Precise Ablation Zone Segmentation on CT Images after Liver Cancer Ablation using Semi-automatic CNN-based Segmentation

Quoc Anh Le[1], Xuan Loc Pham[2], Theo van Walsum[3], Viet Hang Dao[4,5], Tuan Linh Le[6], Daniel Franklin[7], Adriaan Moelker[3]†, Vu Ha Le[1,2], Nguyen Linh Trung[1], Manh Ha Luu[1,2,3]

[1] AVITECH, VNU University of Engineering and Technology, Vietnam
[2] Department of Radiology and Nuclear Medicine, Erasmus MC, the Netherlands
[3] FET, VNU University of Engineering and Technology, Vietnam
[4] Internal Medicine Faculty, Hanoi Medical University, Vietnam
[5] The Institute of Gastroenterology and Hepatology, Vietnam
[6] Diagnostic Imaging and Interventional Radiology Center, Hanoi Medical University Hospital, Vietnam
[7] School of Electrical and Data Engineering, University of Technology Sydney, Australia

Version typeset May 27, 2024

Corresponding author: Manh Ha Luu. E-mail: halm@vnu.edu.vn
Corresponding address: 707 E3, 144 Xuan Thuy road, Cau Giay district, Hanoi, Vietnam
Postal code: 100000

Abstract

Background: Ablation zone segmentation in contrast-enhanced computed tomography (CECT) images enables the quantitative assessment of treatment success in the ablation of liver lesions. However, fully-automatic liver ablation zone segmentation in CT images still remains challenging, such as low accuracy and time-consuming manual refinement of the incorrect regions.

Purpose: Therefore, in this study, we developed a semi-automatic technique to address the remaining drawbacks and improve the accuracy of the liver ablation zone segmentation in the CT images.

Methods: Our approach uses a combination of a CNN-based automatic segmentation method and an interactive CNN-based segmentation method. Firstly, automatic segmentation is applied for coarse ablation zone segmentation in the whole CT image. Human experts then visually validate the segmentation results. If there are errors in the coarse segmentation, local corrections can be performed on each slice via an interactive CNN-based segmentation method. The models were trained and the proposed method was evaluated using two internal datasets of post-interventional CECT images ($n_1 = 22$, $n_2 = 145$; 62 patients in total) and then further tested using an external benchmark dataset ($n_3 = 12$; 10 patients).

Results: To evaluate the accuracy of the proposed approach, we used Dice Similarity Coefficient (*DSC*), average symmetric surface distance (*ASSD*), Hausdorff Distance (*HD*), and volume difference (*VD*). The quantitative evaluation results show that the proposed approach obtained mean *DSC*, *ASSD*, *HD*, and *VD* scores of 94.0%, 0.4 mm, 8.4 mm, 0.02, respectively, on the internal dataset, and 87.8%, 0.9 mm, 9.5 mm, and -0.03 respectively, on the benchmark dataset. We also compared the performance of the proposed approach to that of five well-known segmentation methods; the proposed semi-automatic method achieved state-of-the-art performance on ablation segmentation accuracy, and on average, 2 minutes are required to correct the segmentation. Furthermore, we found that the accuracy of the proposed method on the benchmark dataset is comparable to that of manual segmentation by human experts ($p = 0.55$, t-test).

Conclusions: The proposed semi-automatic CNN-based segmentation method can be used to effectively segment the ablation zones, increasing the value of CECT for assessment of treatment success. For reproducibility, the trained models, source code, and demonstration tool are publicly available at https://github.com/lqanh11/Interactive_AblationZone_Segmentation.

Contents

I. Introduction

Liver cancer has a high mortality rate, and its incidence increases yearly[1]. According to statistics from GLOBOCAN 2020, liver cancer ranked fifth in the number of new cases and third in the number of cancer deaths[2]. Early detection and treatment of liver cancer are crucial for improving treatment outcomes[3]. Thermal ablation such as radiofrequency ablation (RFA) and microwave ablation (MWA) are considered as curative treatment options for patients with early-stage liver cancer can not undergo an open surgical procedure[4] and can be performed with minimal discomfort for the patient, and with a short recovery time[5]. RFA and MWA also have a low risk of complications compared to surgery and a low risk of side effects compared to chemotherapy and radiation therapy[6].

During the procedure, the interventionist uses ultrasound (US) or CT to guide the insertion of a thin needle through the patient's skin and into the target lesion[7]. Once the needle tip is placed in position, heat is generated at the needle tip to destroy the malignancy by creating a region of cell destruction, also known as the ablation zone. A major drawback of RFA and MWA is the high recurrence rate after treatment, especially in the local ablation site. It has been reported that ablation of tumors with a size ranging from 2 to 5 cm in diameter results in a recurrence rate of 26.4%[8].

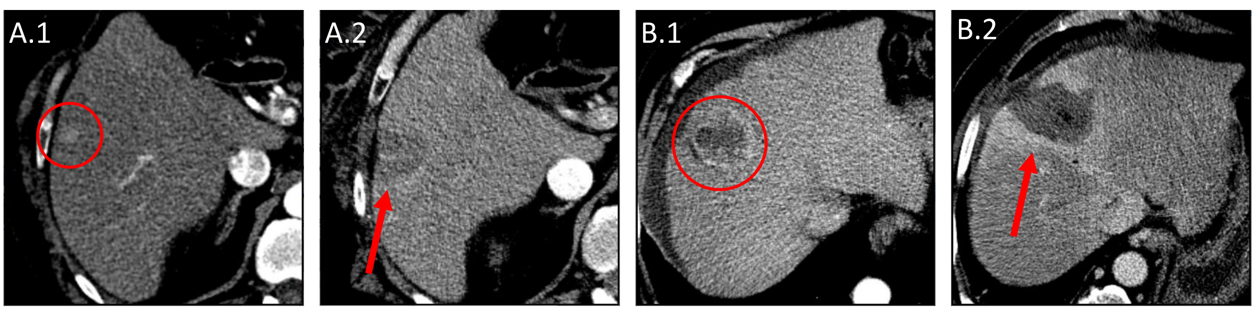

Figure 1: The liver tumor (red circle) in the pre-intervention CT image (A.1 and B.1) and its corresponding ablation zone (red arrow) in the post-intervention image (A.2 and B.2).

To evaluate the ablation, a CECT scan is usually performed at the end of the intervention to visualize the ablation zone (see Figure 1)[9]. By assessing the ablation zone in the CECT image, the physician can determine whether the procedure was completely successful, whether additional ablation needs to be performed, and what the safety margins of the ablation zone correspond to the lesion[10]. It is especially important to accurately assess the margins of the ablation zone. In current clinical practice, this assessment is performed visually by the interventional radiologist. Precise segmentation of the ablation zone in the CECT image enables quantitative assessment and may provide improved confidence in the outcome.

While manual segmentation of the ablation zone is tedious and time-consuming and thus is not feasible in clinical practice, computer-aided methods can be used to greatly reduce the required cognitive load. The main challenge of precise ablation zone segmentation is that the post-interventional CT images are noisy, with inhomogenous intensities inside the ablation zone. In addition, for clinical use, the segmentation processing time should be sufficiently fast (in order of minutes). Although there have been many studies developing methods for liver lesion segmentation[11], only a few studies have focused on ablation zone segmentation. Egger et al. (2015)[12] presented a semi-automatic method, based on an interactive, graph-based contouring approach, for segmenting the ablation zone in twelve post-interventional CECT images. The obtained segmentation is compared to manual segmentations performed by two medical experts, achieving a mean Dice Similarity Coefficient (*DSC*) score of 77%. Wu et al. (2021)[13] applied a region-growing method to segment the ablation zone in the CT image, followed by fuzzy c-mean clustering and then refined by cyclic morphological processing, which achieved the mean *DSC* core of 75%. Recently, deep learning-based methods have been applied to segment the ablation zone. He et al. (2021)[14] used a multi-scale patch-based 3D Residual Attention U-Net (RA-UNet) to segment the ablation zone in multiphase CT images, achieving median *DSC* scores of 83% and 89% for arterial and portal venous phase, respectively. It was reported that the hepatic enhancement in the arterial phase is not sufficient to discriminate the ablation zone's edge, resulting in the difficulty for the segmentation. Anderson et al. (2022)[15] investigated Hybrid-WNet to segment the ablation zone in CECT images, reporting a median *DSC* score of 79% and a median surface distance of 0.76 mm. From the reported state-of-the-art results, it is clear that precise segmentation of the ablation zone in CT images remains a challenging problem. In addition, to the best of our knowledge, none of the previous studies evaluated the methods on external datasets. Therefore, the purpose of this study is to propose and assess an effective method for precise ablation zone segmentation. In addition, we will evaluate the method using both internal and external datasets to verify the performance of the method.

In the last decade, deep learning-based methods, especially Convolutional Neural Networks (CNN) and Transformers, have demonstrated their indisputable effectiveness across a variety of fields including medical image analysis. Ronneberger et al. (2015)[16] introduced U-Net, which consists of a combination of encoder-decoder structures for efficient automatic segmentation of medical images. Since then, numerous variants of U-Net have been proposed, such as Residual U-Net[17], which integrates the residual path into the original structure, and H-DenseUNet[18], which combines 2D and 3D Dense U-Net. Isensee et al (2021)[19] further demonstrated the effectiveness of U-Net in medical image segmentation via the release of nnU-Net, which has become a well-known platform for automatic training and organ segmentation pipelines. Recently, the Vision Transformer technique (ViT) proposed by Dosovitskiy et al. (2020)[20] demonstrated its potential in computer vision, and is now being applied to the problem of medical image segmentation. For instance, UNETR[21] leverages the U-shaped encoder-decoder architecture but replaces the CNN-based encoding branch with ViT. Chen et al. (2021)[22] also realize the potential of the ViT-based encoder scheme and introduce TransUNet, but instead of a pure ViT encoding branch, the authors propose a hybrid architecture where a multi-resolution CNN is initially utilized to produce input feature maps for the ViT encoding block. CoTr[23] further improves the hybrid architecture in TransUNet by fully leveraging the multi-scale feature maps from the CNN instead of only using the lowest resolution. Subsequently, these multi-scale feature maps are flattened and encoded by Deformable ViT, which greatly reduces the complexity of ViT and therefore significantly boosts its performance.

Unlike automatic segmentation methods, *semi-automatic* segmentation methods require some level of human interactions to complete the segmentation procedure. The interaction can be point-clicking, scribbling, rectangle/circle initial drawing, and manual tuning of parameter values. Several semi-automatic segmentation methods have been developed for medical image segmentation[24]. Region-growing is one of the most popular semi-automatic segmentation methods[25]. Region-growing is initialized by a seed point with a predefined threshold interval and then expands within a connected region. The contrast between the object and the background acts as an important factor for a successful segmentation. Chan and Vese (2001)[26] introduced an interactive segmentation method based on a level-set algorithm, where the user provides an initial contour around the object. The method can successfully segment objects without clear boundaries. Grab-cut[27] is a well-known interactive segmentation method in which the user provides source and sink regions via manual interactions. Furthermore, several conventional methods such as the Robust Statistics Segmenter[28], Otsu & Picking[29], and Geodesic Segmenter[30] were investigated for interactive segmentation of objects in medical images.

