makes UNet+++ perform better in detail processing, especially in more accurate segmentation of target areas in complex backgrounds.
3. Attention_UNet：Attention_UNet introduces an attention mechanism based on UNet, which enables the network to pay more attention to important feature areas and ignore irrelevant background information. This mechanism significantly improves

segmentation accuracy, especially in tasks that require high-precision segmentation. In this study, Attention_UNet performed well in liver segmentation and provided more accurate initial input for lesion segmentation in the two-stage segmentation.

4. ResUNet: ResUNet combines the advantages of the residual network (ResNet) and solves the gradient vanishing problem in deep network training by introducing residual connections. This structure makes ResUNet more robust and stable when processing complex and high-resolution images.

5.5 Summary

The results show that due to the great similarity between the diseased liver and the normal liver, the one-stage algorithm needs to complete regression and classification tasks at the same time, so it is difficult to segment the selected liver and liver tumors in terms of accuracy compared to the two-stage algorithm, and the accuracy on the validation set (test set) without network training has worse recognition accuracy. In this project, by stacking two neural networks, semantic segmentation of the lesion part is performed on the basis of semantic segmentation of the liver, and the segmentation accuracy of the lesion part has reached 0.934Dicecoefficient.

In addition, the study shows that although the top segmentation algorithms perform well in some tumor detection tasks, there is still room for improvement, and research in this direction still has great research potential..existUsing the UNet frameworkFocusing on the task of liver and lesion segmentation,The quality of the final segmentation effect often depends on UNetFeature extraction part under the frameworkEncoderThe quality of feature extraction is positively correlated.

In summary, through experiments and comparisons of the one-stage and two-stage algorithms, we found that the two-stage algorithm has obvious advantages in processing liver lesion segmentation tasks. CycleGAN and 3D_UNet have certain limitations in one-stage segmentation, and it is difficult to achieve ideal segmentation effects. The two-stage algorithm greatly improves the segmentation performance and generalization ability of the model by processing and segmenting key areas in steps. In particular, the combination of Attention_UNet and standard UNet further improves the segmentation accuracy, demonstrating its potential in complex medical image segmentation tasks. Future research can continue to optimize the two-stage algorithm and explore more effective backbone structures and training strategies to further improve segmentation accuracy and provide more reliable technical support for clinical diagnosis.

Chapter 6
Conclusion

6.1 Conclusion

In this study, we evaluated the performance of the one-stage and two-stage algorithms in detail for the liver lesion segmentation task. Through experimental analysis, we draw the following conclusions:

1. Limitations of the One-Stage Algorithm: CycleGAN and 3D_UNet show poor generalization on untrained test sets. CycleGAN is not suitable for instance segmentation tasks with strong input-output correspondence due to its cycle consistency loss mechanism. Although 3D_UNet has advantages in three-dimensional feature extraction, its high complexity and insufficient training data lead to poor generalization and robustness of the model.
2. Advantages of the two-stage algorithm：By first semantically segmenting the liver part of the image and then segmenting the liver lesions based on this, the two-stage algorithm significantly improves the segmentation accuracy and robustness. The experimental results show that Attention_UNet as the backbone performs best in the two-stage segmentation task, showing the significant effect of introducing the attention mechanism on improving segmentation accuracy.
3. Performance of different backbones: In the two-stage segmentation experiment, the Dice coefficient of Attention_UNet reached 0.936, outperforming other backbones (UNet, UNet+++, ResUNet). Attention_UNet can focus more on important feature areas and ignore irrelevant background information, thus significantly improving the segmentation effect.

Through detailed comparison and analysis of different algorithms, this study determined the effectiveness of the two-stage algorithm in liver lesion segmentation and recommended Attention_UNet as the best backbone, providing reliable technical support for the liver lesion segmentation task.

6.2 Future Works

Although this study has achieved certain results, there are still some improvements and research directions that deserve further exploration:

1. Optimizing model structure: Further optimize and improve the structure of Attention_UNet, explore more attention mechanisms and residual connections, and enhance the model's ability to process complex images and edge details.
2. Improve computational efficiency: While ensuring segmentation accuracy, optimize the computational efficiency of the model, reduce training and inference time, and make the model more suitable for actual clinical applications.
3. Cross-dataset validation: Evaluate the generalization ability and robustness of the model by validating it on more datasets of different sources and types.
4. Explore applications of nnUNet：Future research can explore the application of nnUNet (Universal Medical Image Segmentation Network). nnUNet is an adaptive

segmentation framework that can automatically adjust the model structure and training strategy to adapt to different segmentation tasks. By combining the advantages of nnUNet, it is expected to further improve the segmentation accuracy and robustness.

5. Research on multimodal data: In the future, the research scope can be expanded to explore the application of multimodal data (such as the combination of CT and MRI) in liver lesion segmentation. Multimodal data can provide more information and may significantly improve the segmentation effect.

Through further optimization and expansion, the two-stage algorithm and Attention_UNet model proposed in this study are expected to be applied in more medical image segmentation tasks, providing more reliable and efficient technical support for clinical diagnosis and treatment.

List of References

1. [World Health Organization] World Health Organization. "Cancer." https://www.who.int/news-room/fact-sheets/detail/cancer.

2. [Heimann 2009] T. Heimann, et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," IEEE Transactions on Medical Imaging, vol. 28, no. 8, pp. 1251-1265, 2009.

3. [Ronneberger 2015] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015.

4. [LeCun 2015] Y. LeCun, et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

5. [Liu 2017] L. Liu, et al., "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11-26, 2017.

