Liver tumor segmentation method based on U-Net architecture: a review

Biao Wang[1, *] Chunfeng Yang[2, *]

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, PR China
[2]Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, School of Computer Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, PR China

Abstract

Liver cancer is a disease with a high incidence and high probability of deterioration, and for the rapid diagnosis of liver disease, computed tomography (CT) scans must be used to segment the liver tumors. For the past few years, with the rapid development of deep learning, many deep learning methods for liver tumor segmentation using abdominal CT images have appeared, and the clinical application of these methods is of important significance for computer-aided diagnosis of liver tumors. The U-Net, with its unique U-shape network structure, exhibits excellent performance in medical image segmentation field and has been extensively utilized in various medical image segmentation applications. In this paper, we summarize the researches of U-Net and its improved networks in CT image segmentation of liver tumors by deep learning methods and classify various U-Net-based convolutional neural networks (CNNs) into 2D (two-dimensional), 3D (three-dimensional), and 2.5D (2.5-dimensional). In this paper, 2D, 3D, and 2.5D convolutional neural networks are summarized. In addition, this paper summarizes the advantages and disadvantages as well as the improvement methods of each type of network, which provides a useful reference for the studies of deep learning based on liver tumor segmentation field. Finally, this paper envisions future research trends for deep learning segmentation methods in the context of liver tumors.

*Corresponding author. Email: wangbiao21@home.hpu.edu.cn, chunfeng.yang@seu.edu.cn

1. Introduction

Liver tumors endanger the lives and health of people all over the world, with extremely high morbidity and mortality rates, making them the second most lethal cancer in the world. Currently, human abdominal computed tomography (CT) is often used for the diagnosis and treatment planning of liver cancer, and the first step of the computer-aided diagnostic system is to separate tumors from adjacent organs and tissues in CT images [1], i.e., liver tumor segmentation. Although this task has attracted the attention of many scholars, segmenting tumors from liver CT images automatically remains a demanding task, mainly due to the varied shapes of liver tumors and the low contrast and unclear borders with adjacent organs and tissues [2]. As can be seen from **Figure 1**, the muscle around the liver tumor and the small difference in gray scale contrast between other organs (e.g., heart, kidneys, etc.) and the liver tumor result in blurred edges, posing a challenge in segmenting the liver tumor. At the same time, the size and location of liver tumors vary between individuals, and liver tumors are spread over multiple slices of CT images with subtle differences between different slices, all of which pose a great challenge to the liver tumor segmentation task.
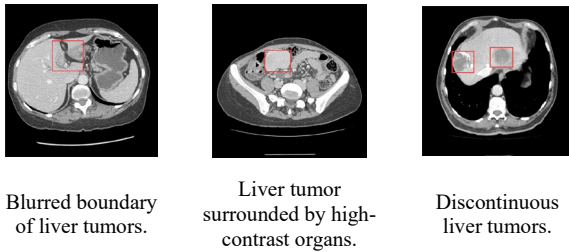
Blurred boundary of liver tumors.

Liver tumor surrounded by high-contrast organs.

Discontinuous liver tumors.

Figure 1. Slices of liver tumor

Deep learning has been developing rapidly in the last decade[3-5]. In 2015, Long et al. [6] proposed the FCN (Fully Convolutional Network) for image segmentation with great success. Due to the fruitful results of deep learning in the image area, the application of deep learning in medical image segmentation has gradually become a hot spot for researchers. In the same year, Ronneberger et al. [7] proposed U-Net for biomedical image segmentation, which showed excellent performance in the medical image segmentation field with its unique architecture of encoder-decoder, and since then, U-Net and its enhanced networks are frequently utilized in various medical image segmentation tasks. Liver tumor segmentation using deep learning methods can be classified into 2D, 3D, and 2.5D approaches, among which the 2D method requires the lowest performance of hardware devices during training and is the worst among the three methods in terms of training results, which is due to the fact that the 2D segmentation network does not fully utilize the three-dimensional information between the CT slices, resulting in the prediction of the model trained using the 2D method in predicting the results of the liver between consecutive slices and the liver tumor continuity. As a result, the model trained with the 2D method did not have enough continuity of the liver tumor between consecutive slices in the prediction results, i.e., it was not smooth enough and had a tearing sensation. To fully utilize the three-dimensional information from CT slices, in the training process of the liver tumor segmentation model, people try to use three-dimensional convolution kernels for feature extraction. Since the 3D convolution kernel can move in 3D space, it is able to make better use of the spatial feature information from the CT slices, so that the segmentation accuracy of the 3D technique is usually better than that of the 2D technique, and the degree of smoothing between the slices is also better than that of the 2D technique, the smoothness between slices is also higher. But the number of parameters in the 3D convolution kernel is greatly greater than that of the 2D convolution kernel, so when the equipment conditions are poor, because of the large number of parameters, the 3D method can often lead to memory overflow during the training process, and we have to preprocess the input CT image for cropping, which may lead to a reduction in segmentation accuracy. In order to solve the contradiction that 3D methods have a good