Recently, interactive CNN-based segmentation methods have been investigated and shown to outperform traditional interactive methods, achieving higher accuracy with fewer user interactions. DeepCut[31] and ScribbleSup[32] are among the first interactive CNN-based segmentation frameworks which embed user-provided bounding boxes or scribbles into CNN models. DeepIGeoS[33] performed interactive segmentation by using geodesic distance transforms of scribbles as additional input channels to CNNs. Furthermore, Wang et al. (2018)[34] proposed an image-specific fine-tuning method to make a CNN model adaptive to a specific test image. Although promising results have been reported, the method still has some limitations. For example, it requires the user to provide a bounding box for the object and scribbling on the background and foreground for the correction which may be inconvenient in practice. Luo et al. (2021)[35] introduced MIDeepSeg, a minimally interactive segmentation method based on a CNN and exponentiated geodesic distance, to segment both seen and unseen objects appearing in the training dataset. However, the framework requires the user to click several points at the edges of the object on each slice the object presents to create a ROI for the object, which may be inconvenient to use in applications with tight timing constraints. Sun et al. (2022)[36] proposed a graph convolutional neural network for segmentation tasks; however, the requirement of clicking points at the boundaries or dragging predicted points is also not well-suited to time-critical tasks. Sofiiuk et al. (2022)[37] proposed a click-based interactive segmentation, Reviving Iterative Training with Mask Guidance (RITM), that iteratively uses the differences of the previous prediction segmentation and the ground truth to provide additional prior information to train the model and improve segmentation prediction accuracy.

Generally, fully automatic segmentation methods are convenient and fast for global segmentation of the entire image. However, they frequently have small segmentation errors that need to be manually corrected. Theoretically, for CNN-based approaches, the more data involved in training the deep learning models, the more accurate the model can become. However, it is not clear how much data should be used for a specific medical image segmentation application. Furthermore, collecting large amounts of data in the medical image field with accurate segmentations is challenging. In contrast, a human expert can control an interactive method to segment a region correctly. Nevertheless, using interactive segmentation methods to fully segment a structure may not be practical because of the excessive amount of interactions required. In addition, we hypothesize that clicking points inside a region is more convenient than clicking points at the edges, or scribbling using the mouse. *Therefore, our key idea to solve the problem of ablation zone segmentation is to utilize an interactive CNN-based segmentation method in which the user clicks points at the incorrect segmentation regions to refine the segmentation provided by an automatic segmentation method.*

The overview of the remainder of the paper is as follows: Section II. describes the proposed interactive CNN-based segmentation method in detail. Subsequently, Section III. describes extensive experiments to assess the performance of the proposed solution. Next, the experimental results are discussed in Section IV.. Finally, Section V. summarizes the findings of this study.

II. Methods

In order to precisely segment the ablation zone in a CECT image, our strategy is to combine automatic segmentation with interactive segmentation. Firstly, the automatic segmentation method is applied to segment the ablation zone in the whole CT volume. Next, a human expert reviews the automatic segmentation and then uses RITM[37] as a CNN-based interactive segmentation method to fix the incorrect segmentations via clicking points at the local error locations. Finally, the segmentations are combined by mixing the probability maps at the local location of the two methods. The underlying assumption is that the more accurate the automatic segmentation is, the fewer human interactions are needed for error correction. In this study, we evaluate several automatic segmentation frameworks for the initial segmentation in Section III.D.1.. In addition, the interactive method enables click-based segmentation which is one of the simplest interaction types. The pipeline of the proposed approach is illustrated in Figure 2. The following sections describe each component in the pipeline in detail.

II.A. Automatic segmentation

In the first step of the proposed pipeline, an automatic segmentation network is utilized to segment the ablation zone from the CT image. In this study, we evaluate well-known CNN-based and Transformer-based networks to find a suitable one, aiming for fast inference time and high accuracy. Here are descriptions of four segmentation methods.

- 3D U-Net, introduced by Cciccek et al. (2016)[38], is an extension of U-Net architecture designed to segment objects in 3D data by processing them with corresponding 3D operations. 3D U-Net consists of an encoder-decoder structure with a skip connection to capture high-level and low-level features of the 3D image and produce full-resolution segmentation. The 3D U-Net has been extensively used in various medical image segmentation tasks.

- UNETR, proposed by Hatamizadeh et al. (2022)[21], leverages the potential of ViT in sequence representation learning, making it highly effective in segmenting objects in images. UNETR utilizes the strength of the U-Net architecture but replaces the CNN-based encoder with the Transformers-based encoder. The original image is split into 3D patches and a linear projection of these patches is applied to produce the input for the ViT-based encoder. UNERT has demonstrated state-of-the-art performance in several medical image segmentation tasks.

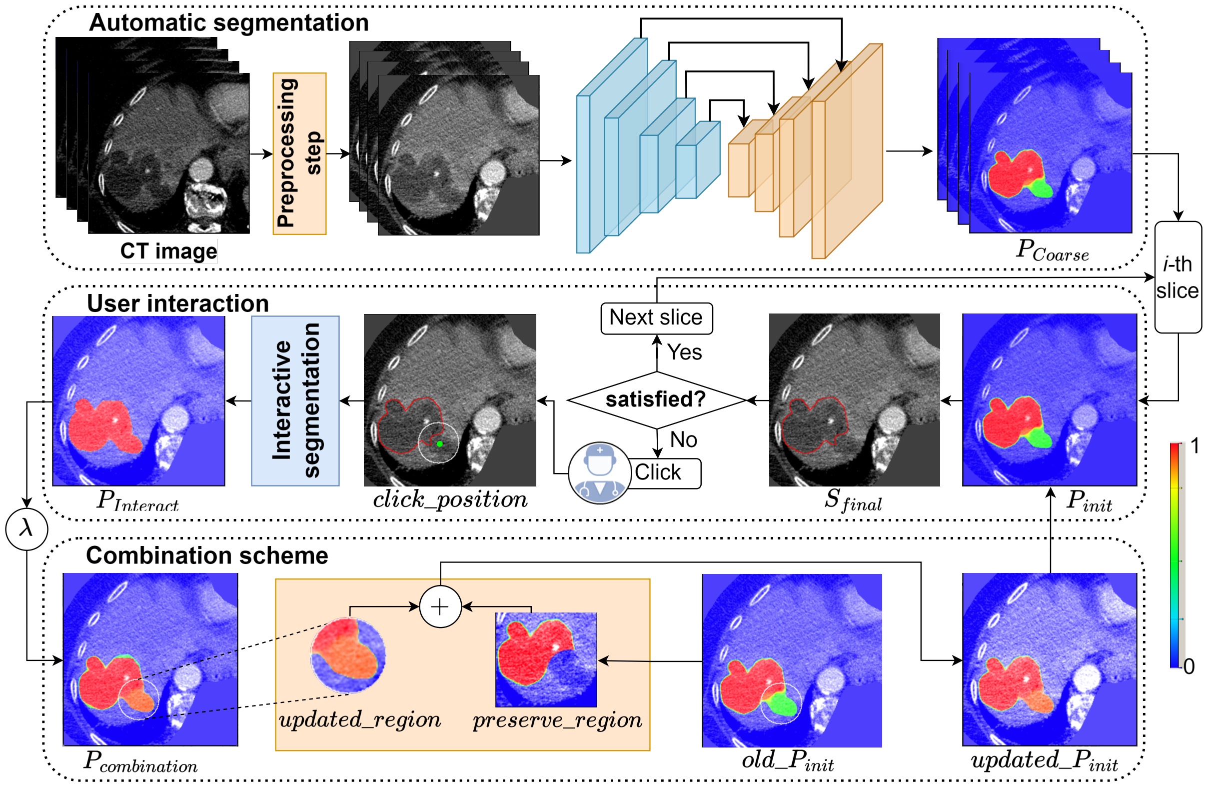

Figure 2: The pipeline of the proposed approach for the semi-automatic ablation zone segmentation. The 3D coarse segmentation is predicted using nnU-Net, which is then locally corrected in each slice by the user using the interactive segmentation via the combination scheme. The probability map, the output of segmentation models, overlapped with the CT image and visualized by a color map. The color bar indicates the probability prediction value of the segmentation. The white circle marks the region of interest.

- nn-UNet, developed by Isensee et al.(2021)[19], focuses on optimizing and improving the performance of the U-Net architecture by introducing various enhancements. nn-UNet employs novel data augmentation techniques, training strategies, and model configurations to achieve better segmentation results.

- CoTr, proposed by Xie et al. (2021)[23], also consists of an encoder-decoder structure like other segmentation networks. In the encoder part, CNNs and Transformers are used in the CoTr's architecture. The CNN's role is to extract feature representations from the input image. Then, the deformable Transformer (DeTrans) is used to model long-range dependencies within the extracted feature maps. Combining the strengths of CNNs and Transformers, CoTr addresses complex tasks requiring local and global context understanding in image analysis.

II.B. Interactive CNN-based segmentation

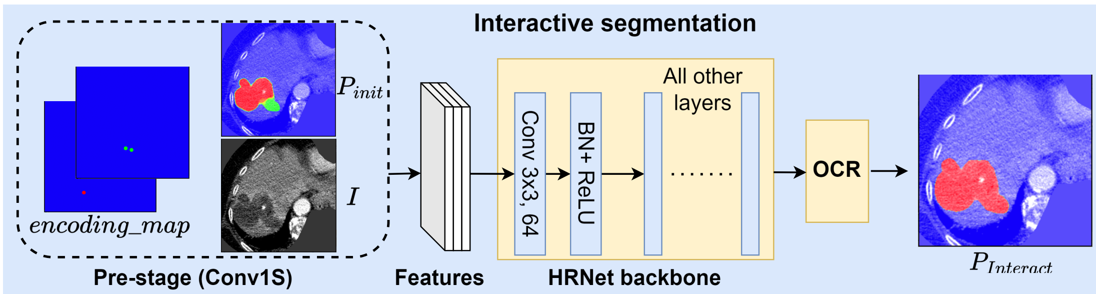

Figure 3: The architecture of RITM. User click points are encoded in binary disks. The positive and negative click points are shown in green and red in the *encoding_map*, respectively.

We aim to use slice-based segmentation for the correction as in clinical routine: the medical expert needs to check every single axial slice containing the ablation zone after the treatment[12].

