Three-Direction Fusion for Accurate Volumetric Liver and Tumor Segmentation

Feng Zhan, Wenwu Wang, *Senior Member, IEEE*, Qian Chen, Yina Guo, *Senior Member, IEEE*, Lidan He, and Lili Wang

***Abstract*—Biomedical image segmentation of organs, tissues and lesions has gained increasing attention in clinical treatment planning and navigation, which involves the exploration of two-dimensional (2D) and three-dimensional (3D) contexts in the biomedical image. Compared to 2D methods, 3D methods pay more attention to inter-slice correlations, which offer additional spatial information for image segmentation. An organ or tumor has a 3D structure that can be observed from three directions. Previous studies focus only on the vertical axis, limiting the understanding of the relationship between a tumor and its surrounding tissues. Important information can also be obtained from sagittal and coronal axes. Therefore, spatial information of organs and tumors can be obtained from three directions, i.e. the sagittal, coronal and vertical axes, to understand better the invasion depth of tumor and its relationship with the surrounding tissues. Moreover, the edges of organs and tumors in biomedical image may be blurred. To address these problems, we propose a three-direction fusion volumetric segmentation (TFVS) model for segmenting 3D biomedical images from three perspectives in sagittal, coronal and transverse planes, respectively. We use the dataset of the liver task provided by the Medical Segmentation Decathlon challenge to train our model. The TFVS method demonstrates a competitive performance on the 3D-IRCADB dataset. In addition, the t-test and Wilcoxon signed-rank test are also performed to show the statistical significance of the improvement by the proposed method as compared with the baseline methods. The proposed method is expected to be beneficial in guiding and facilitating clinical diagnosis and treatment.**

This work was supported by National Natural Science Foundation of China under Grant 62271341 and 61301250, China Scholarship Council under Grant [2020]1417, Key Research and Development Project of Shanxi Province under Grant 201803D421035, Natural Science Foundation for Young Scientists of Shanxi Province under Grant 201901D211313, Shanxi Scholarship Council of China under Grant HGKY2019080 and 2020-127, Shanxi Province Postgraduate Excellent Innovation Project Plan under Grant 2021Y679, Open project of Guangdong Provincial Key Laboratory of Digital Signal and Image Processing in 2021, Natural Science Foundation of Fujian Province under Grant 2021J01243 and 2020J01937. (Corresponding author: Yina Guo).

F. Zhan is with the School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China, the College of Engineering, Fujian Jiangxia University, Fuzhou 350108, China, the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China, and also the Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou 515063, China. (e-mail: fzhan@foxmail.com).

W. Wang is with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7XH, U.K. (e-mail: w.wang@surrey.ac.uk).

Q. Chen and Y. Guo are with the School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China, the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China, and also the Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou 515063, China. (e-mail: chenqian@stu.tyust.edu.cn, zulibest@tyust.edu.cn).

L. He is with the First Affiliated Hospital, Fujian Medical University, Fuzhou 350004, China. (e-mail: hld2399@fjmu.edu.cn).

L. Wang is with the Union Hospital, Fujian Medical University, Fuzhou 350001, China. (e-mail: fjwll1@fjmu.edu.cn).

***Index Terms*—Liver and tumor segmentation, Fusion volumetric segmentation, Three-direction volumetric datasets, Three-direction spatial information, Boolean algebra.**

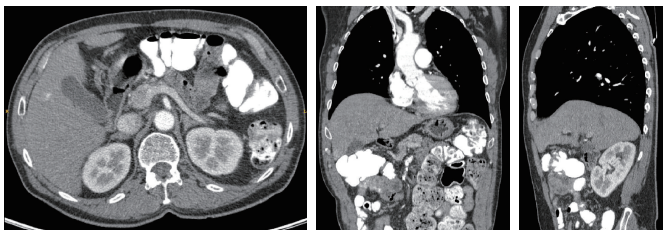
I. Introduction

THE liver is an essential glandular organ for humans, participating in digestion, metabolism, excretion, detoxification, and immunity. Liver cancer is one of the five most frequently diagnosed cancers and the fourth major cause of cancer death globally [1]. Over the past 20 years, biomedical image analysis has been a core constituent of biomedical diagnosis [2]. Biomedical image segmentation, such as liver and tumor segmentation, is one of the most critical areas in biomedical image analysis, which is a prerequisite for quantitative analysis of a region of interest (ROI) [3], enabling successful clinical diagnosis and treatment of the liver [4], [5]. The detection and delineation of liver and tumor are one of the most important steps in current clinical practice supporting diagnosis, tissue quantification, and treatment planning [6]. This is because the accurate semantic segmentation results contain location, shape and volume of the ROI, and important physiological or pathological information, which play a vital role in clinical diagnosis, treatment, and prognosis, especially in computer-aided diagnosis, image-guided interventions, radiotherapy, and downstream applications such as radiomics [7]. In addition, with the segmentation results, the effect of the clinical treatment can be evaluated more accurately to improve the survival rate of patients [6].

The meaningful knowledge can be obtained from biomedical image and clinical data by semantic segmentation, including manual segmentation and machine segmentation [8], [9], [10]. Manual segmentation is a very tedious task that relies on hand-crafted features and is incapable of addressing the automatic segmentation challenge. Therefore, machine segmentation with minor or no human involvement is preferable. Some improved algorithms based on convolutional neural networks (CNNs) offer good segmentation performance [11], such as U-Net [12], Framing U-Net [13], CAC-SPP [14], H-DenseUNet [15], mU-Net [16], RA-UNet [17], KiU-Net [18], Song et al. [19], Di et al. [20], LiTS-Net [21], HFRU-Net [22] and nnU-Net [2].

However, existing 3D biomedical image segmentation methods often focus on spatial information over the vertical axis, limiting the understanding of the relationship between organ, tumor and their surrounding tissues. In addition, the spatial information in biomedical images mined from a single direction

is not complete. In human anatomy, an organ or tumor can be observed from three directions to acquire more information, as shown in Fig. 1. There is an increasing demand on studying how to effectively excavate suspicious tumors in the image, find out the scope of the tumors, and finally show the results to clinicians with visualization. Accordingly, it is crucial to study effective three-direction biomedical image segmentation methods.



(a) Transverse plane (b) Coronal plane (c) Sagittal plane

Fig. 1. Different tomographic meanings observed from different anatomical planes.

The aim of this paper is to develop a method for three-direction fusion based volumetric segmentation of a 3D biomedical image, motivated by recent advances in 2D multi-perspective segmentation [23], which utilized 2D views (slices) of 3D computed tomography (CT) volume in organ-attention networks. Nevertheless, spatial information could be easily lost when analyzing 3D biomedical images from a 2D perspective. Furthermore, efficient three-direction data fusion is a practical challenge in 2D and 3D three-direction segmentation problems. In order to address these difficulties, a three-direction fusion volumetric segmentation (TFVS) method is proposed, and our contributions are summarized as follows.

1) TFVS model. A new model of biomedical image segmentation is proposed in order to exploit three-direction spatial information and enhance segmentation results.
2) TFVS method. Based on the TFVS model, we investigate the various combinations and their performance in predicting segmentation results for 3D biomedical images.
    a) **Datasets.** Inspired by 2D multi-perspective methods, new 3D three-direction datasets are constructed for this study, where the organ and tumor spatial information are obtained from three directions of vertical, sagittal, and coronal axes. More spatial information can be leveraged for image segmentation by using 3D three-direction datasets for training, validating and testing of the network.
    b) **Three-direction method.** We embed sagittal and coronal, and transverse planes to the encoder-decoder structure to fully use three-directional features. The overall performance can be improved by combining spatial information from vertical, sagittal, and coronal axes directions.
    c) **Fusion method.** We propose a fusion method for volumetric segmentation to address boundary ambiguity and the invasion depth of tumors in biomedical image segmentation. The prediction results are processed and analyzed using intersection and union on the test data. By comparing the prediction results, clinicians can gain comprehensive insights into cancer and receive anatomical information for further diagnosis and treatment.

The remainder of the paper is organized as follows. Section II describes the background. Section III formulates the mathematical model for the segmentation problem. Section IV presents our proposed method for the problem of 3D biomedical image segmentation. Section V discusses datasets, baseline methods and performance metrics. Section VI shows experimental results. Section VII concludes the paper and draws potential future research directions.