segmentation effect but high requirements on hardware resources, the compromise 2.5D method was later adopted. The 2D method relies completely on the intra-slice information and therefore cannot fully utilize the spatial information, and the idea of 2.5D is to use a few neighboring slices of the slice that are input to the network model as the segmentation network's input feature. Compared to 3D method, 2.5D has a smaller number of parameters but utilizes less spatial information than the 3D method. This paper classifies deep learning-based liver tumor segmentation methods into three groups: 2D, 3D, and 2.5D convolutional neural networks are based on the different dimensions of the segmentation data, at the same time the advantages and disadvantages of each type of network are summarized in detail. Finally, the future development direction is discussed. **Figure 2** illustrates the process of liver tumor segmentation based on deep learning methods.
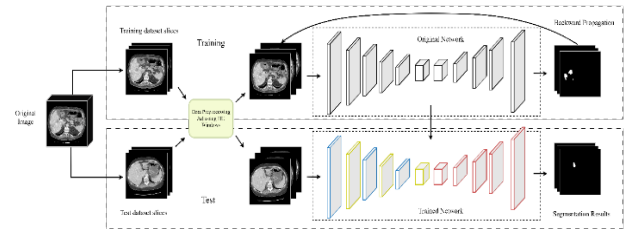

Figure 2. Segmentation flowchart for a deep learning approach

2. Datasets and evaluation indicators

2.1. Datasets

In order to advance medical image segmentation techniques and to provide an objective comparison of emerging liver tumor segmentation methods, in addition to the LiTS2017 [8] dataset published by MICCAI, commonly used datasets include 3DIRCADb (for comparing algorithms for 3D image reconstruction), ATLAS2023 [9], MICCAI 2015, Codala [10], TCGA-LIHC (The Cancer Genome Atlas Liver Hepatocellular Carcinoma) [11], and Midas, as well as several in-house datasets used by researchers.

The LiTS2017 dataset is the official dataset provided by the competition for the liver and liver tumor segmentation, jointly organized by two organizations, MICCAI and ISBI. The LiTS2017 dataset contains a total of 201 human CT scans. Out of these 201 CT images, 131 have manual annotation provided by the official, and the other 70 only have raw CT scans and are not manually annotated. The reason for this is that these 70 were used as a performance test dataset for the competition. The number of tumors per case ranged [0 ~ 75], with tumor sizes ranging from 38 cubic millimeters to 349 cubic millimeters. The data came from many institutions with different equipment, so these data vary somewhat in

imaging quality. The resolution of these slice samples ranged from 0.55 millimeters to 1.0 millimeters, the distance between slices ranged [0.45 millimeters ~ 6.0 millimeters], and the number of slices ranged [42 ~ 1026].

The 3DIRCADb dataset consists of the 3DIRCADb-01 dataset and the 3DIRCADb-02 dataset. The 3DIRCADb-02 dataset does not contain liver tumors, and 75% of the cases in the 3DIRCADb-01 dataset have liver tumors, with the number of tumors is between 1 and 46. And the image resolution was 512 × 512 pixels, the sizes were [16.3 ~ 24.9 cm, 12.0 ~ 18.6 cm, 11.0 ~ 20.2 cm], and the voxel sizes were [0.56 ~ 0.87 cubic millimeters, 0.56 ~ 0.87 cubic millimeters, 1.6 ~ 4.0 cubic millimeters].

The ATLAS2023 dataset consists of T1 CE-MRI liver scans from 90 patients with unresectable hepatocellular carcinoma, as well as 90 liver and liver tumor segmentation masks, divide into training and test datasets, with 60 and 30 patients in the training datasets and test datasets respectively. And the CE-MRI of the ATLAS2013 dataset comprises 3D images of the chest and abdomen in 44 to 136 transverse slices, covering the whole liver and tumor. The pixel pitch of each slice of 0.68 × 0.68 square millimeters to 1.41 × 1.41 square millimeters and a slice thickness of 2 millimeters to 4 millimeters. The ground truth of these CT images was created by an experienced MRI radiologist who manually outlined the liver and tumor contours on selected CE-MRI transverse slices.

The MICCAI 2015 dataset consists of 3631 slices segmented with corresponding labeling data. The voxels are [512 × 512 × 85 ~ 512 × 512 × 198] pixels. The CodaLab dataset comprises 131 contrast-enhanced images, each with a resolution of 512 × 512 pixels, and the voxel size interval of each CT image ranges from 0.64 to 0.84 millimeters. The TCGA-LIHC dataset consists of 1,688 patient cases. The data in MIDAS involves metastatic cysts and tumors of varying sizes, all at a resolution of 512 × 512 pixels. The TCGA-LIHC dataset contains 1,688 patient cases, and the data in MIDAS involve metastatic cysts and tumors of varying sizes, all at a resolution of 512 × 512 pixels.