As post-interventional CT images are characterized by inhomogeneous intensities and noise in the ablation zone (see Figure 1), traditional methods may not perform well in such conditions. It has been demonstrated that CNN-based segmentation methods are able to deal with inhomogeneous regions and noise[39]. Therefore, we use RITM[37], a 2D interactive CNN-based segmentation method, to revise the prediction of the automatic segmentation and obtain the final ablation zone segmentation. The RITM consists of two parts: the click encoding block and the backbone, as shown in Figure 3. The idea of the interactive segmentation network is to encode the user clicks and feed them into the network's backbone to generate a prediction. We encode the user click in a binary disk, which achieved effective performance compared to other click encoding schemes[40].

As a segmentation network, RITM also contains a semantic segmentation backbone. In this study, we choose High-Resolution Net combined with Object-Contextual Representations (HRNet+OCR) as the backbone of the interactive segmentation method. The HRNet+OCR is a promising architecture specifically designed for producing high-resolution outputs[34]. The HRNet+OCR backbone model was pre-trained using the ImageNet dataset, meaning the backbone's input is a three-channel image. However, in the interactive segmentation, the input includes additional features such as a guided mask and the click-encoding map. To adapt the pre-trained model, we employ a convolution block known as Conv1S, introduced by Sofiiuket al. (2022)[37]. The architecture of Conv1S is designed to ensure that the output feature's channel matches the input of the backbone's first convolutional layer (64 channels).

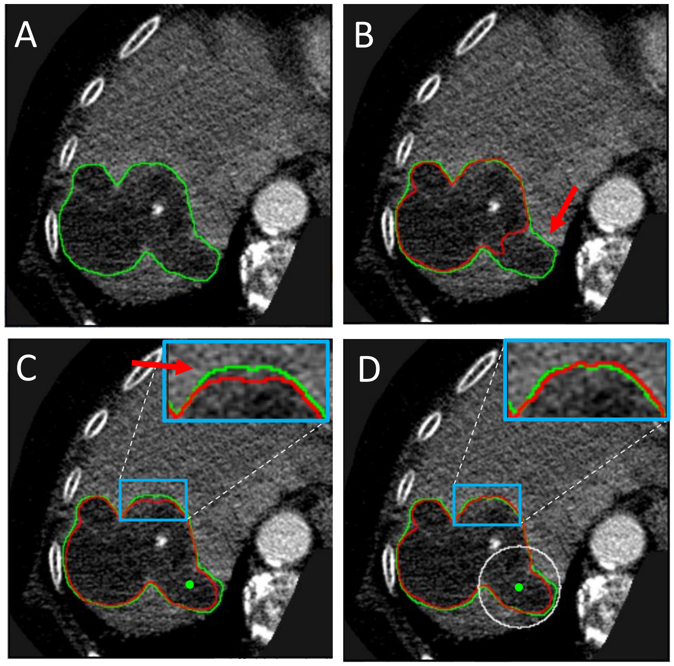

Figure 4: Illustration of interactive segmentation using the click-based semi-automatic method w and w/o the combination scheme: the original image with ground truth segmentation (green contour) of an ablation zone (A), the segmentation (red contour) obtained from the automatic model (B), the segmentation (red contour) obtained from the original interactive model (C), and the segmentation (red contour) obtained using the proposed combination scheme (D). The green dot represents the user's click. The white circle is the ROI. The red arrows mark the mislabeled regions.

II.C. Combination scheme

Since the segmentation obtained from the automatic model might have mislabeled regions (see Figure 4.B), the interactive model is applied to revise the prediction of the automatic segmentation. Nevertheless, the segmentation obtained from the interactive model may also have mislabeled regions far from the clicked points (see Figure 4.C). Consequently, we combine the two CNN-based segmentations using their probability predictions with a spatial constraint from the clicked points. The main idea is to construct a weighted voter between the two probability predictions within the regions of interest (ROIs), assuming that the user clicks inside the incorrect segmentation regions. The pseudo-code of the combination scheme is shown in Algorithm 1.

Firstly, the automatic segmentation network coarsely predicts the ablation zone from the CT slice $I \in \mathbb{R}^{512\times512}$ to obtain the probability prediction $P_{coarse} \in \Omega^{512\times512}, \Omega \triangleq [0,1]$ which is then defined as the initial prediction P_{init} for the combination scheme. Next, the initial prediction is thresholded to obtain the final segmentation $S_{final} \in \Theta^{512\times512}, \Theta \triangleq \{0,1\}$. If the segmentation is not satisfying, the

Algorithm 1 Combination scheme

Input: CT slice: I, Coarse prediction: P_{coarse}, Kernel size: K, Weighted value: λ, Threshold value: $thrsh$
Output: Final segmentation: S_{final}
$P_{init} \leftarrow P_{coarse}$
$S_{final} \leftarrow P_{init} > thrsh$
while S_{final} is unsatisfied **do**
    / User define a click in the mislabeled region */*
    $encoding\_map$, $click\_positon \leftarrow$ USER_CLICK()
    $P_{interact} \leftarrow$ PREDICTOR(I, P_{init}, $encoding\_map$)
    $cirle\_mask\_1$,$cirle\_mask\_0 \leftarrow$ CRICLE($click\_position, K$)
    / Update P_{init} */*
    $P_{combination} \leftarrow \lambda \times P_{interact}$ + (1-λ) $\times P_{init}$
    $updated\_region \leftarrow P_{combination} \times cirle\_mask\_1$
    $preserve\_region \leftarrow P_{init} \times cirle\_mask\_0$
    $P_{init} \leftarrow updated\_region + preserve\_region$
    / Thresholding to get S_{final} */*
    $S_{final} \leftarrow P_{init} > thrsh$
end while

user corrects the mis-segmented regions using a negative point click to correct a false positive region or a positive point click to correct a false negative region. Based on the clicked points, the interactive segmentation network generates a probability prediction (PREDICTOR), which refers to the interactive prediction $P_{interact} \in \Omega^{512\times512}$. In addition, we define ROIs from the positions of user clicks by dilating the clicking positions with a kernel size K (CIRCLE). The initial probability prediction is then updated by combining the previous initial probability prediction P_{init} and the interactive probability prediction $P_{interact}$ with a weighted parameter λ within the local correcting ROIs. Subsequently, the ablation zone segmentation is corrected only in the ROIs. To this end, the process is repeated until the final ablation zone segmentation S_{final} is satisfied. The effect of weighted parameter λ and kernel size K on the segmentation accuracy will be assessed in Section III.D.2..

III. Experiments and results

III.A. Datasets

III.A.1. The details of the datasets:

This study involved three datasets, comprising a total of 179 contrast-enhanced CT scans, from two medical centers.

The first dataset, denoted as EMC$_A$ dataset, includes CT scan images (arterial phase) from 22 patients who underwent ablation treatment of liver lesions. The CT scans were acquired at Erasmus MC using Siemens CT scanners and reconstructed with an axial matrix size of 512×512, with a pixel spacing of 0.8 mm, a slice thickness of 3 mm, and the number of slices ranged from 40 to 70 slices.

The second dataset, referred to as EMC$_B$ dataset, was reused from our prior study[41] and consists of 145 multiphase CT scans from 40 patients who underwent ablation treatment of liver cancer at Erasmus MC. The CT scan was acquired while patients were in the intra-intervention room and patients went to the medical center for the follow-up procedure. The CT scans were reconstructed with an axial matrix size of 512×512, with the slice thickness ranging from 0.4 to 5 mm, the pixel spacing ranging from 0.59 to 0.98 mm, and the number of slices ranging from 19 to 672 slices.

The third dataset, the Benchmark dataset, was obtained from the Medical University of Leipzig, Saxony, Germany and included 12 CT scans from 10 patients[12]. The CT scans are acquired in the portal venous phase, reconstructed with an axial matrix size of 512×512, pixel spacing ranging from 0.68 to 0.78 mm, slice thickness ranging from 1 to 2 mm, and the number of slices ranging from 52 to 232.

Each CT scan contains one to three ablation zones. The mean diameter of the ablation zones in the training set was 55 $\pm$ 18.6 mm. In 12 CT scans in the training set, the needle is visible. In the validation set, the mean diameter of the ablation zones is 54.6 $\pm$ 23.8 mm. Three CT scans in the validation set contain the needle. In the EMC testing set, the mean diameter of the ablation zones was 54.5 $\pm$ 18.4 mm. The needle was visible in seven CT scans in this set. For the Benchmark dataset, the mean diameter of the ablation zones was 62.2 $\pm$ 16.3 mm. The needle is visible in six CT scans in this dataset.

The specifics of data division can be found in Table 1.

Table 1: Number of CT volumes and slices for training, validation, and testing used in this study. The numbers in parentheses are the number of 2D slices.

Dataset		# Patients	Arterial	Portal venous	Total
Training	EMC$_A$	18	18	-	18
	EMC$_B$	29	36 (5638)	35 (5379)	71 (11017)
Validation	EMC$_A$	4	4	-	4
	EMC$_B$	18	5 (862)	13 (1353)	18 (2215)
Testing	EMC$_B$	11	31 (3006)	25 (3141)	56 (6147)
	Benchmark	10	-	12 (1525)	12 (1525)

III.A.2. Annotation and ground-truth

For the ablation zone of EMC_A and EMC_B datasets, a technician manually created segmentations using Mevislab version 3.4.3. Next, the segmentations were corrected/verified by a physician, serving as the ground truth. In the Benchmark dataset, the ground truth of the ablation zones was obtained from two medical experts who manually drew segmentations using Mevislab[12].

III.A.3. Preprocessing

In the preprocessing step of the CT scan, we limited the prediction range to the liver region. Firstly, liver segmentation is automatically acquired using a nnU-Net segmentation network, which has been trained on the LiTS dataset[11]. Then, the liver segmentation is dilated with a kernel size of (30, 30, 1). The dilated liver segmentation is used as a mask for the ablation zone segmentation.

In the interactive segmentation, we truncated the CT image using a HU range, which is defined in Section III.D.2. for an optimal value. Then, we converted the clipped CT images into a three-channel format using the *OpenCV* library for input into the interactive segmentation model.

The preprocessing step of the automatic segmentation methods is performed using the default settings described in the original articles. For nnU-Net, nnU-Net (fine-tuning), and CoTr, these models are implemented within the nnU-Net framework. Consequently, they share similar preprocessing steps, which are automatically determined based on the characteristics of the training dataset. Firstly, all data is cropped to retain only the region containing nonzero values. Subsequently, the data is resampled to the median voxel spacing of the entire training dataset, where third-order spline interpolation and nearest interpolation are applied for the image data and segmentation label, respectively. Following resampling, the data is normalized by clipping intensity values to the [0.5, 99.5] percentiles of the entire training dataset's intensity range. This is based on z-score normalization, which is computed based on the mean and standard deviation of all intensity values collected from the dataset.[42]. For 3D U-Net and UNETR, the models are implemented using the MONAI library. The preprocessing steps include various data transformations, which are resampled to a fixed spacing, clipped, and scaled the intensity into the range of 0 to 1[43]. Since the training samples are not extensive, data augmentation is also applied to the data to prevent the overfitting problem, which are random crop, random rotate, random flip, and random elastic deformations[42].