II. Background

In this section, we discuss the background work proposed for multi-perspective biomedical image segmentation.

From ancient times to the present, medical diagnoses and treatments have been inseparable from anatomy. The widespread availability of biomedical imaging has stimulated the development of biomedical imaging anatomy [24]. Most biomedical data used in clinical practice are stereo image data, such as CT images [25], magnetic resonance imaging (MRI) [26] and positron emission tomography-computed tomography images [27]. The use of 3D biomedical images has gained significant interest as they offer improved assistance to clinicians in understanding the internal structure and functioning of a patient's body, as compared to 2D biomedical images. Recently, inspired by the massive success of deep learning, several researchers considered the application of deep learning for organ and tumor segmentation [28], [29]. The application of deep learning methods to multi-perspective segmentation of organs and tumors can be classified approximately into several categories, such as multi-planar data augmentation, 2D multi-perspective segmentation, 2.5D multi-perspective segmentation, and 3D multi-perspective segmentation methods.

Multi-planar data augmentation methods. Deep learning is a data-driven approach that utilizes input data to perform various tasks. To improve biomedical image segmentation when working with a small sample size, multi-planar data augmentation methods are employed [30]. The effectiveness of this approach is demonstrated in a study by Perslev et al. [31], who used multi-planar data augmentation to enhance segmentation accuracy in a 2D architecture.

2D multi-perspective segmentation methods. Different segmentation results may be obtained when the structures are viewed from different directions [23]. However, 2D segmentation methods lack contextual and spatial information, such as inter-slice correlations within 3D biomedical images. To overcome this limitation, some researchers have proposed 2D multi-perspective biomedical image segmentation methods. For example, Wang et al. [23] utilized 2D views (slices) of CT volume in organ-attention networks. Liu et al. [32] substituted multi-direction fusion modules for convolutional blocks, which is limited in obtaining deep multi-direction integration in the prediction process. Xia et al. [33] employed $n \times n \times 1$ convolutional kernels to derive diversified features

in different sub-networks. Sundaresan et al. [34] proposed a triplanar ensemble network consisting of three 2D networks.

2.5D multi-perspective segmentation methods. To combine the 3D information of multi-perspective 2D models, 2.5D methods achieve an effective balance between model complexity and segmentation accuracy [35], [36].

3D multi-perspective segmentation methods. To address the limitations of 2D segmentation methods in using spatial information, 3D methods have been developed. These methods are more effective in identifying the relationship between organs, tumors, and surrounding tissue [37]. However, current 3D methods are limited to spatial information over the vertical axis, resulting in suboptimal segmentation performance. By improving the flatness of the interpolated 3D medical volumes in all three directions (i.e. axial, sagittal, and coronal), segmentation completeness and accuracy can be significantly enhanced [38]. In addition, the segmentation performance can be improved through larger patch sizes during training, which increases the ability of the network to exploit contextual information [2]. However, the memory constraints of current GPUs mean that some 3D methods would have to be trained using small patch sizes [39].

To address such limitations, we build the TFVS model that improves organ and tumor segmentation accuracy by capturing spatial information from all three directions of a 3D biomedical image, using a fusion volumetric segmentation method, and leveraging 3D three-direction datasets.

III. Mathematical Model

This section mainly focuses on the problems in segmenting organs and tumors from a 3D biomedical image, such as the lack of three-direction spatial information, boundary ambiguity, and the invasion depth of tumors. In this section, a mathematical model is presented for the above problems.

We consider the task of TFVS for 3D biomedical image. Given the dataset $\mathbb{D} = \{\mathbb{D}_1, \mathbb{D}_2\}$, which is composed of training data $\mathbb{D}_1$ and test data $\mathbb{D}_2$, where in the training data $\mathbb{D}_1 = \{\mathbb{D}_{tra}, \mathbb{D}_{val}\}$, $\mathbb{D}_{tra}$ is used for training, and $\mathbb{D}_{val}$ is used for validation. Each labeled data pair is $(\mathbf{V}_k, \mathbf{G}_k)_{k=1}^{N} \in \mathbb{D}$, $\mathbf{V}_k \in \mathbb{R}^{W\times H\times D}$, $\mathbf{G}_k \in \mathbb{R}^{W\times H\times D}$, where $\mathbf{V}_k$ is the k-th volume data and $\mathbf{G}_k$ is the associated ground truth.

A 3D biomedical image is defined as follows,

$$\mathbf{V}_X = \bigcup_{j=1}^{W} \mathbf{A}_{X_j}, \mathbf{V}_Y = \bigcup_{j=1}^{H} \mathbf{A}_{Y_j}, \mathbf{V}_Z = \bigcup_{j=1}^{D} \mathbf{A}_{Z_j}, \quad (1)$$

where W, H and D are the width, height and depth, respectively. $\mathbf{V}_X \in \mathbb{R}^{H\times D\times W}$, $\mathbf{V}_Y \in \mathbb{R}^{W\times D\times H}$ and $\mathbf{V}_Z \in \mathbb{R}^{W\times H\times D}$ are used to represent 3D image from three perspectives: coronal, sagittal and vertical axes. $\mathbf{A}_{X_j} \in \mathbb{R}^{H\times D}$, $\mathbf{A}_{Y_j} \in \mathbb{R}^{W\times D}$ and $\mathbf{A}_{Z_j} \in \mathbb{R}^{W\times H}$ are the j-th 2D image in the sagittal, coronal and transverse planes, respectively.

Biomedical image segmentation aims to find contours separating organs from tumors. The segmentation problem of 3D medical images can be treated as the problem of solving function mapping, as follows,

$$\mathbf{\Omega}_{3D_X} = f(\mathbf{V}_X), \mathbf{\Omega}_{3D_Y} = f(\mathbf{V}_Y), \mathbf{\Omega}_{3D_Z} = f(\mathbf{V}_Z), \quad (2)$$

where $\mathbf{\Omega}_{3D_X}$, $\mathbf{\Omega}_{3D_Y}$, and $\mathbf{\Omega}_{3D_Z}$ are used to represent the results of 3D biomedical image segmentation. The results of 3D biomedical image $\mathbf{\Omega}_{3D_X}$ ($\mathbf{\Omega}_{3D_Y}$, $\mathbf{\Omega}_{3D_Z}$) are divided into n distinct and disjoint non-empty sub-regions $\mathbf{\Omega}_{3D_{X_0}}, \mathbf{\Omega}_{3D_{X_1}}, ..., \mathbf{\Omega}_{3D_{X_{n-1}}}$ (($\mathbf{\Omega}_{3D_{Y_0}}, \mathbf{\Omega}_{3D_{Y_1}}, ..., \mathbf{\Omega}_{3D_{Y_{n-1}}}$), ($\mathbf{\Omega}_{3D_{Z_0}}, \mathbf{\Omega}_{3D_{Z_1}}, ..., \mathbf{\Omega}_{3D_{Z_{n-1}}}$)) by some 3D biomedical image segmentation methods. The relation symbol $H(\mathbf{\Omega}_{3D_{X_i}})$ ($H(\mathbf{\Omega}_{3D_{Y_i}})$, $H(\mathbf{\Omega}_{3D_{Z_i}})$) represents the relevant attributes of set elements, $H(\mathbf{\Omega}_{3D_{X_i}}) = True$ ($H(\mathbf{\Omega}_{3D_{Y_i}}) = True$, $H(\mathbf{\Omega}_{3D_{Z_i}}) = True$) denotes that each element in the set $\mathbf{\Omega}_{3D_{X_i}}$ ($\mathbf{\Omega}_{3D_{Y_i}}$, $\mathbf{\Omega}_{3D_{Z_i}}$) has the same attributes, and $H(\mathbf{\Omega}_{3D_{X_i}}) = False$ ($H(\mathbf{\Omega}_{3D_{Y_i}}) = False$, $H(\mathbf{\Omega}_{3D_{Z_i}}) = False$) denotes that each element in the set $\mathbf{\Omega}_{3D_{X_i}}$ ($\mathbf{\Omega}_{3D_{Y_i}}$, $\mathbf{\Omega}_{3D_{Z_i}}$) does not have the same attributes.