In addition, a large number of researchers have used in-house datasets to validate segmentation performance. Drozdzal et al. [12] used a dataset containing 135 abdominal enhanced CTs; each image was segmented with the corresponding labeled data. 58 of these images had the corresponding labeling of the tumor portion. The images had an image resolution of 512 x 512 pixels, with the size of each pixel ranging between [0.53, 1.25 mm]. The gray value of each pixel is within [3000, 13500], and the slice thickness is between [0.5, 5.01 mm]. Sheba [13] obtained medical data from 2009–2014 containing 182 portal staging 2D CT scans, of which 53 images of cysts, 64 images of migrating tumors, and 65 images of hemangiomas. All slices were 512 × 512 pixels in size, with each pixel size ranging from [0.71 millimeters, 1.17 millimeters] and slice thicknesses ranging from [1.25 millimeters, 5 millimeters]. The dataset used by Roth et al. [14] consisted of 331 CT contrast-enhanced images with an image resolution size of 512 × 512 pixels. The number of

CT images in the test dataset ranged [460 ~ 1177]. The CT images in the training dataset consisted of 263 ~ 1061 slices. The dataset used by Sun et al. [15] was labeled by two radiologists from the First Hospital of Jilin University after acquiring multiphase CT images on a GE high-speed CT machine.

2.2. Evaluation indicators

In liver tumor segmentation tasks, we typically use the following evaluation metrics to assess model performance: DICE [16], VOE, Jaccard [17], ASSD, RMSD, MSSD, RVD and Accuracy. **Table 1** summarizes the common evaluation metrics listed above. Information on each metric includes the metric name, effect, range of values, frequency of use, and units.

Table 1. Evaluation indicators

Indicator name	Effect	Range of values	Unit
DICE	Duplication rate between results and labels	[0,1]	%
Jaccard	Similarities and differences between results and labels	[0,1]	%
VOE	The degree of overlap between the segmentation results and the ground truth, indicating the error rate	[0,1]	%
RVD	Difference in volume between segmentation results and labels	$(-\infty, +\infty)$	%
ASSD	Mean surface distance between the result and the symmetrical position between the labels	$(0, +\infty)$	mm
MSSD	Maximum surface distance between the result and the symmetrical position of the labeling tag	$(0, +\infty)$	mm
RMSD	The root mean square of the surface distance between the segmentation result and the symmetric position	$(0, +\infty)$	mm
Accuracy	Correctly predicted positive instances account for the sample of all positive instances predicted	[0,1]	%

3. 2D Network Segmentation Methods

The segmentation image data of the 2D CNN is a single 2D slice of the abdominal CT image, and the 2D network's segmentation results are obtained by the 2D convolution operation in the network. Finally, all the 2D network's segmentation results are superimposed to get the final results. The 2D network is characterized by low model complexity, fast operation speed, and superior segmentation performance, but its segmentation accuracy is limited by the discontinuity between image label layers. A typical representative of 2D networks is U-Net for biomedical image segmentation, proposed by Ronneberger

et al. [7] in 2015, which is improved and extended on the basis of FCN. U-Net contains contraction paths for capturing contextual information and symmetric expansion paths for precise localization. With data augmentation, the U-Net network will achieve more accurate segmentation with fewer training images. The structure of the U-Net network for liver tumor segmentation is shown in **Figure 3.**
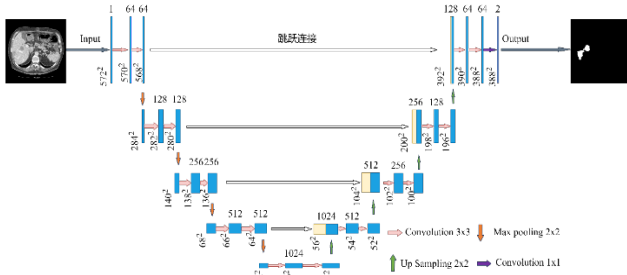
Figure 3. CT image segmentation method for liver tumor based on the U-Net network

The U-Net network is based on FCN expansion and modification, it is a classical fully convolutional network, this network splices the results of downsampling into the upsampling, which better preserves the position information of the original image, in the liver and its tumors segmentation task, the size, dimensions, and position of liver tumors puts high requirements on the training of the model, and U-Net solves this problem very well, therefore U-Net model is one of the most successful models in the task of medical image segmentation, many researchers have improved on this network, which has significantly facilitated medical image segmentation research [18]. The most common way is to use classical convolutional neural network backbones with pre-training parameters, such as VGG [19], ResNet [20], DenseNet, GhostNet [21], etc., instead of an encoder that implements migration learning [22].