III.B. Hardware and implementation in details

III.B.1. Implementation of the CNNs

This study is conducted using a Ubuntu 20.04 workstation, with an Intel® Core™ i9-10900K CPU, 64GB RAM, RTX 8000 GPU. The source code is implemented in Python 3.6 with Pytorch 1.10 integrated with CUDA 11.3. The implementations of automatic segmentation models and RITM are based on their authors' GitHub repositories [a] [b] [c] [d].

We implemented well-known CNN-based and Transformer-based networks, including 3D U-Net[38], nnUNet[19], UNETR[21], and CoTr[23] for automatic segmentation performance comparison. We trained CNN-based and Transformer-based models with the training EMC$_B$ dataset mentioned in Section III.A.. For nnU-Net and CoTr, two models were constructed using the same self-configuration framework for training and testing. For 3D U-Net and UNETR, we trained the models using the tutorials from MONAI with default parameters. To assess the performance of the automatic segmentation model when more data is involved in the training model, we conducted an experiment with nnU-Net. We trained the nnU-Net model with the EMC$_A$ dataset and then employed the fine-tuning technique to train the model with the EMC$_B$ dataset; we refer to this experiment as nnU-Net (fine-tuning). All automatic segmentation models were trained for 1000 epochs, and the training time for nnU-Net and CoTr was approximately 30 hours, while the training time for U-Net and UNETR was about 25 hours.

To train the RITM model, an interactive sampling procedure is required. We reused the procedure described in the original RITM paper, in which the sampled point is obtained by applying a morphological erosion operation of the mislabeled region[37]. In addition, we used the $DiceCE$ loss function, which was used in the automatic segmentation network nnU-Net. We reuse the RITM model, which was trained on the COCO+LVIS dataset[44,45] as the pre-trained model for the ablation zone segmentation task. Note that before the training stage, the backbone HRnet was already pre-trained with the ImageNet dataset. We then employed the fine-tuning technique to train the RITM model with the EMC$_B$ training dataset. We trained RITM for 500 epochs with the default parameters from the original article.

III.B.2. Ablation zone segmentation demonstration tool

For an easy demonstration, we adopt the demonstration tool, which was created by Sofiiuk et al. (2022)[37], using the $Tkinter$ library. Nevertheless, the original demonstration tool was initially designed for 2D images, which necessitates modifications for working with CT scans. The most significant modification involves embedding the proposed method, incorporating 3D CT scans and enabling users to adjust the display of slices in the 3D CT scan using the mouse. The demonstration tool and video are publicly available at Github [e].

III.C. Experiments setup and results

III.D. Evaluation criteria

In this study, we use the Dice Similarity Coefficient (*DSC*), Average Symmetric Surface Distance (*ASSD*), Hausdorff Distance (*HD*), and Volume difference (*VD*) as metrics for evaluation of the proposed methods.

- **Dice similarity coefficient (DSC)**: Suppose *A* and *B* represent the ground truth and predicted segmentation of the ablation zone of a 3D CT image, respectively. The *DSC* measures how good the overlap between *A* and *B* is. The *DSC* value of 0 means no overlap, and 100 means perfect overlap. The more overlap between *A* and *B*, the closer the *DSC* score to 100%.

$$\mathit{DSC}(A,B) = \frac{2\,|A \cap B|}{|A| + |B|} \times 100\% \quad . \tag{1}$$

- **Average symmetric surface distance (ASSD)**: Suppose *S(A)*, *S(B)* represent all surface voxels on the ground truth (*A*) and the predicted ablation zone segmentation (*B*). Voxel v_A and v_B are arbitrary voxels belonging to *A* and *B*, respectively. We define the shortest path from v_A to *S(B)* or v_B to *S(A)* as follows:

$$d(v_{A'}, S(B)) = \min_{v_{BA} \in S(B)} ||v_A - v_{BA}|| \quad , \tag{2}$$

$$d(v_{B'}, S(A)) = \min_{v_{AB} \in S(A)} ||v_B - v_{AB}|| \quad , \tag{3}$$

where v_{BA} means the point in *S(B)* that draws the shortest distance from point v_A and similar for v_{AB}. The *ASSD* metric measures the average gap between the boundary of *A* and *B*. The formula for *ASSD* is written as follows:

$$\mathit{ASSD}(A,B) = \frac{\sum_{v_A \in S(A)} d\left(v_{A'}, S(B)\right) + \sum_{v_B \in S(B)} d\left(v_{B'}, S(A)\right)}{S(A) + S(B)} \quad . \tag{4}$$

- **Hausdorff distance (HD)**: The *HD* shows the maximum distance between the boundary of *A* and *B*. The formula for *HD* is as follows:

$$\mathit{HD}(A,B) = \max\left(\max_{v_A \in S(A)} d\left(v_{A'}, S(B)\right), \max_{v_B \in S(B)} d\left(v_{B'}, S(A)\right)\right) \quad . \tag{5}$$

- **Volume Difference (VD)**: Suppose V_A and V_B represent the volumes of the ground truth and the predicted segmentation of the ablation zone in a 3D CT image, respectively. The *VD* metric measures the volume difference between these two volumes without considering their overlap. A *VD* value close to 0 indicates that the size of the prediction closely matches the size of the ground truth. The formula for *VD* is as follows:

$$\mathit{VD}(A, B) = \frac{2(V_B - V_A)}{V_B + V_A} \quad . \tag{6}$$

- **Precision**: The formula for Precision is written as:

$$Precision = \frac{TP}{TP + FP} \quad , \tag{7}$$

where *TP* is the number of correctly identified ablation voxels, and *FP* is the number of over segmentation voxels of the predicted ablation zone segmentation.

- **Recall**: The formula for Recall is written as:

$$Recall = \frac{TP}{TP + FN} \quad , \tag{8}$$

where *FN* is the number of incorrectly identified ablation voxels.

In addition, we also use the Area under a curve metric (*AUC*) to evaluate the performance of the segmentation methods.

III.D.1. Automatic ablation zone segmentation

In this experiment, we investigate the performance of four state-of-the-art automatic segmentation networks on ablation zone segmentation: 3D U-Net[38], UNETR[21], nnU-Net, nnU-Net (fine-tuning)[19] and CoTr[23]. The experimental results regarding the comparison of the automatic ablation zone segmentation of four well-known methods are summarized in Table 2. The evaluation is based on the three metrics: *DSC*, *HD* and *ASSD*, and on the two test sets from the EMC$_B$ testing dataset, with the final segmentation achieved using a threshold of 0.5. We also list the number of failed cases, when there is no overlap between the ground truth and the predicted segmentation. The highest mean *DSC* values are 81.2% and 88.4% for CT images acquired in the arterial and portal venous phases, respectively.

Regarding the processing time, UNETR has the lowest processing time (4-5 seconds for a CT volume on average), while nnU-Net (fine-tuning) requires a slightly longer processing time of approximately 7

Table 2: Performance comparison of ablation zone segmentations among the well-known automatic segmentation methods. The bold numbers are the highest mean scores.

Dataset	Method	DSC	HD (mm)	ASSD (mm)	VD	# failure	Processing time (s)
EMC (Arterial) n = 31	**3D U-Net**	61.7 ± 18.7	66.7 ± 78.1	8.2 ± 8.8	0.61 ± 0.39	13	4.4 ± 3
	UNETR	48.8 ± 20.8	203.1 ± 64.8	16.7 ± 11.2	0.56 ± 0.46	9	3.8 ± 2
	nnU-Net	**81.2 ± 12.8**	34.1 ± 29.7	**2.9 ± 4.5**	**0.26 ± 0.29**	1	31.9 ± 11.7
	CoTr	76.6 ± 20.6	39.4 ± 29.7	3.6 ± 5.8	0.34 ± 0.39	1	25.3 ± 9.1
	nnU-Net (fine-tuning)	80 ± 15.6	**29.9 ± 30.3**	3.4 ± 5.4	0.3 ± 0.33	2	7.1 ± 3.5
EMC (Portal venous) n = 25	**3D U-Net**	58.4 ± 26.9	125.3 ± 94.9	12.5 ± 16.1	0.62 ± 0.54	1	5.3 ± 3.8
	UNETR	52.9 ± 20.9	240 ± 84.2	23.4 ± 21.4	0.34 ± 0.58	0	4.9 ± 4
	nnU-Net	86.1 ± 13.7	44.3 ± 37.3	2 ± 2.1	0.15 ± 0.3	0	33 ± 12.8
	CoTr	87 ± 14.2	42.4 ± 36.3	1.9 ± 2.3	0.16 ± 0.31	1	28.2 ± 9.7
	nnU-Net (fine-tuning)	**88.4 ± 11.2**	**25.6 ± 20.2**	**1.4 ± 1.5**	**0.13 ± 0.26**	0	6.6 ± 4.7

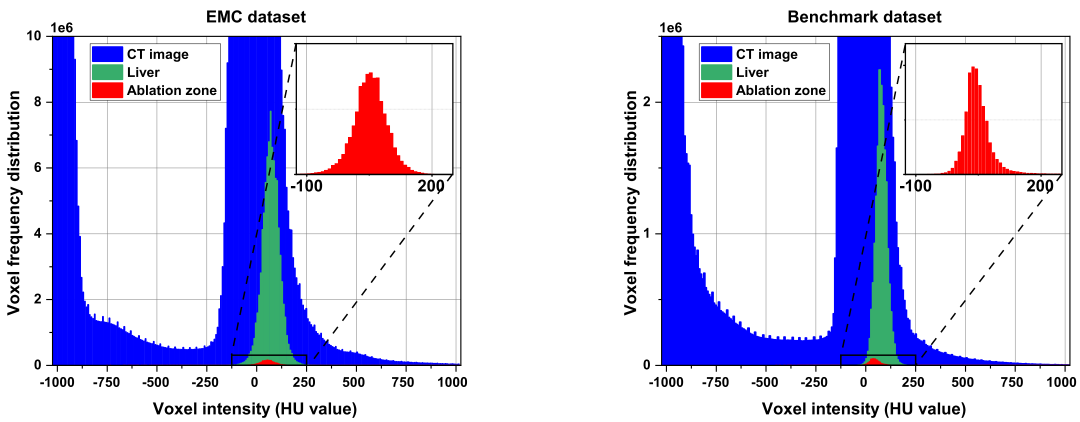

Figure 5: The histogram of voxel intensity in the CT scans in the EMC dataset (left) and the Benchmark dataset (right).

seconds for a CT volume, on average. Furthermore, nnU-Net trained by EMC dataset only has the longest processing time– an average of approximately 30 seconds for a CT volume. The reason is that the nnU-Net framework is designed for patch-based segmentation, which means that the framework needs to define the patch size and the patch separation strategy based on the training dataset. The nnU-Net (fine-tuning) performs a preprocessing stage based on the EMC_A dataset, therefore the preprocessing strategy is different from that of nnU-Net and CoTr, resulting in a lower number of split patches. Thus, the processing time of nnU-Net (fine-tuning) is significantly reduced compared to nnU-Net and CoTr. As a result, we choose the nnU-Net (fine-tuning) model as the automatic method in the proposed approach since it has shown high accuracy and sufficiently fast processing time.