$$\mathbf{\Omega}_{3D_X} = \bigcup_{i}^{n-1} \mathbf{\Omega}_{3D_{X_i}}, \mathbf{\Omega}_{3D_{X_i}} \cap \mathbf{\Omega}_{3D_{X_j}} = \varnothing, \forall i, j, i \neq j, \quad (3)$$

$$\mathbf{\Omega}_{3D_Y} = \bigcup_{i}^{n-1} \mathbf{\Omega}_{3D_{Y_i}}, \mathbf{\Omega}_{3D_{Y_i}} \cap \mathbf{\Omega}_{3D_{Y_j}} = \varnothing, \forall i, j, i \neq j \quad (4)$$

and

$$\mathbf{\Omega}_{3D_Z} = \bigcup_{i}^{n-1} \mathbf{\Omega}_{3D_{Z_i}}, \mathbf{\Omega}_{3D_{Z_i}} \cap \mathbf{\Omega}_{3D_{Z_j}} = \varnothing, \forall i, j, i \neq j, \quad (5)$$

that is, the subsets obtained from the segmentation should cover the whole segmentation result and not overlap with each other. In our work, the value of n in Eq. (3), (4) and (5) is taken as 3, where $\mathbf{\Omega}_{3D_{X_0}}$ ($\mathbf{\Omega}_{3D_{Y_0}}$, $\mathbf{\Omega}_{3D_{Z_0}}$) is background, $\mathbf{\Omega}_{3D_{X_1}}$ ($\mathbf{\Omega}_{3D_{Y_1}}$, $\mathbf{\Omega}_{3D_{Z_1}}$) is organ, and $\mathbf{\Omega}_{3D_{X_2}}$ ($\mathbf{\Omega}_{3D_{Y_2}}$, $\mathbf{\Omega}_{3D_{Z_2}}$) is tumor.

A deep neural network has been used recently to achieve function approximation in biomedical image segmentation. There is a correlation between slices in 3D biomedical images. By transforming 2D convolution in a CNNs model into 3D convolution, 3D biomedical image segmentation can be achieved. Implementing the 3D network is memory-demanding, making it infeasible to take the whole 3D voxel as input. To address this issue, we can divide it into a number of 3D patches as input. However, cropping may reduce the maximum receptive field of the network, resulting in missing global and spatial information. These may lead to inaccurate segmentation results. Therefore, further improvement is required.

Different structures and texture patterns from various viewing directions often generate nonidentical segments [23]. We can utilize logical operations in Boolean algebra, including logic AND, logic OR and logic XOR, to determine the correlation of three segmentation results at the exact location for addressing problems with insufficient emphasis and border ambiguity. Formally, logic AND is the logical intersection of three prediction results from the pixel-level perspective, combined into a prediction result,

$$\mathbf{\Omega}_{3D_1}^{and} = \mathbf{\Omega}_{3D_{X_1}} \cap \mathbf{\Omega}_{3D_{Y_1}} \cap \mathbf{\Omega}_{3D_{Z_1}} \quad (6)$$

and

$$\mathbf{\Omega}_{3D_2}^{and} = \mathbf{\Omega}_{3D_{X_2}} \cap \mathbf{\Omega}_{3D_{Y_2}} \cap \mathbf{\Omega}_{3D_{Z_2}}, \quad (7)$$

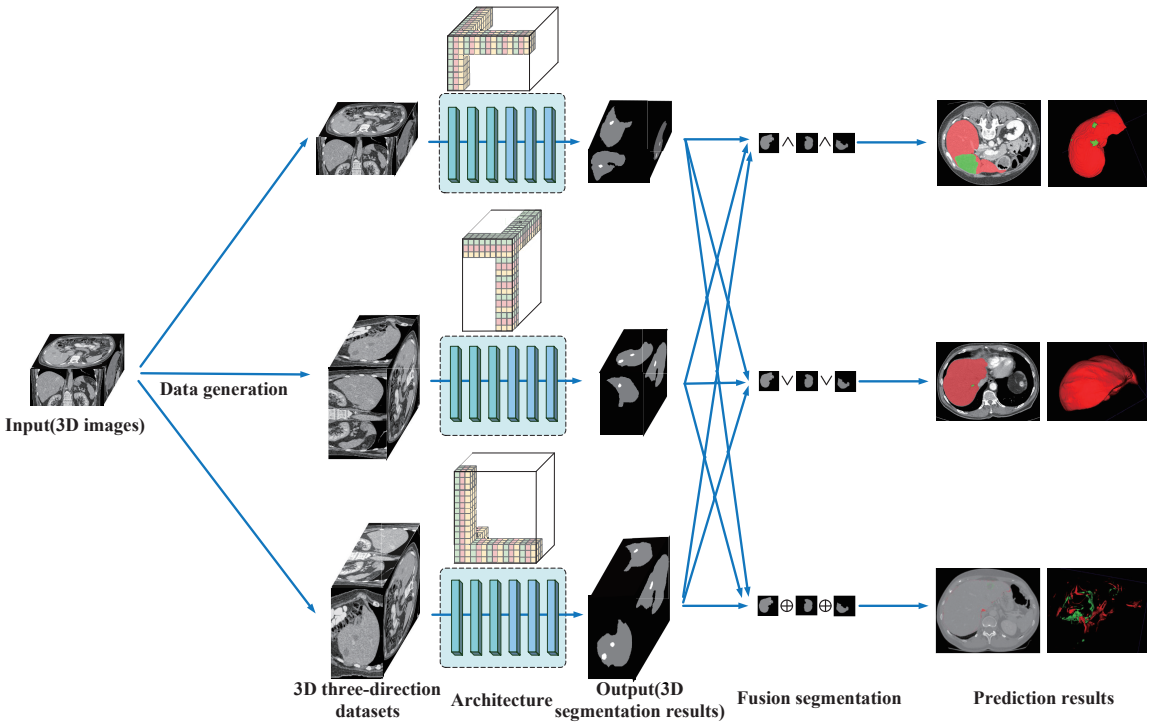


Fig. 2. The overall framework of the proposed TFVS algorithm for liver and liver tumor segmentation model.

where $\mathbf{\Omega}_{3D_1}^{and}$ and $\mathbf{\Omega}_{3D_2}^{and}$ are the results of logic AND on organ and tumor, and $\cap$ is the element-wise logic AND operation.

Logic OR is the logical union of the three predicted segmentation results from the pixel-level perspective, combined into a prediction result,

$$\mathbf{\Omega}_{3D_1}^{or} = \mathbf{\Omega}_{3D_{X_1}} \cup \mathbf{\Omega}_{3D_{Y_1}} \cup \mathbf{\Omega}_{3D_{Z_1}} \tag{8}$$

and

$$\mathbf{\Omega}_{3D_2}^{or} = \mathbf{\Omega}_{3D_{X_2}} \cup \mathbf{\Omega}_{3D_{Y_2}} \cup \mathbf{\Omega}_{3D_{Z_2}}, \tag{9}$$

where $\mathbf{\Omega}_{3D_1}^{or}$ and $\mathbf{\Omega}_{3D_2}^{or}$ are the consequences of logic OR on organ and tumor, and $\cup$ is the element-wise logic OR operation.

Logic XOR is the logical exclusive-or of three predicted segmentation results from the pixel-level perspective, combined into a predicted result,

$$\mathbf{\Omega}_{3D_1}^{xor} = \mathbf{\Omega}_{3D_{X_1}} \oplus \mathbf{\Omega}_{3D_{Y_1}} \oplus \mathbf{\Omega}_{3D_{Z_1}} \tag{10}$$

and

$$\mathbf{\Omega}_{3D_2}^{xor} = \mathbf{\Omega}_{3D_{X_2}} \oplus \mathbf{\Omega}_{3D_{Y_2}} \oplus \mathbf{\Omega}_{3D_{Z_2}}, \tag{11}$$

where $\mathbf{\Omega}_{3D_1}^{xor}$ and $\mathbf{\Omega}_{3D_2}^{xor}$ are the consequences of logic XOR on the organ and tumor, and $\oplus$ is the element-wise logic XOR operation.