Considering the problem of low resolution feature information duplication, the skip connection of U-Net was improved by Seo et al. [23], who proposed the mU-Net, combining high-level features related to the target to improve liver tumor segmentation. This network introduces residual paths with back-convolution and activation operations to the skip connection of the U-Net network to obtain high-level global feature information for small target input data and high-level features with high-resolution edge information for large target input data. In the case of small target inputs, the features in the skip connection and the residual path will not be fused, and the model can extract the global features better than U-Net. In order to efficiently transfer feature information from different layers, Wang et al. [24] proposed AFD-UNet, which is an adaptive fully dense neural network that connects different layers of the U-Net shared encoder and the corresponding decoder structure by adding horizontal

connection paths. This network can efficiently and adaptively utilize shallow and deep features, taking full advantage of the output results of each layer, and then automate learning. Tran et al. [25] found that most of the models based on U-Net ignored the output features of the convolutional units in the nodes, so they used it as a skip connection to provide more features for the decoder node and the next convolutional node, which improved liver tumor segmentation performance. Steven [26] and Xiao [27] et al. were inspired by residual structure and dense connection and replaced the coding and decoding units of U-Net with residual structures and dense connection modules to improve network feature extraction. Ghofrani et al. [28] extended U-Net by combining the advantages of ConvLSTM, Dense Convolution, and Residual Block, replacing the traditional convolution module with a recursive and residual approach in the encoding phase, adding dense connections in the fifth layer of the maximal receptive field, and improving the skip connections using ConvLSTM. Zhao et al. [29] improved the U-Net model by using the same padding after each convolution operation without changing the image scale or truncating the edges of the image. To alleviate the problem of the unknown depth of the optimal network, Zhou et al. [30] proposed UNet++, which is based on the traditional five-layer architecture of U-Net and integrates U-Net at different depths. UNet++ is able to collaboratively learn the optimal depth for the network at hand by using different depths of U-Net networks. Meanwhile, UNet++ redesigns the skip connection to interconnect multiple feature layers at the same scale two by two, and the sub-network of the decoder can fuse feature information at different semantic scales, thus realizing highly flexible feature fusion. Based on UNet++, Gao et al. [31] proposed a liver tumor segmentation method that is a nested U-Net's adaptive feature extraction method, which further improves gradient propagation and feature retention by combining UNet++ with extended dense short connections within convolutional blocks.

Another popular way to improve the U-Net network is to introduce attention mechanism between the encoder and decoder [32] to focus on the region of interest. For the first time, Oktay et al. [33] proposed Attention U-Net by combining the attention mechanism with U-Net and applying it to the task of pancreas segmentation. Attention U-Net has been demonstrated on multiple datasets to improve U-Net's performance. Li et al. [34] added the attention structure to UNet++ and redesigned the dense hopping connection, which uses the attention to enable the extraction of features at different layers in combination with the relevance of the task. Attention UNet++ speeds up the prediction of the network but also causes moderate performance degradation and shows good performance in liver tumor segmentation tasks. Wang et al. [35] found that U-Net could not fully exploit the useful feature information of the channel and could not make full use of the contextual feature information. So, they proposed an improved U-Net network with residual concatenation, channel attentional block, and hybrid extended attentional convolutional layers

that can accurately and efficiently perform different medical image segmentation tasks and perform well in liver tumor segmentation. Pang et al. [36] found that arbitrary superposition of feature maps makes CNNs mimic human cognition and visual attention in specific visual tasks very inconsistent. To alleviate the problem of CNNs lacking a reasonable feature selection mechanism, they developed a new efficient network, TANet, based on the U-Net architecture for liver tumor segmentation by embedding adaptive features in the tumor attention layer through multifunctional modules. Li et al. [37] proposed ANU-Net, which is a nested segmentation model attention mechanism, which employs a deeply supervised encoder-decoder structure, redesigned the dense skip connections, and introduced an attention mechanism to the nested convolutional blocks, to fuse features extracted from different layers with task-relevant decisions.

The multi-network liver tumor segmentation method is a medical image segmentation method with relatively excellent segmentation performance, the essence of which is to utilize multiple network architectures to achieve liver tumor CT image segmentation task. According to whether the multiple network architectures are taken in serial or parallel mode in the liver tumor segmentation method, they can be further classified into cascaded U-Net and dual-path U-Net. This method combines multiple underlying networks to increase feature extraction capability. Higher segmentation accuracy is obtained while maintaining lower computational costs and memory consumption. Gruber et al. [38] used two successively improved U-Nets; the first model was used to perform liver segmentation, and then the results of liver segmentation were fed into the second network for tumor segmentation. They used a hybrid loss function to combine the results of the two networks to further improve segmentation accuracy. Li et al. [39] proposed a network consisting of 2 parallel U-Nets, where one parallel branch is a coding U-Net and the other parallel branch is a segmentation U-Net, and trained the coding U-Net first to obtain the labeled features that contain the segmentation information (shape and position) of the encoding. This information is then used to train the segmentation U-Net. The goal is to preserve the segmentation features of the liver and liver tumor. To accurately segment liver tumors, Christ et al. [40] proposed cascaded U-Net+3D CRF to increase the segmentation accuracy of U-Net using 3D conditional random fields.