III.D.2. Define optimal model for click-based interactive segmentation

a. HU truncation range assessment:

Table 3: Performance assessment of interactive segmentation model with several the HU truncation range on CT images. NoC@85% and NoC@90% are the average number of required clicks to achieve mean *DSC* scores of 85% and 90%, respectively.

Model	HU truncation ranges	NoC@85	NoC@90
RITM + DiceCE (Baseline)	-100 to 200	**3.62**	**6.01**
	-160 to 240	3.64	6.24
	-100 to 400	3.82	6.53
	-1024 to 1024	4.26	6.46

In this experiment, we demonstrate the value of the HU truncation range on ablation zone segmentation. First, we plot the histogram of the voxel intensity of two datasets to show the distribution of ablation zone intensity in the CT image in Figure 5. It can be seen that the range of -100 to 200 HU contains the most ablation zone voxel intensity (larger than 99%). Furthermore, we evaluate the impact of HU truncation on the performance of the interactive segmentation model. Four HU truncation ranges are employed for this purpose. Firstly, the HU range of -100 to 200 is utilized by He et al. for automatic ablation zone segmentation[14]. Secondly, the HU range of -160 to 240 is widely used in various methods participating in the Liver Tumor Segmentation Benchmark (LiTS)[11]. Thirdly, the HU range of -100 to 400 is often used for liver segmentation[11]. Finally, the HU range of -1024 to 1024 represents the entire HU range of a CT image. We use the RITM (baseline) model with DiceCE loss to perform this experiment. The models are trained using truncated CT image datasets. We evaluate using 100 2D images randomly selected from the validation set. Table 3 indicates that the range of -100 to 200 HU achieved the minimum number of clicks required compared to other HU truncation ranges. Hence, we used the HU range of -100 to 200 for truncating the CT image in the interactive segmentation model.

b. Weight & kernel size optimization:

In this section, we examined the impact of weight λ and kernel size K on the mean number of clicks required to achieve *DSC* scores of 85% and 90% (referred to as NoC@85% and NoC@90%). To achieve this, λ is varied from 0.1 to 0.9 and K from 10 to 190 pixels, and the results were evaluated using 100 2D images randomly selected from the validation set. We experimented with the values of λ and K in two strategies: keeping the value fixed and changing adaptively based on the number of clicks. In the fixed strategy, the values of λ and K are fixed for the segmentation revising process. In the adaptive strategy, each click point provided by the user increases the λ by 10% and decreases K by 10%. A 10% change per click is substantial enough to alter the parameters meaningfully and avoid instability in the segmentation refinement process. The findings in Figure 6 indicate that when λ is small (e.g., 0.1 and 0.3), the interactive network requires more clicks. When λ exceeds 0.5, there is no difference in the

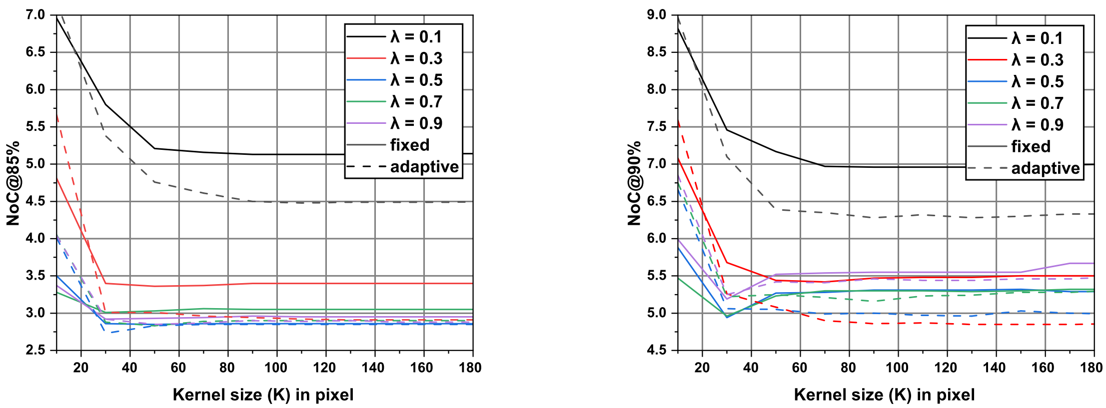

Figure 6: The effect of the mean number of clicks to achieve DSC of 85% (left) and 90% (right) w.r.t the weighted value (λ) and kernel size (K) in fixed strategy (solid line) and adaptive (dash line) strategy.

results. The reason is that when λ is less than 0.5, the weight of interactive segmentation is smaller than that of automatic segmentation. Thus, it requires more clicks in the ablation zone regions in which the automatic model predicts with low confidence scores. Regarding the kernel size K, the fewest clicks are required when K is 30 pixels. For K is less than 30, the ROI may not cover all the large mislabeled regions (e.g., Figure 4.B). When K exceeds 30, it produces an ROI with a large coverage area; when the interactive network has mislabeled regions, it subsequently affects the final segmentation. For the strategies of value adjustment, we observed that there were no significant differences in the minimum value of NoC@85% (with a difference of 0.13) and NoC@90% (with a difference of 0.08) between the fixed and adaptive strategies. Furthermore, utilizing fixed values can offer considerable advantages in simplicity, flexibility, and user proactivity during the process of revising segmentation. Based on these considerations, we selected fixed values of $\lambda = 0.5$ and $K = 30$ for the proposed method.

c. Interactive methods comparison:

This section presents the results of an experiment conducted on the EMC testing set to evaluate the impact of the loss function and combination scheme on the mean number of clicks required to achieve a mean *DSC* score of 85% and 90% (referred to as NoC@85% and NoC@90%). Based on the study of Sofiiuk et al. (2022), the RITM architecture outperformed several interactive segmentation methods[37]. Therefore, we use the RITM network structure to perform this experiment. The experiment involved three models: the baseline model, which is an interactive segmentation model without a guided mask from automatic segmentation; the automatic initial model, which is an interactive segmentation model with a guided mask from automatic segmentation; and the combination scheme model, which is

Table 4: Performance comparison of the loss functions and combination schemes on 2D CT images of the ablation zone. NoC@85% and NoC@90% are the average number of required clicks to achieve mean *DSC* scores of 85% and 90%, respectively. SPC is second per click.

Model		NoC@85%	NoC@90%	SPC (s)
Baseline	**RITM + NFL**	5.13	8.03	0.061
	RITM + DiceCE	4.47	7.23	0.055
Automatic Initial	**RITM + NFL**	4.9	7.86	0.054
	RITM + DiceCE	4.07	6.85	0.051
Combination scheme	**RITM + NFL**	4.79	6.98	0.057
	RITM + DiceCE (selected)	**3.74**	**6.22**	0.054

an interactive segmentation model that utilizes the combination scheme described in section II.C.. To compare the loss function, we trained the RITM model using Normalized Focal Loss (*NFL*), which was used in the original work by Sofiiuk et al. (2022)[37], and using *DiceCE* loss, a loss function that has been used in various medical image segmentation studies. The results, as shown in Table 4, indicate that using *DiceCE* loss performs better than using the *NFL* loss in terms of the mean number of clicks required. Additionally, the selected combination scheme achieved the highest results. As a result, we selected the combination scheme model (RITM + DiceCE) for further evaluation.

In the next experiment, we compare the performance of the proposed method with the baseline interactive segmentation method (RITM) and conventional approach (manual segmentation) on a pilot dataset. The pilot dataset contains 10 CT volumes, which are randomly selected from the testing set, and contains from one to three ablation zones per volume. Two medical image analysis technicians (3 years and 1 year of experience), referred to as User 1 and User 2, respectively, utilized the developed tool with the two interactive segmentation methods to segment the ablation zone in the pilot dataset. Additionally, two users manually annotated the ablation zone slice-by-slice using the Mevislab software. Two users perform the ablation zone annotation until the satisfaction is met. The ablation zone appears as a non-enhancing area of low attenuation in the CT image. The users segment the ablation zone slice-by-slice by delineating the attenuation as a typical procedure[13]. To assess the impact of the interaction on the segmentation accuracy, the *DSC* score of the whole 3D CT volume is recalculated when a new click is provided by a technician. We also plot the mean *DSC* score of manual segmentation from two technicians. The results are shown in Figure 7 (left). It can be seen that, for both of the technicians, using the baseline interactive method requires an average of more than 250 clicks to achieve a mean *DSC* score of 88%. In contrast, for the proposed method, both of the technicians require averages of 53 and 94 clicks to achieve saturated mean *DSC* scores of approximately 91.1% and 92.4%, respectively, indicating that the proposed method outperformed RITM (baseline) in terms of *DSC* score with the same amount of clicks. Furthermore, mean *DSC* scores of 92.4% and 90.9% are achieved by User 1 and User 2 in manual segmentation, respectively. We also achieved the mean *DSC* of 92.4% between the ablation zone manually annotated by the two technicians. The experiment demonstrated a high level of inter-observation agreement between the manual segmentation annotations made by two technicians. The annotation time for each case was recorded. The average annotation time is shown in Figure 7 (right). The results experimental show that using the proposed method, the average annotation time is reduced by approximately 40% and 60% compared with the baseline interactive method and manual segmentation, respectively.

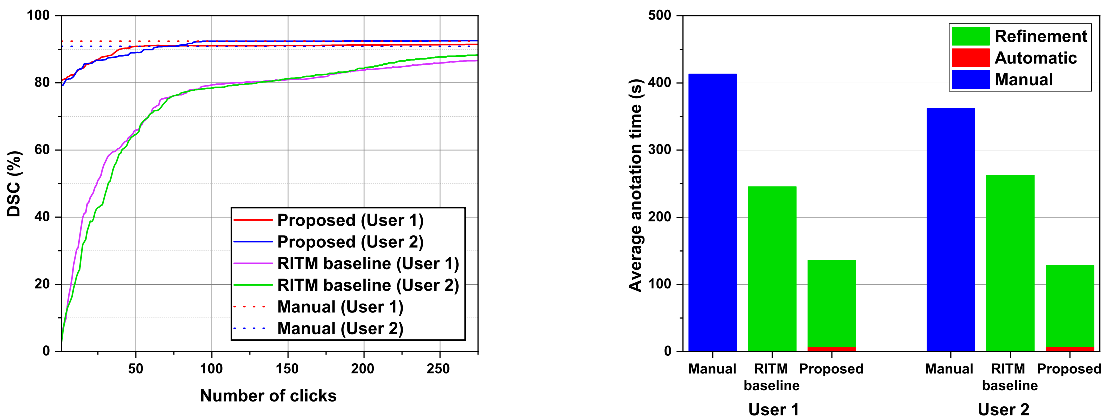

Figure 7: Experiment on the pilot dataset. Mean *DSC* scores w.r.t the number of clicks by two users for segmenting the ablation zones in a CT volume (left). Average annotation time by two users in three manners: Manual, RITM baseline assistance, and proposed method assistance (right). RITM baseline is the interactive segmentation model trained using NFL loss.