IV. Three-direction fusion volumetric segmentation method

In this section, we propose a three-direction method by changing the orientation of the network input data, and obtaining intra-slice and inter-slice from sagittal, coronal and transverse planes, followed by jointly optimizing the hybrid features for precise liver and tumor segmentation. The TFVS model is presented to improve medical image segmentation by exploiting directional information in 3D images. The overall framework of the proposed TFVS algorithm for liver and tumor segmentation model is shown in Fig. 2, which consists of three steps. The first step is to generate 3D three-direction datasets $\mathbb{D}_{\boldsymbol{X}}$, $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{Z}}$. The sagittal plane $\bigcup_{j=1}^{W} \mathbf{A}_{X_j}$, coronal plane $\bigcup_{j=1}^{H} \mathbf{A}_{Y_j}$, and transverse plane $\bigcup_{j=1}^{D} \mathbf{A}_{Z_j}$ are taken as the first plane. The second step is to train the models on the 3D three-direction datasets. The pre-trained models are used to predict segmentation results $\mathbf{\Omega}_{3D_{X_i}}$, $\mathbf{\Omega}_{3D_{Y_i}}$ and $\mathbf{\Omega}_{3D_{Z_i}}$ $(i = 0, 1, 2)$, respectively. The third step is to highlight liver and tumor $\mathbf{\Omega}_{3D_i}^{and}$, $\mathbf{\Omega}_{3D_i}^{or}$ and $\mathbf{\Omega}_{3D_i}^{xor}$ $(i = 1, 2)$ through the intersection and union of the segmentation result. A detailed description of each step is given below.

A. Generation of 3D three-direction datasets

In engineering and biomedical imaging, 3D objects are often observed from multiple perspectives, which provide visual information from various segmentation surfaces. Spatial information between slices in 3D biomedical images can be obtained through 3D modeling, which enhances the performance of image segmentation [17], [18]. To further improve this process, we propose training three different models from three perspectives: vertical, sagittal, and coronal axes. We can train the model with images from each perspective by utilizing the same network structure and adjusting the network input. These models can then be used for 3D biomedical image prediction from the corresponding viewpoint.

Human anatomy contains three mutually perpendicular axes: vertical, sagittal, and coronal. The human body, organs, tissues, lesions, and tumors can be formed into different slices

according to these axes, each providing unique cross-sectional information. For example, the liver can be observed in the vertical, sagittal, and coronal planes, allowing for examination of its relationship with other organs such as the kidney, heart, and lungs in each respective direction. Biomedical images such as CT and MRI allow us to examine the human body as a 3D rigid object, which can be rotationally adjusted using 3D rigid rotation methods. To generate 3D three-direction datasets, we follow the steps outlined in this paper.

1) Transverse plane angle dataset (dataset $\mathbb{D}_{\boldsymbol{Z}}$). The connection between the upper and lower slices is linked according to the transverse plane. The given dataset is denoted as $\mathbb{D}_{\boldsymbol{Z}}$.
2) Sagittal plane angle dataset (dataset $\mathbb{D}_{\boldsymbol{X}}$). The NIFTI format data of dataset $\mathbb{D}_{\boldsymbol{Z}}$ are converted to NumPy arrays using the Nibabel Python library. The NumPy arrays of the 3D image data are rotated by 90 degrees along Y-axis, as detailed in Eq. (12), with β set to 90 degrees. Finally, the modified data is saved in NIFTI format to create dataset $\mathbb{D}_{\boldsymbol{X}}$, which provides spatial information from the data.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} cos\beta & 0 & -sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ sin\beta & 0 & cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{12}$$

3) Coronal plane angle dataset (dataset $\mathbb{D}_{\boldsymbol{Y}}$). The dataset $\mathbb{D}_{\boldsymbol{Y}}$ is obtained by rotating each 3D image data in the dataset $\mathbb{D}_{\boldsymbol{Z}}$ by 90 degrees along the X-axis. The above implementation can be defined as Eq. (13). The value of α in Eq. (13) is taken as 90 degrees. The generated dataset $\mathbb{D}_{\boldsymbol{Y}}$ is intended to facilitate access to the spatial information of the data.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha & sin\alpha & 0 \\ 0 & -sin\alpha & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{13}$$

With the above calculation, we can get the planar schematic diagram of datasets $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$, $\mathbb{D}_{\boldsymbol{X}}$ shown in Fig. 3.

Fig. 3. Planar schematic diagrams (a, b, c) of datasets $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$ observed from the transverse plane.

B. Segmentation network architecture and training details

After the previous section, we obtain the required data and divide the training data into three sets. We develop a segmentation network architecture for this assignment. The segmentation network architecture is depicted in Fig. 4. The encoder sub-network on the left comprises five convolutional layers, while the decoder sub-network on the right contains five deconvolution layers. Each layer consists of several modules. The encoder path of the segmentation network uses residual learning to improve the network performance [40]. Residual learning can overcome the potential degradation problem associated with a network with increased depth. We employ stacked residual units to build deeper convolutional encoders. Also, we remove the second LeakyReLU block from the residual unit and add a convolutional module after the residual unit, which is used for feature aggregation. This convolution module shares the same parameters as the convolution module inside the residual unit. This design can reduce computational complexity. The max-pooling layer is replaced by a strided convolution layer with step size 2 to enhance the network accuracy and reduce the model size. By deepening the layers, information can be transmitted hierarchically, and the problems to be learned at each layer are decomposed into simpler problems that are easier to solve. After completing the decoder sub-network, a $1\times1\times1$ convolutional block is utilized to convert the feature map generated by the segmentation network to the corresponding result.

Our segmentation network is built on PyTorch 1.10.0 framework with CUDA 11.4 and cuDNN 8.2.1.32. The loss function used in the proposed method is the product of cross entropy (CE) loss and Dice loss [41]. The CE loss is skewed toward the detection of larger target areas and seeks to improve segmentation at the image level. However, the inclusion of a Dice component in the loss function facilitates the detection of small tumours and decreases false negatives, especially in regions with a significant class imbalance. The loss function is

$$\mathcal{L} = \lambda_1 \mathcal{L}_{CE} + \lambda_2 \mathcal{L}_{Dice} \tag{14}$$

with

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{c=1}^{C} \sum_{i=1}^{N} g_i^c \log s_i^c, \tag{15}$$

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_{c=1}^{C} \sum_{i=1}^{N} g_i^c s_i^c}{\sum_{c=1}^{C} \sum_{i=1}^{N} g_i^c + \sum_{c=1}^{C} \sum_{i=1}^{N} s_i^c}, \tag{16}$$

where λ_1 and λ_2 are the hyper-parameters weighting the contributions of the CE loss and Dice loss, respectively, set empirically as $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$ in our experiments. The term g_i^c is the ground truth binary indicator representing the class label c of voxel i, while s_i^c is the corresponding predicted segmentation probability.

With a larger patch, more contextual information can be exploited by the network. However, due to the memory constraints of the available GPUs, we choose the batch size to be 2, which considers a tradeoff between the memory requirment and the segmentation performance. In both the convolution and deconvolution layers, we use 3D convolution kernels in the network. The network parameters are optimized by the stochastic gradient descent algorithm. We chose to run a fixed number of epochs, i.e. 1000, during training, where each epoch includes 250 training iterations. The initial learning rate is set at 0.005. As illustrated in Fig. 4, kernels of size $1 \times 1 \times 1$,