In addition to using the improved U-Net, researchers have found that combining the U-Net-based model with some nonparametric methods can further increase segmentation accuracy, and some of the commonly used nonparametric methods include level-set algorithms and graph-cutting algorithms. Alirr et al. [41] first used two U-Nets to segment livers and tumors of interest in the liver area, and then further refined the segmentation results using a local level set approach. Similarly, Zhang et al. [42] first used a U-Net based on 2D slices to coarsely position the liver, then refined the liver segmentation results using a 3D block-based FCN with coarsely localized tumors, and then further refined the tumor segmentation results using a novel level-set approach. To gain more advanced semantic feature information and reduce information loss, Liu et al. [10] proposed an improved model by increasing the depth of the U-Net, not replicating the activation features after convolution, and only replicating the pooling layer features when performing skip connections. Based on this improved model, combined with the graph-cut algorithm, they proposed a GIU-Net with better segmentation performance. In the graph-cut algorithm, based on the sequence context information and the output probability distribution graph of the improved U-Net, the graph-cut energy function is constructed. In summary, it can be seen that these methods are often based on the improved U-shaped network to gain preliminary segmentation results of liver tumors and then use nonparametric methods to obtain final fine segmentation results.

These improved 2D models show better performance in liver tumor segmentation tasks, and they also facilitate the design of data enhancement methods while maintaining lower memory requirements. However, due to the use of 2D convolutional kernels, they cannot capture spatial feature information along the z-axis, which may degrade the performance of volumetric segmentation.

4. 3D Network Segmentation Methods

Since medical images are spatially 3-dimensional, 2-dimensional neural networks cannot learn the spatial feature information of the 3-dimensional CT images in feature extraction. In contrast, the input feature of the 3-dimensional CNN is the whole volumetric CT image, and the whole segmentation result can be obtained at one time after the 3-dimensional convolutional operation in the network. The 3-dimensional CNN can sufficiently utilize the spatial information of the volumetric CT image and thus effectively solve the inter-layer discontinuity problem of the segmentation result of the 2-dimensional neural network.

A typical representative of 3D networks is V-Net for 3-dimensional medical image segmentation proposed by Milletari et al. [43] in 2016, which is different from U-Net in that V-Net is a 3D method and a new objective function based on Dice coefficients is used in V-Net to overcome the case of severe imbalance between foreground and background voxels, and its segmentation accuracy is much improved compared with that of U-Net with a great improvement. V-Net can predict the segmentation result of the whole volume image at one time after end-to-end training and is popularly used in medical image segmentation task with superior segmentation performance. After V-Net, researchers have proposed some improved networks based on V-Net. Since V-Net is a 3D method, its large number of parameters requires high hardware. In order to solve this issue, Lei et al. [44] designed a residual bottleneck module IRB (inverted residual bottleneck) instead of the ordinary convolutional block to construct the encoder and decoder, and the IRB block adopts deep convolution and point convolution

operations to decrease the voxels' number, while features can be fully extracted by decoupling cross-channel correction and spatial correlation. Accordingly, they proposed a lightweight LV-Net that uses 3D depth supervision in the training phase to improve the final loss function, thus better distinguishing between foreground and background regions and drastically reducing the number of voxels in the V-Net. Zhang et al. [45] proposed an improved V-Net with the combined use of region-based and the loss functions based on distance , which solves the problem of model performance degradation due to the high imbalance in the foreground and background voxels' numbers by training the model with the region-based loss function and three loss functions based on distance in the V-Net separately.

V-Net can be regarded as a 3D variant network of the U-Net network. Besides, there is another common 3D variant network of U-Net proposed by Cicek et al. [46],which is named 3D U-Net. Although they are both 3D variants of U-Net, the V-Net network is proposed for volumetric CT images and introduces residual connectivity, while 3D U-Net just replaces all the 2D convolution operations in U-Net with 3D convolution operations. Based on 3D U-Net, researchers have designed some improved networks. Mohagheghi et al. [47] found that using a hybrid loss function combining dice loss and data-driven loss (DDL) can improve the 3D U-Net's performance by integrating the a priori shape knowledge in DDL. In addition, this approach enhances the generalization ability and robustness of the hybrid network. In order to enable semantic features to adaptively change, Jin et al. proposed [48] the RA-UNet model, which introduces a stacked attention mechanism module on top of the 3D U-Net network and makes full use of the spatial feature information of the CT image by applying an attentional residual mechanism to perceive the features, accurately extracting the liver region and segmenting the tumor from the liver. The model is based on the U-shaped architecture and captures contextual feature by combining low-level and high-level feature maps. Most of these U-Net-based models just use the pairing information between samples and labels without using the information in the labels as supervised information; therefore, Song et al. [49] proposed a supervised BSU-Net network with bottleneck features for liver tumor segmentation task to improve the accuracy of the algorithm. The model consists of an encoding U-Net without skip connection and a segmentation U-Net with skip connection. The encoding U-Net is first trained as an auto-encoder to obtain the encoding of the ground truth mappings, which is then used as additional supervision to train the segmentation U-Net. Dou et al. [50] combined a conditional random field for conventional segmentation with a 3D segmentation approach to take full advantage of the benefits of each method. Similar to Dou et al., Lu et al. [51] combined the graph segmentation method of traditional segmentation with a 3D segmentation method. After them, other scholars have successively proposed some improved 3D

segmentation models around attention injection, convolutional improvement, and lightweighting.