III.D.3. Semi-automatic segmentation performance

In this section, we investigate the performance of the proposed method on 3D CT images with human interaction. A technician uses the interactive segmentation tool, which integrates the proposed method, to correct the ablation zone region in the 3D CT images. The evaluation on the EMC$_B$ dataset, arterial phase CT subset shows that the proposed segmentation method obtained a mean *DSC*, *HD*, *ASSD*, and *VD* of 92.3%, 6.5 mm, 0.5 mm, and 0.05, respectively. These metrics for the portal-venous subset are 94%, 8.4 mm, 0.4 mm, and 0.02, respectively (see Table 5). The paired t-tests to those of the nnU-Net (fine-tuning) obtained p- values which are less than 0.01, suggesting that the proposed method statistically significantly improves the segmentation accuracy of the automatic method. In addition, the

Table 5: Performance comparison of ablation zone segmentations between the automatic/semi-automatic segmentation and proposed methods on the EMC_B dataset. The bold numbers are the highest mean values.

Dataset	Method	DSC	HD (mm)	ASSD (mm)	VD	# fails	Processing time (s)
EMC_B (Arterial)	nnU-Net (fine-tuning)	80 ± 15.6	29.9 ± 30.3	3.4 ± 5.4	0.3 ± 0.33	2	7.1 ± 3.5
	MONAI Label (Deepedit)	50.4 ± 23	99.9 ± 82.6	16.6 ± 20.1	0.64 ± 0.41	7	2.1 ± 1.1
	Proposed	**92.3 ± 3.6**	**6.5 ± 3.2**	**0.5 ± 0.3**	**0.05 ± 0.13**	0	121.5 ± 103.1
EMC_B (Portal venous)	nnU-Net (fine-tuning)	88.4 ± 11.2	25.6 ± 20.2	1.4 ± 1.5	0.13 ± 0.26	0	6.6 ± 4.7
	MONAI Label (Deepedit)	60.3 ± 18.2	159.7 ± 137.2	35.1 ± 55.1	0.14 ± 0.42	2	2.4 ± 1.2
	Proposed	**94.0 ± 2.2**	**8.4 ± 5.9**	**0.4 ± 0.2**	**0.02 ± 0.06**	0	126.7 ± 105.8

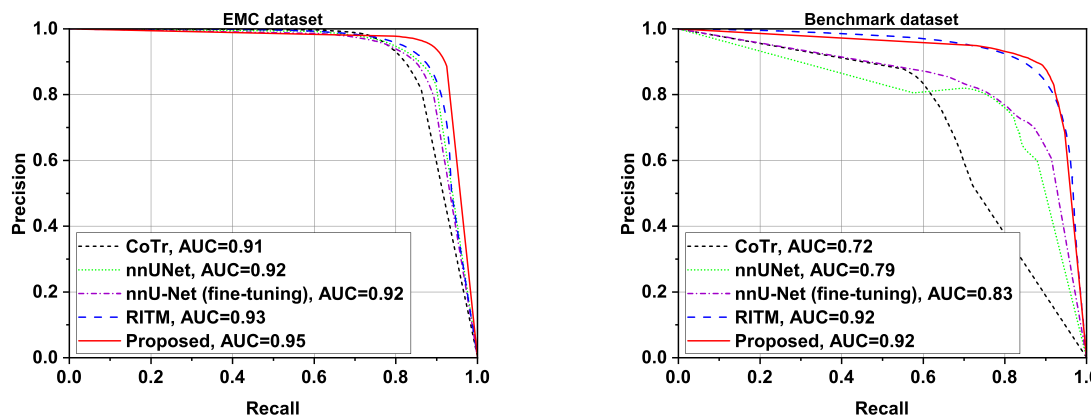

Figure 8: The Precision-Recall curve of ablation zone segmentation on the EMC dataset (left) and the Benchmark dataset (right).

proposed method successfully segmented all of the lesions in the EMC_B dataset while the nnU-Net fine-tuning model failed 2 cases in the EMC_B arterial subset. Moreover, the Precision-Recall curve of the proposed method shows highly precise ablation zone segmentation compared to the automatic segmentation methods: CoTr, nnU-Net, nnU-Net fine-tuning (as depicted in Figure 8). Specifically, the proposed method's AUC scores are 0.92 and 0.95 for the Benchmark dataset and the EMC_B dataset, respectively, which are greater than those of the other automatic methods. Examples of ablation zone segmentation by the proposed method and the other methods on EMC and Benchmark dataset are in Figure 10.

To further assess the segmentation accuracy of the proposed method on the Benchmark dataset, we compared the proposed method with inter-observer manual segmentation and the other well-known segmentation methods, including CoTr, nnU-Net, nnU-Net (fine-tuning) and Graph-based contouring[12]. Two experts labeled the Benchmark dataset. We use the labels created by the first expert as the

Table 6: Performance comparison of ablation zone segmentations on the Benchmark dataset. The bold numbers are the best scores. The statistics of Graph-based contouring method is listed from the original paper by Egger et al. (2015)[12].

Method	DSC	HD (mm)	ASSD (mm)	VD	# failure	Processing time (s)
3D U-Net	62.9 ± 21.5	87.9 ± 78.8	7.8 ± 6.9	0.33 ± 0.64	1	4.7 ± 1.5
UNETR	60.8 ± 13.8	260.3 ± 75.3	25.2 ± 7.5	-0.02 ± 0.43	2	4.5 ± 2.2
nnU-Net	76.1 ± 18.4	40.5 ± 41	6.4 ± 8.8	-0.06 ± 0.46	0	29.8 ± 7.9
CoTr	78 ± 17	47 ± 45	7.3 ± 9.5	-0.19 ± 0.28	3	23.1 ± 7.6
nnU-Net (fine-tuning)	77.2 ± 16.6	33.9 ± 40.3	5.8 ± 8.2	-0.07 ± 0.39	0	5.5 ± 4.5
MONAI Label (Deepedit)	52.3 ± 20.8	105.8 ± 100	14.2 ± 13.8	0.42 ± 0.75	2	-
Proposed	***87.8 ± 6.8***	***9.5 ± 6.9***	***0.9 ± 0.5***	***-0.03 ± 0.07***	***0***	***134.3 ± 82.8***
Manual (inter-observer)	**88.8 ± 3.3**	**8.6 ± 3.4**	**0.8 ± 0.2**	**-0.02 ± 0.07**	**0**	-

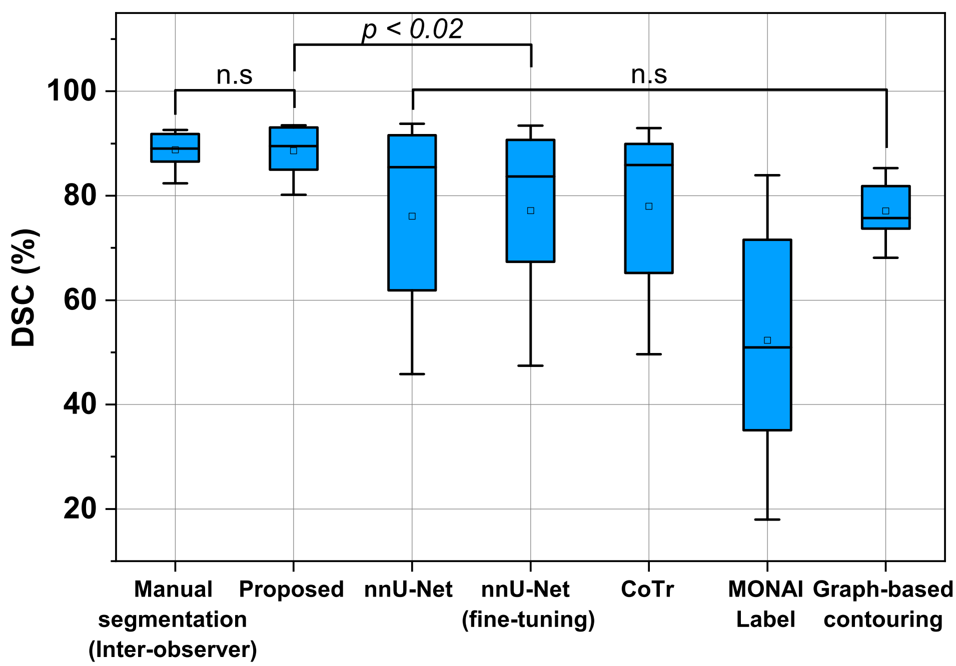

Figure 9: The boxplot of DSC scores among the manual segmentation, the proposed method, the automatic methods, and the classical interactive method[12] for ablation zone segmentation on the Benchmark dataset.

ground truth, while the labels created by the second expert are used to represent the inter-observer manual segmentation. From Table 6, the mean *DSC*, *HD*, *ASSD*, and *VD* scores achieved by the proposed method were 87.8%, 9.5mm, 0.9 mm, and -0.03, respectively. The inter-observer manual segmentation achieved mean *DSC*, *HD*, *ASSD*, and *VD* scores of 88.8%, 8.64 mm, 0.8 mm, and -0.02, respectively. In addition, we applied a *t*-test on the *DSC* scores of the methods. As shown in Figure 9, there is no statistically significant difference between the proposed method and inter-observer manual