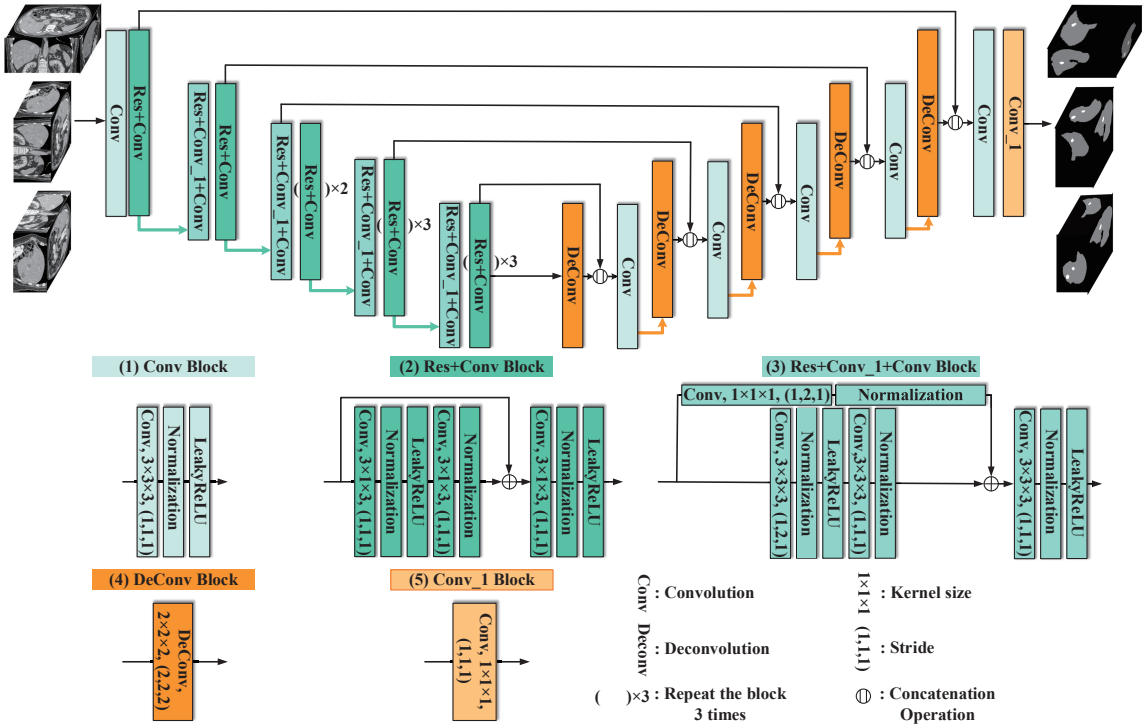


Fig. 4. The network architecture for liver and liver tumor segmentation. Several key network blocks from different parts are illustrated at the bottom.

$3\times3\times3$ and $3\times1\times3$ are used in the convolution layers, while the kernel of size $2\times2\times2$ is used in the deconvolution layers. When all the models have been trained, five folds are combined for obtaining the final predictions. In other words, the best model is chosen according to cross-validation performance.

The 3D three-direction datasets $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$ are used to evaluate the proposed TFVS model. Based on the segmentation network, the training is performed separately, and three different pre-trained models from three perspectives: vertical, sagittal and coronal axes are obtained, respectively. On the training set, the experiment is conducted in the way of five folds cross validation, with four folds as the training set and the remaining fold as the validation set. Moreover, the three pre-trained models are used to segment the test set from three perspectives. $\boldsymbol{\Omega}^1_{3D}$ and $\boldsymbol{\Omega}^2_{3D}$ are obtained by fusion segmentation, which addresses problems with insufficient emphasis and boundary ambiguity.

The facilities used to carry out the experiments include AMD Ryzen 7 3700X 8-Core Processor, NVIDIA Geforce RTX 3080 Graphics Card, 112 GB memory, and Ubuntu 20.04 LTS 64-bit Operating System. The network architecture is constructed using the well-known deep learning framework, PyTorch [42]. Following the work of Wang et al. [23], we train the network for each direction. We follow the recommendation to train the backbone network on the 3D unidirectional dataset for experimental comparison.

C. Fusion method of volumetric segmentation

In multi-perspective biomedical image segmentation, it is necessary to consider how to merge multiple objects. Several fusion methods have been proposed for improving segmentation accuracy [32], [43], [44]. They mainly discuss fusion methods in terms of network construction. Although existing algorithms have made great improvements on biomedical image segmentation, they rarely consider it from the perspective of data processing. Moreover, image edge is often blurred in biomedical images [45].

In this paper, liver and its tumor are predicted from three perspectives: vertical, sagittal and coronal axes. Different perspectives may lead to different results. The three-direction perspectives provide complementary information, which potentially can improve the segmentation performance. Therefore, attention is paid to accuracy and comprehensiveness in the fusion process. Inspired by the fusion network [45], we can address the problems of insufficient emphasis and boundary ambiguity through the intersection and union of the segmentation results, as given in Eq. (6), (7), (8), (9), (10) and (11). The fusion method for volumetric segmentation is summarized in Algorithm 1.

Logic AND is an operation of taking the intersection among the segmentation results, which can further improve the prediction confidence and emphasize the core position of the object by eliminating the intra-class inconsistency. Logic OR is an operation of merging the prediction results of medical image segmentation, in order to improve the recall and marginal details of organs and tumors. In the subsequent experiments, we adopt logic OR to obtain better segmentation results. Logic XOR is an XOR operation applied on the segmentation results from the three perspectives, to improve the exposure of marginal details in the concerned object and highlight the regional boundary of the three prediction results.

Algorithm 1 The algorithm of fusion for volumetric segmentation

Input: The input image $\mathbf{V}_X$, $\mathbf{V}_Y$ and $\mathbf{V}_Z$ are given in Eq. (2).

Output: Enhanced results of $\mathbf{\Omega}^{and}_{3D_i}$, $\mathbf{\Omega}^{or}_{3D_i}$ and $\mathbf{\Omega}^{xor}_{3D_i}$, i=1,2.

1. *The input image to array conversion.*
$\mathbf{Array}_{3D_X}$, $\mathbf{Array}_{3D_Y}$ and $\mathbf{Array}_{3D_Z}$.
2. *Transformation of array.*
The prediction results are mapped into the original image direction. $\mathbf{Array}'_{3D_X}$ and $\mathbf{Array}'_{3D_Y}$ are given in Eq. (13) and (15), where the value of α and β are taken as 270 degrees.
3. *Data fusion.*
$\mathbf{Array}'_{3D_X} \wedge \mathbf{Array}'_{3D_Y} \wedge \mathbf{Array}_{3D_Z}$,
$\mathbf{Array}'_{3D_X} \vee \mathbf{Array}'_{3D_Y} \vee \mathbf{Array}_{3D_Z}$,
$\mathbf{Array}'_{3D_X} \oplus \mathbf{Array}'_{3D_Y} \oplus \mathbf{Array}_{3D_Z}$.
4. *Array to enhanced results conversion.*
$\mathbf{\Omega}^{and}_{3D_i}$, $\mathbf{\Omega}^{or}_{3D_i}$ and $\mathbf{\Omega}^{xor}_{3D_i}$, i=1, 2.

It provides more details of the regional border, thus enabling better observation of an organ and tumor. As a result, clinicians can more quickly understand the relationship between the organ and tumor. Therefore, it is possible to merge different spatial information using logic AND, logic OR and logic XOR, through the intersection and union of the three segmentation results. This can facilitate the clinicians in observing the patterns from the data, highlighting organs and tumors, as well as the details of the regional border.

V. Experimental setup

In this section, we discuss the design for performing 3D image segmentation experiments, including datasets, baseline methods, and performance metrics.

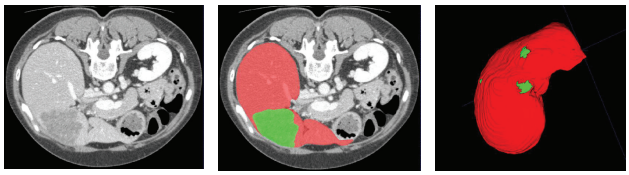
A. Experimental datasets

We evaluate the proposed approach for liver and tumor segmentation using public datasets, including the Liver dataset provided by Medical Segmentation Decathlon (MSD) [46] and the 3D Image Reconstruction for Comparison of Algorithm Database (3Dircadb) provided by IRCAD [47]. We take these two datasets and create the training, validation and testing sets for our experiments. Our segmentation target contains liver and tumor.

1) Liver dataset: Liver dataset is one of the MSD challenge datasets. We use the training data of Liver dataset to train the proposed model. This dataset has 201 contrast-enhanced 3D abdominal CT scans, with 131 scans for training with ground-truth labels, and 70 scans for testing. The challenge is to address the label imbalance issue with a large liver and tiny tumor target. Several works that are based on U-Net performed well for this challenge [5].

2) 3Dircadb dataset: To evaluate the generalization capability of the TFVS model, we utilize a training set comprising 111 cases from the Liver dataset. This training set excludes data from the 3Dircadb dataset, and the model is subsequently tested on the 3Dircadb dataset. The format of the 3Dircadb dataset does not follow the data structure of the Liver dataset, which should be converted for the following examination process. Processing for data formation is implemented using Python (3.6 and 3.8, Python Software Foundation).

The images in the Liver dataset and the 3Dircadb dataset are in 512×512 pixels. The number of CT scan slices can vary from a few dozen to around one thousand slices. Visualizations are performed using ITK-SNAP version 3.6 (the interactive image segmentation software, www.itksnap.org) [48]. An example for the visualization of an image from the Liver dataset is shown in Fig. 5.