In short, the 3D U-Net-based variational network can enhance the segmentation performance of liver tumor by using a hybrid loss function to better distinguish foreground and background and improve the network structure to extract more feature information. However, using 3D convolution operations requires a lot of memory during the calculation process, and training 3D networks often requires a lot of time and resources. In addition, the learning process of 3D networks often requires a sufficiently large dataset to fully converge the network.

5. 2.5D Network Segmentation Methods

2D convolutional neural networks are unable to utilize the interlayer continuity information of image labels, while 3D convolutional neural networks are limited by their large computational cost. To solve these problems, we propose a class of 2.5D convolutional neural networks. The input data for 2.5D networks are multiple adjacent slices of volumetric CT images to utilize the interlayer continuity information of the image labels. The 2.5D liver tumor segmentation structure is shown in **Figure 4**.
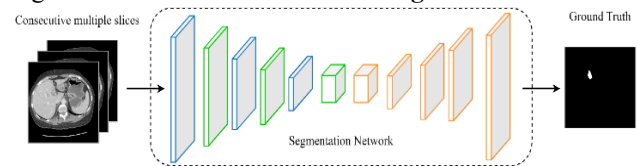
Figure 4. Segmentation method flow using three consecutive slices

The 2.5D networks generally contain 2D and 3D convolutional operations to achieve different functions. For example, Li et al.[52] proposed H-DenseUNet, a hybrid densely connected for liver tumor segmentation task. And this model consists of a 2D DenseUNet to obtain intra-slice features and a 3D DenseUNet for summarizing volumetric contexts. The hybrid feature information fusion layer optimizes both intra-slice representations and inter-slice features based on an automated context algorithm. In order to simultaneously utilize intra-slice semantic feature information and inter-slice continuity feature information to extract discriminative features, Wang et al. [53] proposed a 2.5D segmentation network, and this network consists of a multi-branch decoder for learning the features of a specific slice and an attention block for slice-centric, which is a densely connected dice loss function to normalize the intra-slice segmentation results to continuity. Zhang et al. [54] utilized a scaling approach to allow the segmentation network to focus only on useful localities, which reduces the parameters in the segmentation model and thus reduces the hardware resource requirements. Ben-Cohen et al. [55] changed the original FCN to a 2.5D FCN and introduced the idea of generative adversarial to

improve segmentation results. The Triplanar FCN based on FCN was proposed by Wang et al. [56] to take advantage of 3D spatial feature information and integrate the results in three dimensions. proposed Triplanar FCN by making full use of 3D spatial information to segment in each of the three dimensions and integrating the results. Ahn et al. [57] input three consecutively sliced images into the network as three channels and performed the segmentation task in the center region of the images. The encoder component is based on a modified Xception model that includes down-sampled layers and null-separable spatial pyramid pooling units, and the decoder part is a series of bilinear up-sampled layers connected to the encoder's skip connections. In order to attain a balance between computational cost and segmentation accuracy, and the utilization of 3D context information, Zhang et al. [58] designed a new 2.5D network that encodes the interlayer information in a 3D convolutional context and reconstructs the high-resolution result with 2D deconvolution. This structure can achieve effective multidimensional feature extraction without increasing the computational effort and increase the segmentation capability and efficiency of the model.

6. Summary and prospects

6.1. Summary

This paper classifies CNNs for liver tumor segmentation into three categories based on the input data, which has different dimensions: 2D, 3D, and 2.5D convolutional neural networks. 2D networks have low model complexity and fast operation speed, but their accuracy is limited by the discontinuity between image label layers; 3D networks can fully utilize the discontinuity between image label layers. The 3D network can fully utilize the continuity between the layers of image labels, but its model complexity is high and the computational cost is large, while the 2.5D network can fully utilize the continuity between the layers of image labels. A 2.5D network can not only utilize the continuity between the layers of image labels to improve segmentation but also maintain a lower model complexity and faster computational speed. **Table 2** details the characteristics of 2D, 2.5D, and 3D networks. The researchers have also proposed solutions to some common problems that may exist in different networks. For example, the issue of a severe imbalance between foreground and background in an image can be solved by using a hybrid loss function; more effective features are obtained by improving the network's structure, resulting in higher segmentation performance and improving the efficiency of the network.