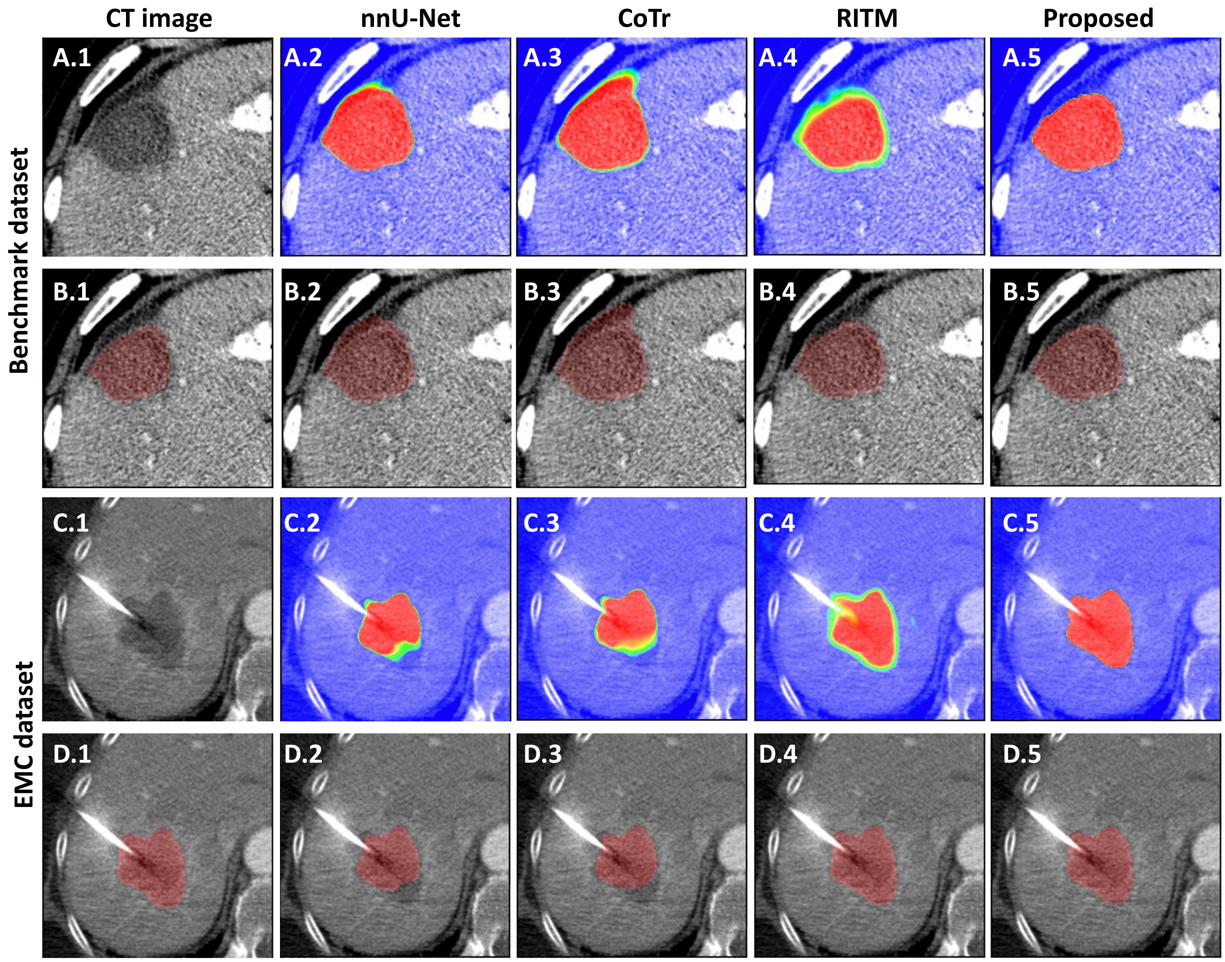

Figure 10: Examples of ablation zone segmentation on EMC and Benchmark dataset of the methods with the segmentation ground truths (B.1 and D.1). The original images (A.1 and C.1) are overlaid by the probability predictions (A.2-5; C2-5) and the thresholded segmentations of each corresponding method (B.2-5; D2-5).

segmentation (p-value $=$ 0.55). While the performances of CoTr, nnU-Net, nnU-Net (fine-tuning), and Graph-based contouring are not statistically significantly different, the proposed method obtained a statistically significantly better performance compared to those of the methods (p-values $\leq$ 0.02). Furthermore, the mean processing time to correct the ablation zone is 134 seconds.

IV. Discussion

In this study, we have proposed and evaluated a semi-automatic method for accurate segmentation of the ablation zone in the post-interventional liver tumor ablation CT images. The ablation zone segmentation accuracy was compared to five state-of-the-art segmentation methods using both internal and publicly available external datasets. Extensive experiments were carried out to assess the performance of the proposed method. In addition, we also developed a tool for demonstrating the effectiveness of the method. The demonstration tool and the source code were made publicly available for research purposes.

Table 2 displays the ablation zone segmentation accuracy of four automatic segmentation methods. The results indicate that nnU-Net performs better than the other baseline methods in automatic segmentation, showing the effectiveness of the self-configuration framework for automatic ablation zone segmentation. Furthermore, using the *t*-test, we found no statistically significant differences between the results of nnU-Net and CoTr (with a *p*-value ranging from 0.2 to 0.5). The accuracy of both methods is comparable to those of the state-of-the-art ablation zone segmentation method reported by He et al. (2021)[14] and Anderson et al. (2022)[15]. We also found that the accuracy of the methods is reduced when using arterial phase CT images compared to the portal venous phase, which is consistent with the study by He et al. (2021)[14]. However, the performance of the automatic segmentation methods suggests that they are still unreliable and insufficient for clinical use. In addition. Table 6 shows that the accuracy of the method decreases when performed on the external dataset, indicating that ablation zone segmentation is still a challenge for fully automatic methods. On the other hand, the proposed semi-automatic segmentation method achieves state-of-the-art performance on ablation zone segmentation and outperforms the other methods, yielding mean *DSC* scores of 92.3%, 94.0%, and 87.8% on the Arterial EMC$_B$, Portal venous EMC$_B$ and Benchmark datasets, respectively (see Table 5, 6, and 8), which are remarkably better than the mean *DSC* scores reported by Anderson et al. (2022)[15] (79%). The means of *ASSD* cores of the proposed method on both the EMC dataset and Benchmark dataset are less than 1 mm, which is smaller than the ideal ablation zone safety margin of 10 mm[46] and equivalent to the median surface distance reported by Anderson et al. (2022)[15] (0.76 mm).

From the experiment with MONAI Label, we see that MONAI Label, a state-of-the-art semi-automatic segmentation method for medical images[47], yielded low accuracy with the embedded click-based method. This is because the default preprocessing setting of the MONAI Label is investigated for multiple organ segmentation (BTCV Challenge), which may not be optimal for a single class (ablation zone segmentation only). In the resampling step, a fixed spacing is applied for the entire data, and a large spacing (spacing of [1.5, 1.5]) for the axial plane resampling makes lost information. Additionally, the patch-based segmentation with a small patch size (the size of [96, 96, 32]) results in a class imbalance during the training phase. Since the ablation zone region is small compared to the whole CT volume, the number of patches that contain the background only is larger than the number of patches that contain the ablation zone. This is the evidence to explain the performance of models implemented using the MONAI library (3D U-Net, UNETR, and MONAI Label) achieved low accuracy. In contrast, the proposed method has the advantage of the self-configuration framework (nnU-Net), which can automatically adapt the preprocessing step based on the training data. Furthermore, using a 2D interactive segmentation model (RITM) for the refinement of the ablation zone in a slice-based manner supports the technician in the refinement of the ablation zone as a typical procedure.

Figure 9 also indicates that the proposed method achieved the best performance compared to other automatic methods. Our observation is that the EMC_B dataset is less noisy than the Benmchark dataset. Thus, the performance of the methods on the EMC_B dataset is higher than that on the Benchmark dataset. In addition, the manual segmentations of the Benchmark dataset were created by two experts. By quantitatively evaluating the inter-observer variability of manual segmentation, we found that the proposed method achieved a segmentation accuracy comparable with the inter-observer variation in terms of *DSC*. The *p*-value of 0.55 (*t*-test) suggests that there is no statistically significant difference between the proposed method and the inter-observer manual technique (see Figure 9).

From Figure 7, we can see that the proposed method needs a lower number of required clicks compared to the original RITM. This is because the proposed method takes advantage of the automatic segmentation to reduce the workload for the user. This indicates that the performance of the automatic method is also an essential factor in reducing the number of clicks. The larger the number of clicks is, the more time and inconvenient it is for the operator to obtain a good segmentation. From Table 5 and 6, we can see that the average processing time of the proposed method for a volume is around 2 minutes, which is small compared to the average operation time of a MWA/RFA session of 112 -149 minutes[48]. Moreover, it can be seen from Figure 7 that the mean saturated *DSC* scores of both users are slightly different. Since the proposed method requires human interaction, we suppose the difference is caused by inter-observer variation.

Our study has some limitations. Firstly, although we conducted the study with both internal and external datasets, and achieved state-of-the-art performance, the number of CT images in the external dataset is only 12 CT volumes, thus conclusions on generality should be drawn with care. By sharing our source code and demonstration tool, we expect other researchers can easily reproduce our obtained results and perform testing on larger external datasets. Secondly, the developed demonstration tool was derived from the work by Sofiiuk et al.(2022)[37], which was not originally designed for medical application purposes. As a result, the number of interactions might not be optimal yet. In this study, we consider the tool for the demonstration purpose only. Further studies may require a better design for the user interface and user experience. Another solution which could be considered is to integrate the proposed method with existing medical image tools such as MONAI[43], ITK-SNAP[49] and 3D Slicer[50]. In addition, although the accuracy of the semi-automatic method is comparable with the accuracy of the manual segmentation by experts, it still contains errors in the final segmentation with a mean *HD* score of 9.5 mm (compared that of the 8.6 mm inter-observer score, p-value = 0.67). These errors seem to be the limitations of human-level performance for annotation and evaluation on the Benchmark dataset. We suggest that the interventionist may take the errors into account in assessing the ablation zone. We acknowledge that the time cost of the proposed method, which is about 120 seconds, is higher than automatic segmentation methods that can produce results in a matter of seconds. This difference in time cost is indeed a consideration and can be seen as a limitation when comparing our method to fully automated approaches. However, the proposed method allows for higher accuracy and customization, as users can iteratively refine the segmentation. This is particularly advantageous in cases where automatic methods might struggle with complex or ambiguous regions. Finally, in this study, we mainly focus on developing methods for precise ablation zone segmentation without further investigating the effect of the ablation zone segmentation on the clinical outcome. Nevertheless, Lin et al. (2023)[46] suggested in their recent study that precise ablation zone segmentation has clinical benefits.

The use of deep learning for medical image analysis is massively expanding at present, especially for image segmentation applications[51,52]. A major drawback of deep learning is that it requires a sufficiently large amount of data for effective training of the models. However, it is frequently difficult to acquire a sufficient amount of medical images with labels that are appropriate for a specific application, potentially resulting in sub-optimal performance. For fully automatic CNN-based segmentation methods, the predicted segmentation may therefore contain segmentation errors. However, with simple interactive corrections using the proposed semi-automatic CNN-based method, the accuracy of segmentation can be improved significantly. Therefore, we expect that using the proposed approach, other segmentation problems may be similarly addressed without requiring large amounts of training data.