(a) Original Image (b) Ground Truth (c)3D Rendering

Fig. 5. An image example from the Liver dataset. The ground truth of the original image (a) is shown in 2D projected onto the raw data (b) and in 3D together with a volume rendering of the raw data (c). The red regions indicate liver, while the green ones indicate tumor.

For the liver and tumor segmentation task, we utilize data pre-processing to simplify the complexity of calculation. The intensity values of CT window level and window width are truncated to the range of [-200, +250] Hounsfield units to eliminate the irrelevant information [49]. Only those slices containing liver need to be extracted, while the remaining slices are ignored. Generating the input data based on data cropping and sampling instead of the entire CT volume improves the computational efficiency in 3D biomedical image segmentation, as it reduces the memory consumption and computational resource during the training process. To crop the original CT scans to the liver region, we adjust the 2D images of the transverse plane size to 256×256. The median size in voxels is [256, 256, 362], and median spacing is [1.54, 1.54, 0.5] after data pre-processing.

B. Baseline methods

We compare the TFVS method (trained on the 3D three-direction datasets $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$) with the state-of-the-art algorithms: U-Net (trained on the dataset $\mathbb{D}$ for the best model) [12], Framing U-Net [13], CAC-SPP [14], H-DenseUNet [15], mU-Net [16], Song et al. [19], Di et al. [20], LiTS-Net [21], HFRU-Net [22] and nnU-Net (trained on the dataset $\mathbb{D}$ for the best model) [2].

In order to assess the generalization ability of the TFVS method, we directly apply the model trained on the Liver dataset to the 3Dircadb dataset. To further evaluate the performance of our proposed model and algorithm, we compare it with other published methods. The comparison is restricted to the methods with available references. The prediction results of Framing U-Net, CAC-SPP, H-DenseUNet, mU-Net, Song et al., Di et al., LiTS-Net and HFRU-Net are directly taken from the original papers, while the remaining baseline methods are trained from scratch with the training data (i.e., dataset $\mathbb{D}$, 3D three-direction datasets $\mathbb{D}_{\boldsymbol{X}}$, $\mathbb{D}_{\boldsymbol{Y}}$, $\mathbb{D}_{\boldsymbol{Z}}$).

C. Performance metrics

To evaluate the TFVS method in segmenting liver and tumor, we employ the Dice, Precision, Recall and Jaccard scores, as in [50]. These four metrics are calculated based on the true positive (TP), false positive (FP) and false negative (FN), respectively, as follows,

$$Dice = \frac{2TP}{2TP + FP + FN} \times 100\%, \tag{17}$$

$$Precision = \frac{TP}{TP + FP} \times 100\%, \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \times 100\%, \tag{19}$$

$$Jaccard = \frac{TP}{TP + FP + FN} \times 100\%, \tag{20}$$

where TP is the number of pixels that are identified as target objects and correspond with the ground truth, FP is the number of pixels that are identified as target objects but belong to background, and FN is the number of pixels that are identified as background but not consistent with the ground truth. Dice evaluates the overlap between the predicted and ground truth segments. Precision score reflects the proportion of true positive samples among the algorithmically identified positive cases. Recall score is the correct prediction rate after biomedical image segmentation. Jaccard measures the similarity between the prediction and the ground truth. Higher Dice, precision, recall and Jaccard scores indicate better performance. The t-test and Wilcoxon signed-rank test are also performed to evaluate the statistical significance of the performance difference between the proposed method and the baseline methods.

VI. Results

We present quantitative and qualitative results of our method, comparing it to the state-of-the-art methods. Then, we perform some ablation studies by conducting experiments in various directions. Moreover, we discuss the generalization capability of our method and present a comparison on the 3Dircadb dataset.

A. Overall performance

The correctness of the results obtained by a segmentation method is a critical evaluation index. We present a quantitative comparison using various evaluation metrics for liver and tumor segmentation. Our method's performance is assessed using CT volumes for liver and tumor segmentation from the 3Dircadb dataset, as shown in Table I. The results show that liver segmentation results of the proposed TFVS method on the 3Dircadb dataset in mean values of the Dice is 96.70% that is 5.25%, 2.65%, 2.38%, 1.59%, 0.69%, 0.20%, 0.05% and 0.23% improved than the U-Net, Framing U-Net, CAC-SPP, H-DenseUNet, mU-Net, Song et al. [19], nnU-Net and LiTS-Net, respectively. Our proposed TFVS method has better performance for liver tumor segmentation in mean values of Dice than the U-Net, Framing U-Net, CAC-SPP, H-DenseUNet, and mU-Net algorithms on the 3Dircadb dataset. It is 36.70%, 34.47%, 18.58%, 16.13%, 14.46%, 13.8%, 11.6%, 10.63%, 8.06% and 4.70% improved than the U-Net, Framing U-Net, CAC-SPP, H-DenseUNet, mU-Net, Song et al., Di et al., nnU-Net, LiTS-Net and HFRU-Net, respectively. Table I presents the prediction results of Framing U-Net, CAC-SPP, H-DenseUNet, mU-Net, Song et al., Di et al., LiTS-Net, and HFRU-Net, as reported in their respective papers. Statistical significance tests using t-test are also performed to show the significance of the performance improvement with our proposed TFVS algorithm as compared with the baseline methods. From the results, we can observe the statistically significant differences between our method and the baseline methods for segmenting liver tumors.

TABLE I
THE PERFORMANCE (DICE, %) OF OUR TFVS METHOD AS COMPARED WITH STATE-OF-THE-ART METHODS ON THE 3DIRCADB DATASET. ALL VALUES ARE REPORTED AS MEAN ± STANDARD DEVIATION. BEST SCORES ARE MARKED IN BOLD.

Algorithm	*Dice* (%)	
	Liver	Liver tumor
U-Net [12]	91.45±4.05*	45.90±31.09*
Framing U-Net [13]	94.05±1.20*	48.13±18.44*
CAC-SPP [14]	94.32±1.13*	64.02±7.18*
H-DenseUNet [15]	95.11±1.04*	66.47±6.54*
mU-Net [16]	96.01±1.08	68.14±6.40*
Song et al. [19]	96.5	68.8*
Di et al. [20]	-	71±7*
nnU-Net [2]	96.65±1.43	71.97±24.94*
LiTS-Net [21]	96.47±1.13	74.54±6.21*
HFRU-Net [22]	**97.3***	77.9*
TFVS	96.70±1.15	**82.60±7.96**

* indicates that the results are statistically different from TFVS (t-tests, $p < 0.05$).

Table II, Fig. 6 and Fig. 7 show more detailed comparison results of U-Net, nnU-Net and TFVS. Our proposed TFVS algorithm achieves higher scores than U-Net [12] and nnU-Net [2] on the 3Dircadb dataset. Both nnU-Net and TFVS outperform U-Net in liver segmentation, and TFVS performs better than the others in segmenting liver tumors. The statistical analysis using the Wilcoxon signed-rank test shows that TFVS is most effective at segmenting liver tumors.

For further qualitative comparison, the liver and tumor segmentation outcomes of various methods are illustrated in Fig. 8. From the qualitative evaluation of the three cases, we can observe that there is no apparent difference among the three methods when segmenting images with relatively large lesions. Nevertheless, when the lesion is relatively small, the segmentation result of TFVS is better.

When ground truth cannot be obtained, we can use logic XOR. Without obtaining the ground truth of liver and tumor, the segmentation outcomes of liver and tumor obtained by predicting transverse, sagittal and coronal planes are processed by anisotropy. The positional relationship and the approximate proportion between liver and tumor can be more clearly obtained from the 3D perspective. Absolute difference maps

TABLE II
PERFORMANCE OF LIVER AND LIVER TUMOR SEGMENTATION ON THE 3DIRCADB DATASET. ALL VALUES ARE REPORTED AS MEAN ± STANDARD DEVIATION. BEST SCORES ARE MARKED IN BOLD.