Table 2. The characteristics of 2D,3D, and 2. 5D CNNs

Type	Input data	Advantages	Disadvantages
2D	Single Slice	Low model complexity, fast computing speed, and better segmentation results	Segmentation accuracy is limited by discontinuities between layers of image labels
2.5D	Multiple adjacent slices	Able to utilize the interlayer continuity of image labels, good segmentation effect; lower model complexity; faster computing speed	Compared to 3D networks, which can only utilize the interlayer continuity of image labels
3D	Whole volume image	Able to utilize the interlayer continuity of image labels with better segmentation performance	High model complexity, high computational cost, and more training data required

6.2. Prospects

This paper firstly discusses the necessity and superiority of deep learning especially U-Net network architecture applied to CT image segmentation of liver tumors, then classifies different types of convolutional neural networks, and finally summarizes the advantages and disadvantages of 2D, 3D, and 2.5D networks. In general, the current deep learning methods have superior performance, but there are some limitations and shortcomings, such as the small amount of training data, the imbalance between the number of background and foreground voxels in the training samples, as well as the inter-layer discontinuity of the image labels, the high complexity of the model, and the high computational cost in different types of networks.

The advantages and disadvantages of the existing CNNs-based methods applied to CT image segmentation of liver tumors are listed below as possible future research directions.

(1) To fully utilize the interlayer continuity information of image labels and to solve the issue of excessive computational cost in 3D networks, further research should be conducted on the improvement methods of 2.5D networks.

(2) For the problem of the high imbalance in the number of foreground and background pixels or voxels in an image, it should be further investigated the rational use of the hybrid loss function to better distinguish the foreground from the background.

(3) In order to extract more effective features to improve segmentation performance and reduce redundant operations to improve computational efficiency, improved methods of network structure, such as path connectivity and convolution operations, should be further investigated.

(4) New liver CT image segmentation methods based on deep learning were investigated, such as using segmentation methods to explicitly establish the continuity relationship between different parts by a local segmentation network based on probabilistic networks.

References

[1] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. J. A. I. R. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," vol. 50, pp. 497-537, 2018.

[2] S. J. J. o. C. Luo and Communications, "Review on the methods of automatic liver segmentation from abdominal images," vol. 2, no. 02, p. 1, 2014.

[3] Y.-D. Zhang *et al.*, "Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation," vol. 64, pp. 149-187, 2020.

[4] S. Wang *et al.*, "Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects," vol. 76, pp. 376-421, 2021.

[5] Y. Zhang *et al.*, "Deep Learning in Food Category Recognition," p. 101859, 2023.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015, pp. 234-241: Springer.

[8] P. Bilic *et al.*, "The liver tumor segmentation benchmark (lits)," vol. 84, p. 102680, 2023.

[9] F. Quinton *et al.*, "A Tumour and Liver Automatic Segmentation (ATLAS) Dataset on Contrast-Enhanced Magnetic Resonance Imaging for Hepatocellular Carcinoma," vol. 8, no. 5, p. 79, 2023.

[10] Z. Liu *et al.*, "Liver CT sequence segmentation based with improved U-Net and graph cut," vol. 126, pp. 54-63, 2019.

[11] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," vol. 26, pp. 1045-1057, 2013.

[12] M. Drozdzal *et al.*, "Learning normalized inputs for iterative estimation in medical image segmentation," vol. 44, pp. 1-13, 2018.

[13] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, 2018, pp. 289-293: IEEE.

[14] H. R. Roth *et al.*, "An application of cascaded 3D fully convolutional networks for medical image segmentation," vol. 66, pp. 90-99, 2018.

[15] C. Sun *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," vol. 83, pp. 58-66, 2017.

[16] L. R. J. E. Dice, "Measures of the amount of ecologic association between species," vol. 26, no. 3, pp. 297-302, 1945.

[17] P. J. N. p. Jaccard, "The distribution of the flora in the alpine zone. 1," vol. 11, no. 2, pp. 37-50, 1912.

[18] L. Chen *et al.*, "Liver tumor segmentation in CT volumes using an adversarial densely connected network," vol. 20, pp. 1-13, 2019.

[19] K. Simonyan and A. J. a. p. a. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[21] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1580-1589.

[22] P.-H. Conze *et al.*, "Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks," vol. 117, p. 102109, 2021.

[23] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. J. I. t. o. m. i. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," vol. 39, no. 5, pp. 1316-1325, 2019.

[24] E. K. Wang, C.-M. Chen, M. M. Hassan, and A. J. F. G. C. S. Almogren, "A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain," vol. 108, pp. 135-144, 2020.

[25] S.-T. Tran, C.-H. Cheng, and D.-G. J. I. A. Liu, "A multiple layer U-Net, U n-Net, for liver and liver tumor segmentation in CT," vol. 9, pp. 3752-3764, 2020.

[26] S. Guan, A. A. Khan, S. Sikdar, P. V. J. I. j. o. b. Chitnis, and h. informatics, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," vol. 24, no. 2, pp. 568-576, 2019.

[27] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-unet for high-quality retina vessel segmentation," in *2018 9th international conference on information technology in medicine and education (ITME)*, 2018, pp. 327-331: IEEE.