V. Conclusions

This study has proposed a semi-automatic approach for ablation zone segmentation in thermal treatments of liver cancer. An accurate segmentation is obtained by combining automatic CNN-based segmentation and click-based CNN segmentation methods. Regarding segmentation accuracy, the proposed method is superior to the well-known CNNs in almost all metrics, achieving comparable performance to manual segmentation of human experts on a benchmark dataset, yielding a mean *DSC* score of 87.8% on average. The obtained segmentation accuracy scores of the proposed approach are also better than those of the other methods when applied to the internal dataset, achieving state-of-the-art performance in accuracy (*DSC* score of 94.0% on average), and the method is sufficiently fast for the use in clinical practice. In conclusion, this study has shown the potential of the semi-automatic approach in supporting the interventionist in assessing the treatment outcome of thermal ablation for liver cancer treatment.

Ethical statement

The local medical research ethics committee decided that the Medical Research Involving Human Subjects Act does not apply to this study. The Benchmark dataset is publicly availble for research purpose.

Conflicts of interest

None.

Acknowledgments

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2018.316. We would like to thank NVIDIA for supporting the RTX 8000 GPU for this study. We would like to thank Mr. Yaro Roodenburg from TU Delft for the early experiment on ablation zone segmentation using nnU-Net. We also would like to thank Mr. Quang Van Nguyen for evaluating the developed tool.

References

[1] M. J. Rutherford et al., Comparison of liver cancer incidence and survival by subtypes across seven high-income countries, International Journal of Cancer **149**, 2020–2031 (2021).

[2] H. Rumgay, M. Arnold, J. Ferlay, O. Lesi, C. J. Cabasag, J. Vignat, M. Laversanne, K. A. McGlynn, and I. Soerjomataram, Global burden of primary liver cancer in 2020 and predictions to 2040, Journal of Hepatology **77**, 1598–1606 (2022).

[3] B. Cadier et al., Early detection and curative treatment of hepatocellular carcinoma: a cost-effectiveness analysis in France and in the United States, Hepatology **65**, 1237–1248 (2017).

[4] J. Han, Y.-c. Fan, and K. Wang, Radiofrequency ablation versus microwave ablation for early stage hepatocellular carcinoma: A PRISMA-compliant systematic review and meta-analysis, Medicine **99** (2020).

[5] D. Li, J. Kang, B. J. Golas, V. W. Yeung, and D. C. Madoff, Minimally invasive local therapies for liver cancer, Cancer biology & medicine **11**, 217 (2014).

[6] M. B. Glassberg, S. Ghosh, J. W. Clymer, R. A. Qadeer, N. C. Ferko, B. Sadeghirad, G. W. Wright, and J. F. Amaral, Microwave ablation compared with radiofrequency ablation for treatment of hepatocellular carcinoma and liver metastases: a systematic review and meta-analysis, OncoTargets and therapy **12**, 6407 (2019).

[7] Z. Yu, G. Li, N. Yuan, and W. Ding, Comparison of ultrasound guided versus CT guided radiofrequency ablation on liver function, serum PIVKA-II, AFP levels and recurrence in patients with primary hepatocellular carcinoma, American Journal of Translational Research **13**, 6881 (2021).

[8] J. H. Kim, H. J. Won, Y. M. Shin, P. N. Kim, S.-G. Lee, and S. Hwang, Radiofrequency ablation for recurrent intrahepatic cholangiocarcinoma after curative resection, European journal of radiology **80**, e221–e225 (2011).

[9] G. Antoch, F. M. Vogt, P. Veit, L. S. Freudenberg, N. Blechschmid, O. Dirsch, A. Bockisch, M. Forsting, J. F. Debatin, and H. Kuehl, Assessment of liver tissue after radiofrequency ablation: findings with different imaging procedures, Journal of Nuclear Medicine **46**, 520–525 (2005).

[10] S. Yedururi, S. Terpenning, S. Gupta, P. Fox, S. Martin, C. Conrad, and E. M. Loyer, Radiofrequency Ablation of Hepatic Tumor: Subjective Assessment of the Perilesional Vascular Network on Contrast Enhanced CT Before and After Ablation can Reliably Predict the Risk of Local Recurrence, Journal of computer assisted tomography **41**, 607 (2017).

[11] P. Bilic et al., The liver tumor segmentation benchmark (lits), Medical Image Analysis **84**, 102680 (2023).

[12] J. Egger et al., Interactive volumetry of liver ablation zones, Scientific Reports **5**, 15373 (2015).

[13] P.-h. Wu, M. Bedoya, J. White, and C. L. Brace, Feature-based automated segmentation of ablation zones by fuzzy c-mean clustering during low-dose computed tomography, Medical physics **48**, 703–714 (2021).

[14] K. He, X. Liu, R. Shahzad, R. Reimer, F. Thiele, J. Niehoff, C. Wybranski, A. C. Bunck, H. Zhang, and M. Perkuhn, Advanced deep learning approach to automatically segment malignant tumors and ablation zone in the liver with contrast-enhanced CT, Frontiers in Oncology **11**, 669437 (2021).

[15] B. M. Anderson, B. Rigaud, Y.-M. Lin, A. K. Jones, H. C. Kang, B. C. Odisio, and K. K. Brock, Automated segmentation of colorectal liver metastasis and liver ablation on contrast-enhanced CT images, Frontiers in Oncology **12** (2022).

[16] 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*, pages 234–241, Springer, 2015.

[17] W. Xu, H. Liu, X. Wang, and Y. Qian, Liver segmentation in CT based on ResUNet with 3D probabilistic and geometric post process, in *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 685–689, IEEE, 2019.

[18] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE transactions on medical imaging **37**, 2663–2674 (2018).

[19] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature methods **18**, 203–211 (2021).

[20] A. Dosovitskiy et al., An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).

[21] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, Unetr: Transformers for 3d medical image segmentation, in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022.

[22] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306 (2021).

[23] Y. Xie, J. Zhang, C. Shen, and Y. Xia, Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 171–180, Springer, 2021.

[24] J. Wallner, M. Schwaiger, K. Hochegger, C. Gsaxner, W. Zemann, and J. Egger, A review on multiplatform evaluations of semi-automatic open-source based image segmentation for cranio-maxillofacial surgery, Computer methods and programs in biomedicine **182**, 105102 (2019).

[25] R. M. Haralick and L. G. Shapiro, Image segmentation techniques, Computer vision, graphics, and image processing **29**, 100–132 (1985).

[26] T. F. Chan and L. A. Vese, Active contours without edges, IEEE Transactions on image processing **10**, 266–277 (2001).

[27] C. Rother, V. Kolmogorov, and A. Blake, " GrabCut" interactive foreground extraction using iterated graph cuts, ACM transactions on graphics (TOG) **23**, 309–314 (2004).

[28] Y. Gao, R. Kikinis, S. Bouix, M. Shenton, and A. Tannenbaum, A 3D interactive multi-object segmentation tool using local robust statistics driven active contours, Medical image analysis **16**, 1216–1227 (2012).

[29] N. Otsu, A threshold selection method from gray-level histograms, IEEE transactions on systems, man, and cybernetics **9**, 62–66 (1979).

[30] V. Caselles, R. Kimmel, and G. Sapiro, Geodesic active contours, International journal of computer vision **22**, 61 (1997).

[31] M. Rajchl et al., Deepcut: Object segmentation from bounding box annotations using convolutional neural networks, IEEE transactions on medical imaging **36**, 674–683 (2016).

[32] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, Scribblesup: Scribble-supervised convolutional networks for semantic segmentation, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3159–3167, 2016.

[33] G. Wang et al., DeepIGeoS: a deep interactive geodesic framework for medical image segmentation, IEEE transactions on pattern analysis and machine intelligence **41**, 1559–1572 (2018).

[34] G. Wang et al., Interactive medical image segmentation using deep learning with image-specific fine tuning, IEEE transactions on medical imaging **37**, 1562–1573 (2018).

[35] X. Luo, G. Wang, T. Song, J. Zhang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, and S. Zhang, MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning, Medical image analysis **72**, 102102 (2021).

[36] L. Sun, Z. Tian, Z. Chen, W. Luo, and S. Du, An efficient interactive segmentation framework for medical images without pre-training, Medical Physics (2022).

[37] K. Sofiiuk, I. A. Petrov, and A. Konushin, Reviving iterative training with mask guidance for interactive segmentation, in *2022 IEEE International Conference on Image Processing (ICIP)*, pages 3141–3145, IEEE, 2022.

[38] 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*, pages 424–432, Springer, 2016.

[39] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction, Nature Machine Intelligence **1**, 269–276 (2019).

[40] R. Benenson, S. Popov, and V. Ferrari, Large-scale interactive object segmentation with human annotators, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11700–11709, 2019.

[41] H. M. Luu, W. Niessen, T. Van Walsum, C. Klink, and A. Moelker, An automatic registration method for pre-and post-interventional CT images for assessing treatment success in liver RFA treatment, Medical Physics **42**, 5559–5567 (2015).

[42] F. Isensee et al., nnu-net: Self-adapting framework for u-net-based medical image segmentation, arXiv preprint arXiv:1809.10486 (2018).

[43] M. J. Cardoso et al., MONAI: An open-source framework for deep learning in healthcare, arXiv preprint arXiv:2211.02701 (2022).

[44] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, Microsoft coco: Common objects in context, in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755, Springer, 2014.

[45] A. Gupta, P. Dollar, and R. Girshick, Lvis: A dataset for large vocabulary instance segmentation, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5356–5364, 2019.

[46] Y.-M. Lin, I. Paolucci, C. S. O'Connor, B. M. Anderson, B. Rigaud, B. M. Fellman, K. A. Jones, K. K. Brock, and B. C. Odisio, Ablative margins of colorectal liver metastases using deformable CT image registration and autosegmentation, Radiology , 221373 (2023).

[47] A. Diaz-Pinto et al., Monai label: A framework for ai-assisted interactive labeling of 3d medical images, arXiv preprint arXiv:2203.12362 (2022).

[48] F. Izzo, V. Granata, R. Grassi, R. Fusco, R. Palaia, P. Delrio, G. Carrafiello, D. Azoulay, A. Petrillo, and S. A. Curley, Radiofrequency ablation and microwave ablation in liver tumors: an update, The oncologist **24**, e990–e1005 (2019).

[49] P. A. Yushkevich, Y. Gao, and G. Gerig, ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images, in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 3342–3345, IEEE, 2016.

[50] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support, in *Intraoperative imaging and image-guided therapy*, pages 277–289, Springer, 2013.

[51] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, Medical image analysis **42**, 60–88 (2017).

[52] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, Recent advances and clinical applications of deep learning in medical image analysis, Medical Image Analysis , 102444 (2022).

Notes

† Deceased 31 July 2023

[a]https://github.com/MIC-DKFZ/nnUNet

[b]https://github.com/Project-MONAI

[c]https://github.com/YtongXie/CoTr

[d] https://github.com/SamsungLabs/ritm_interactive_segmentation

[e]https://github.com/lqanh11/Interactive_AblationZone_Segmentation