Algorithm	Metrics of liver segmentation				Metrics of liver tumor segmentation			
	Dice (%)	*Precision (%)*	*Recall (%)*	*Jaccard (%)*	*Dice (%)*	*Precision (%)*	*Recall (%)*	*Jaccard (%)*
U-Net [12]	91.45±4.05*	89.41±5.63*	94.13±6.59*	84.49±6.67*	45.90±31.09*	69.09±34.40*	43.91±29.65*	34.81±26.84*
nnU-Net [2]	96.65±1.43	**96.13±2.57**	97.23±1.18	**93.56±2.60**	71.97±24.94*	83.30±30.98*	66.89±22.69*	60.51±24.08*
TFVS	**96.70±1.15**	95.97±2.16	**97.37±1.23**	93.55±2.40	**82.60±7.96**	**93.74±9.78**	**74.90±11.07**	**71.08±11.38**

* indicates that the results are statistically different from TFVS (Wilcoxon signed-rank tests, $p < 0.05$).

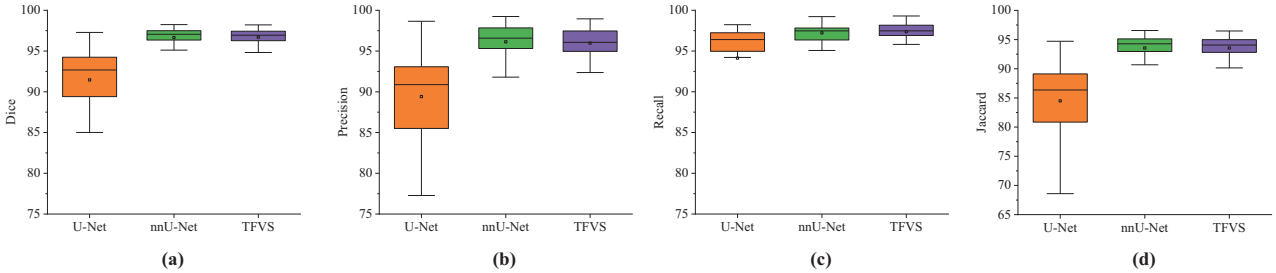


Fig. 6. Performance evaluations of the liver segmentation results (for the 3Dircadb dataset). As in typical box plots, the box represents the first quartile, median and the third quartile from the lower border, middle and the upper boarder, respectively, and the lower and the upper whiskers show the minimum and the maximum values. (a) Dice, (b) Precision, (c) Recall, (d) Jaccard.

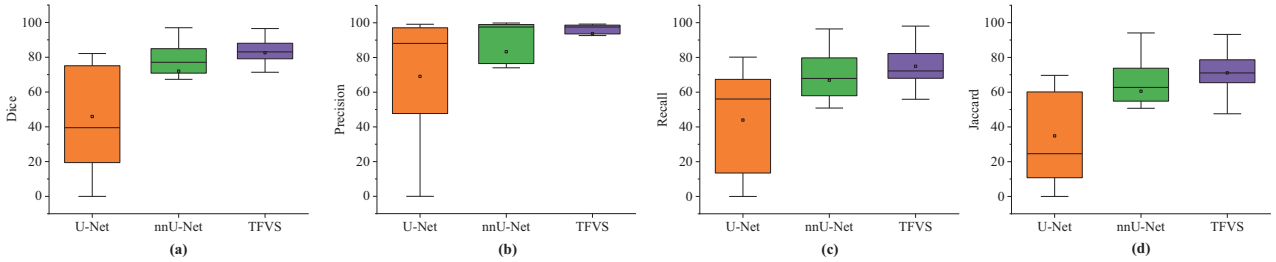


Fig. 7. Performance evaluations of the liver tumor segmentation results (for the 3Dircadb dataset). As in typical box plots, the box represents the first quartile, median and the third quartile from the lower border, middle and the upper boarder, respectively, and the lower and the upper whiskers show the minimum and the maximum values. (a) Dice, (b) Precision, (c) Recall, (d) Jaccard.

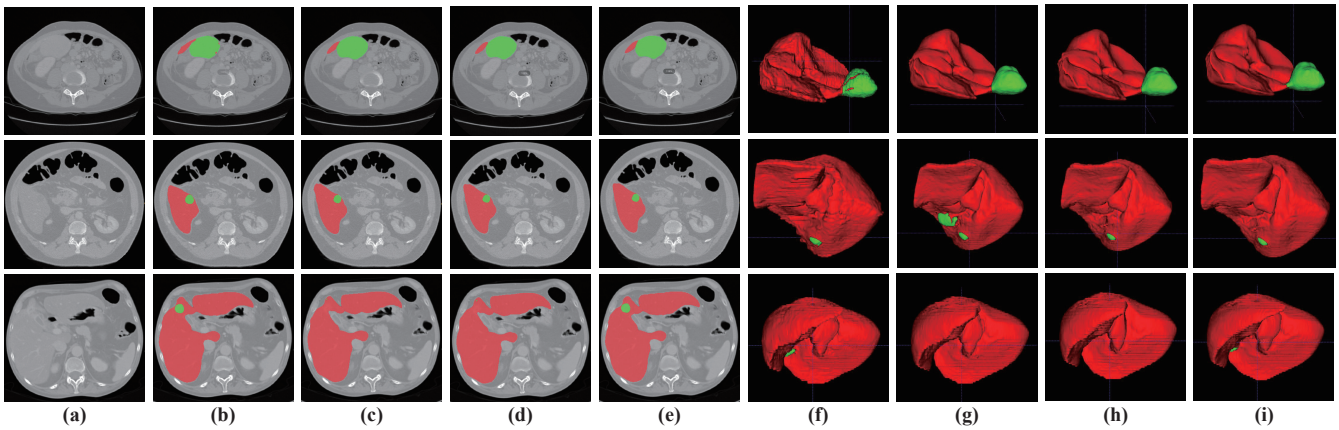


Fig. 8. The visual comparison of liver and tumor segmentation results for the 3Dircadb dataset. The first 5 columns show the 2D results of 3 validation volumes, and the last 4 columns are the corresponding 3D visualization. (a) Original slices (b, f) 2D and 3D ground truth (c, g) U-Net (d, h) nnU-Net (e, i) TFVS. The red regions denote the liver, while the green ones denote tumors.

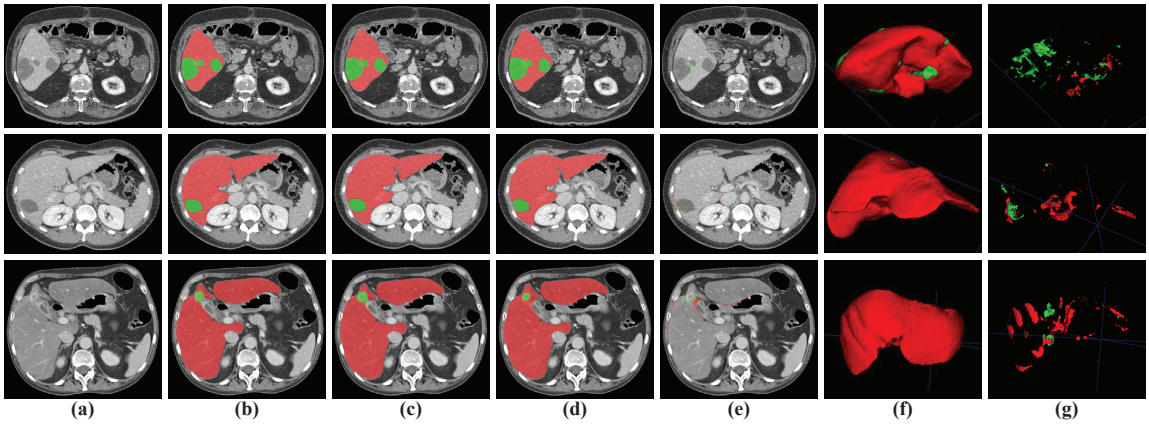


Fig. 9. Absolute difference maps between segmentation results $\mathbf{\Omega}_{3D_Z}$, $\mathbf{\Omega}_{3D_Y}$ and $\mathbf{\Omega}_{3D_X}$. The first 5 columns show the 2D results of 3 validation volumes, and the last 2 columns are the corresponding 3D visualization. (a) Original slice, (b) segmentation result $\mathbf{\Omega}_{3D_Z}$, (c) segmentation result $\mathbf{\Omega}_{3D_Y}$, (d) segmentation result $\mathbf{\Omega}_{3D_X}$, (e, g) 2D and 3D absolute difference map between (b), (c) and (d), (f) 3D prediction result. The red regions denote liver, while the green ones denote tumor.