[28] F. Ghofrani, H. Behnam, and H. D. K. Motlagh, "Liver Segmentation in CT Images Using Deep Neural Networks," in *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 2020, pp. 1-6: IEEE.

[29] W. Zhao *et al.*, "Liver segemtation in CT image with no-edge-cuting UNet," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2020, vol. 9, pp. 2315-2318: IEEE.

[30] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. J. I. t. o. m. i. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," vol. 39, no. 6, pp. 1856-1867, 2019.

[31] Q. Gao, M. J. C. i. B. Almekkawy, and Medicine, "ASU-Net++: A nested U-Net with adaptive feature extractions for liver tumor segmentation," vol. 136, p. 104688, 2021.

[32] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention u-net for lesion segmentation," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, 2019, pp. 683-687: IEEE.

[33] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," 2018.

[34] C. Li *et al.*, "Attention unet++: A nested attention-aware u-net for liver ct image segmentation," in *2020 IEEE international conference on image processing (ICIP)*, 2020, pp. 345-349: IEEE.

[35] Z. Wang, Y. Zou, P. X. J. C. i. b. Liu, and medicine, "Hybrid dilation and attention residual U-Net for medical image segmentation," vol. 134, p. 104449, 2021.

[36] S. Pang, A. Du, M. A. Orgun, Y. Wang, and Z. J. N. N. Yu, "Tumor attention networks: Better feature selection, better tumor segmentation," vol. 140, pp. 203-222, 2021.

[37] C. Li *et al.*, "ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation," vol. 90, pp. 11-20, 2020.

[38] N. Gruber, S. Antholzer, W. Jaschke, C. Kremser, and M. Haltmeier, "A joint deep learning approach for automated liver and tumor segmentation," in *2019 13th International conference on Sampling Theory and Applications (SampTA)*, 2019, pp. 1-5: IEEE.

[39] S. LIa and G. K. Tsoa, "Bottleneck Supervised U-Net for Pixel-wise Liver and Tumor Segmentation."

[40] P. F. Christ *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: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 2016, pp. 415-423: Springer.

[41] O. I. J. J. o. A. C. M. P. Alirr, "Deep learning and level set approach for liver and tumor segmentation from CT scans," vol. 21, no. 10, pp. 200-209, 2020.

[42] Y. Zhang *et al.*, "Deep learning initialized and gradient enhanced level-set based segmentation for liver tumor from CT images," vol. 8, pp. 76056-76068, 2020.

[43] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565-571: Ieee

[44] T. Lei, W. Zhou, Y. Zhang, R. Wang, H. Meng, and A. K. Nandi, "Lightweight v-net for liver segmentation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1379-1383: IEEE.

[45] Y. Zhang, X. Pan, C. Li, and T. J. A. S. Wu, "3D liver and tumor segmentation with CNNs based on region and distance metrics," vol. 10, no. 11, p. 3794, 2020.

[46] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, 2016, pp. 424-432: Springer.

[47] S. Mohagheghi, A. H. J. I. j. o. c. a. r. Foruzan, and surgery, "Incorporating prior shape knowledge via data-driven loss model to improve 3D liver segmentation in deep CNNs," vol. 15, pp. 249-257, 2020.

[48] Q. Jin, Z. Meng, C. Sun, H. Cui, R. J. F. i. B. Su, and Biotechnology, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," vol. 8, p. 605132, 2020.

[49] L. Song, K. Geoffrey, and H. J. E. S. w. A. Kaijian, "Bottleneck feature supervised U-Net for pixel-wise liver and tumor segmentation," vol. 145, p. 113131, 2020.

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

[51] F. Lu, F. Wu, P. Hu, Z. Peng, D. J. I. j. o. c. a. r. Kong, and surgery, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," vol. 12, pp. 171-182, 2017.

[52] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. J. I. t. o. m. i. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," vol. 37, no. 12, pp. 2663-2674, 2018.

[53] S. Wang *et al.*, "Conquering data variations in resolution: A slice-aware multi-branch decoder network," vol. 39, no. 12, pp. 4174-4185, 2020.

[54] L. Zhang and S. C.-H. J. a. p. a. Yu, "Context-aware PolyUNet for liver and lesion segmentation from abdominal CT images," 2021.

[55] A. Ben-Cohen *et al.*, "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," vol. 78, pp. 186-194, 2019.

[56] Z. Wang and G. J. I. J. o. P. E. Wang, "Triplanar convolutional neural network for automatic liver and tumor image segmentation," vol. 14, no. 12, p. 3151, 2018.

[57] Y. Ahn *et al.*, "Deep learning algorithm for automated segmentation and volume measurement of the liver and spleen using portal venous phase computed tomography images," vol. 21, no. 8, p. 987, 2020.

[58] C. Zhang, Q. Hua, Y. Chu, P. J. C. i. B. Wang, and Medicine, "Liver tumor segmentation using 2.5 D UV-Net with multi-scale convolution," vol. 133, p. 104424, 2021.