6. [Dou 2017] Q. Dou, et al., "3D deeply supervised network for automated segmentation of volumetric medical images," Medical Image Analysis, vol. 41, pp. 40-54, 2017.

7. [Kayalibay 2017] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," arXiv preprint arXiv:1701.03056, 2017.

8. [Bilic 2019] P. Bilic, et al., "The Liver Tumor Segmentation Benchmark (LiTS)," IEEE Transactions on Medical Imaging, vol. 38, no. 10, pp. 2776-2791, 2019. doi:10.1109/TMI.2019.2913525.

9. [Kavur 2021] A. E. Kavur, et al., "CHAOS Challenge - Combined (CT-MR) Healthy Abdominal Organ Segmentation," Medical Image Analysis, vol. 69, p. 101950, 2021. doi:10.1016/j.media.2020.101950.

10. [Bilic 2019] P. Bilic, et al., "The Liver Tumor Segmentation Benchmark (LiTS)," IEEE Transactions on Medical Imaging, vol. 38, no. 10, pp. 2776-2791, 2019. doi:10.1109/TMI.2019.2913525.

11. [World Health Organization] World Health Organization, "Cancer," Available at: https://www.who.int/news-room/fact-sheets/detail/cancer.

12. [Heimann 2009] T. Heimann, et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," IEEE Transactions on Medical Imaging, vol. 28, no. 8, pp. 1251-1265, 2009.

13. [Ronneberger 2015] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015.

14. [LeCun 2015] Y. LeCun, et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

15. [Liu 2017] L. Liu, et al., "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11-26, 2017.

16. [Dou 2017] Q. Dou, et al., "3D deeply supervised network for automated segmentation of volumetric medical images," Medical Image Analysis, vol. 41, pp. 40-54, 2017.

17. [Zhou 2018] Z. Zhou, et al., "UNet++: A nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, 2018.

18. [Oktay 2018] O. Oktay, et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

19. [Christ 2017] P. F. Christ, et al., "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," arXiv preprint arXiv:1702.05970, 2017.

20. [Isensee 2021] F. Isensee, et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203-211, 2021.

21. [Gao 2019] Y. Gao, et al., "Focusnet: Imbalanced large and small organ segmentation with an end-to-end deep neural network for head and neck CT images," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2019.

22. [Zhang 2018] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749-753, 2018.

23. [Çiçek 2016] Ö. Çiçek, et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2016.

24. [Milletari 2016] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016 Fourth International Conference on 3D Vision (3DV), IEEE, 2016.

25. [Dolz 2018] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," NeuroImage, vol. 170, pp. 456-470, 2018.

26. [Zhou 2020] X. Zhou, et al., "Hi-Net: hybrid-fusion network for multi-modal MR image synthesis," IEEE Transactions on Medical Imaging, vol. 39, no. 9, pp. 2772-2781, 2020.

27. [Zhu 2017] J. Y. Zhu, et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proceedings of the IEEE International Conference on Computer Vision, 2017.

28. [Heimann 2009] T. Heimann, et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," IEEE Transactions on Medical Imaging, vol. 28, no. 8, pp. 1251-1265, 2009.

29. [LeCun 2015] Y. LeCun, et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.

30. [Liu 2017] L. Liu, et al., "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, pp. 11-26, 2017.

31. [Long 2015] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

32. [Otsu 1979] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

33. [Adams 1994] R. Adams and L. Bischof, "Seeded region growing," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 6, pp. 641-647, 1994.

34. [Canny 1986] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, pp. 679-698, 1986.

35. [Rother 2004] C. Rother, V. Kolmogorov, and A. Blake, ""“GrabCut” interactive foreground extraction using iterated graph cuts," ACM Transactions on Graphics (TOG), vol. 23, no. 3, 2004.

36. [Dou 2017] Q. Dou, et al., "3D deeply supervised network for automated segmentation of volumetric medical images," Medical Image Analysis, vol. 41, pp. 40-54, 2017.

37. [He 2017] K. He, et al., "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision, 2017.

38. [Long 2015] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

39. [Ronneberger 2015] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015.

40. [Bochkovskiy 2020] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

41. [Zhu 2017] J. Y. Zhu, et al., "Unpaired image-to-image translation using cycle-consistent adversarial networks," Proceedings of the IEEE International Conference on Computer Vision, 2017.

42. [Dou 2017] Q. Dou, et al., "3D deeply supervised network for automated segmentation of volumetric medical images," Medical Image Analysis, vol. 41, pp. 40-54, 2017.

43. [Ronneberger 2015] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015.

44. [Dolz 2018] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," NeuroImage, vol. 170, pp. 456-470, 2018.

45. [Dou 2017] Q. Dou, et al., "3D deeply supervised network for automated segmentation of volumetric medical images," Medical Image Analysis, vol. 41, pp. 40-54, 2017.

46. [Zhou 2018] Z. Zhou, et al., "UNet++: A nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, Cham, 2018.

47. [Oktay 2018] O. Oktay, et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

48. [Çiçek 2016] Ö. Çiçek, et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2016.

49. [Zhang 2018] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749-753, 2018.

50. [Zhou 2020] X. Zhou, et al., "Hi-Net: hybrid-fusion network for multi-modal MR image synthesis," IEEE Transactions on Medical Imaging, vol. 39, no. 9, pp. 2772-2781, 2020.

51. [Isensee 2021] F. Isensee, et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203-211, 2021.

52. [Christ 2017] P. F. Christ, et al., "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks," arXiv preprint arXiv:1702.05970, 2017.