TABLE III
QUANTITATIVE SCORES OF THE LIVER AND TUMOR SEGMENTATION RESULTS (TRAINING DATA: $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$ AND $\mathbb{D}_{\boldsymbol{X}}$. TEST DATA: 3DIRCADB DATASET). ALL VALUES ARE REPORTED AS MEAN ± STANDARD DEVIATION.

Algorithm	Training Dataset	Metrics of liver segmentation				Metrics of liver tumor segmentation			
		Dice (%)	*Precision (%)*	*Recall (%)*	*Jaccard (%)*	*Dice (%)*	*Precision (%)*	*Recall (%)*	*Jaccard (%)*
nnU-Net	$\mathbb{D}_{\boldsymbol{Z}}$	96.42±1.60	95.88±2.54	97.26±1.28	93.49±2.66	72.94±25.73	82.92±30.60	68.04±23.14	62.04±24.93
	$\mathbb{D}_{\boldsymbol{Y}}$	96.51±1.44	95.55±2.55*	97.34±1.23	93.30±2.63	72.67±24.46	83.08±30.52	67.81±20.55	61.30±23.97
	$\mathbb{D}_{\boldsymbol{X}}$	96.45±1.64	95.52±2.37*	97.41±1.18*	93.40±2.61	73.04±24.04	82.97±28.56	69.13±23.14	62.47±22.22
nnU-Net (ResNet)	$\mathbb{D}_{\boldsymbol{Z}}$	96.54±1.55	95.81±2.69	97.33±1.18	93.36±2.81	70.37±25.46	84.71±31.46	64.47±23.10	58.69±24.53
	$\mathbb{D}_{\boldsymbol{Y}}$	96.56±1.46	95.84±2.54	97.33±1.12	93.39±2.65	74.17±21.36*	89.43±25.26	64.19±22.25	62.62±20.35
	$\mathbb{D}_{\boldsymbol{X}}$	96.63±1.50	96.35±2.52*	96.97±1.31*	93.44±2.73	74.56±20.92*	90.82±6.68	64.61±22.07	65.13±20.18*
TFVS	$\mathbb{D}_{\boldsymbol{Z}}$	96.59±1.46	95.96±2.63	96.97±1.34	93.45±2.65	80.48±7.22	97.36±3.04	69.36±11.18	67.92±10.54
	$\mathbb{D}_{\boldsymbol{Y}}$	96.61±1.36	95.92±2.25	97.08±1.26	93.47±2.49	80.73±10.76*	93.81±11.76*	71.72±12.93*	68.86±13.94*
	$\mathbb{D}_{\boldsymbol{X}}$	96.66±1.37*	95.96±2.30	96.93±1.30	93.46±2.87	80.75±9.02*	93.52±2.41*	68.88±15.22	67.35±14.26

* indicates that the results on datasets $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$ are statistically different on the dataset $\mathbb{D}_{\boldsymbol{Z}}$ (Wilcoxon signed-rank tests, $p < 0.05$).

between segmentation $\mathbf{\Omega}_{3D_Z}$, $\mathbf{\Omega}_{3D_X}$ and $\mathbf{\Omega}_{3D_Y}$ are illustrated in Fig. 9.

B. Ablation studies

We conduct two ablation studies to show that our proposed TFVS method has a better ability than nnU-Net for the liver and tumor segmentation task. The implementation of the residual network on nnU-Net is denoted as nnU-Net(ResNet).

We conduct the first ablation experiment in which nnU-Net, nnU-Net(ResNet) and TFVS are trained on the 3D three-direction datasets $\mathbb{D}_{\boldsymbol{Z}}$, $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$, respectively. Then the 3Dircadb dataset is segmented with these pre-trained models. Table III shows that the models pre-trained with the 3D three-direction datasets provide different results. The segmentation results by the models pre-trained on datasets $\mathbb{D}_{\boldsymbol{Y}}$ and $\mathbb{D}_{\boldsymbol{X}}$ are statistically different from those obtained on the dataset $\mathbb{D}_{\boldsymbol{Z}}$.

Our second ablation experiment is about the fusion method for volumetric segmentation. We improve the performance of nnU-Net and nnU-Net(ResNet) predictions by combining the fusion method for volumetric segmentation as a post-processing technique. As presented in Table IV, the fusion method significantly enhances the overall performance through incorporating spatial information from three directions. Based on the statistical significance tests, we find that the performance metrics (e.g., Precision and Recall of liver segmentation, Dice, Recall, and Jaccard of liver tumor segmentation) achieved by the nnU-Net method in Table IV are significantly different from the nnU-Net method in Table II.

VII. CONCLUSION

A new method has been presented for the problem of liver and tumor segmentation. Our contributions to this challenging problem are as follows:

Model. We have presented the TFVS model for 3D biomedical image segmentation, aiming to address the issues related to single spatial information, boundary ambiguity, and the invasion depth of tumors.

Method. We have proposed the three-direction fusion volumetric segmentation (TFVS) method, which incorporates deep learning techniques and logical operations in Boolean algebra to enhance the accuracy and effectiveness of the segmentation process. The TFVS method considers the vertical, sagittal,

TABLE IV
RESULTS OF QUANTITATIVE PARAMETERS USING THE FUSION METHOD FOR THE LIVER AND LIVER TUMOR SEGMENTATION ON 3DIRCADB DATASET. ALL VALUES ARE REPORTED AS MEAN ± STANDARD DEVIATION. BEST SCORES AS MARKED IN BOLD.

Algorithm	Metrics of liver segmentation				Metrics of liver tumor segmentation			
	Dice (%)	*Precision (%)*	*Recall (%)*	*Jaccard (%)*	*Dice (%)*	*Precision (%)*	*Recall (%)*	*Jaccard (%)*
nnU-Net [2]	96.54±1.46	95.61±2.68*	**97.53±1.20***	93.34±2.66	73.26±26.26*	83.40±31.78	70.62±20.67*	62.61±25.20*
nnU-Net(ResNet) [2]	96.63±1.41	**96.58±2.37**	97.24±1.05	93.51±2.57	74.63±22.10	90.98±25.32	64.29±21.63	62.84±20.59
TFVS	**96.70±1.15**	95.97±2.16	97.37±1.23	**93.55±2.40**	**82.60±7.96**	**93.74±9.78**	**74.90±11.07**	**71.08±11.38**

* indicates that the results are statistically different from the nnU-Net results in Table II (Wilcoxon signed-rank tests, $p < 0.05$).

and coronal axes of 3D images, making it a robust and reliable solution for 3D biomedical image segmentation. Our experiments on the Liver dataset and 3Dircadb dataset have demonstrated the effectiveness of TFVS in accurately segmenting 3D biomedical images, particularly in the case of liver and tumor segmentation.

Our results show that when segmenting images with smaller tumors, the TFVS method outperforms other state-of-the-art models. Additionally, our second ablation experiment highlights the effectiveness of 3D biomedical image segmentation with three-directional information, which enhances segmentation results by providing more spatial information on organs and tumors. This shows that better segmentation results are achieved by incorporating more information from three orthogonal axes (the axial, sagittal, and coronal axes). The TFVS model helps address issues caused by boundary ambiguity and the invasion depth of tumors encountered in conventional segmentation techniques by utilizing this three-direction information. Consequently, the TFVS method shows enhanced segmentation performance and superior generalization capability when addressing liver and tumor segmentation challenges compared to other state-of-the-art methods.

Although the proposed TFVS method offers promising results, it could be further improved in the following aspects. For example, the use of three-direction spatial information could be extended for larger-scale 3D features. In addition, its performance could be further improved when dealing with small targets, e.g. by leveraging three-direction inter- and intra-slice contexts.

ACKNOWLEDGEMENT

The authors would like to thank the associate editor and anonymous reviewers for their constructive comments for improving this paper.

REFERENCES

[1] A. Forner, "Hepatocellular carcinoma surveillance with mirnas," *The Lancet Oncology*, vol. 16, no. 7, pp. 743–745, 2015.

[2] 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*, vol. 18, no. 2, pp. 203–211, 2021.

[3] J. J. Cerrolaza, M. L. Picazo, L. Humbert, Y. Sato, D. Rueckert, M. Á. G. Ballester, and M. G. Linguraru, "Computational anatomy for multi-organ analysis in medical imaging: A review," *Medical Image Analysis*, vol. 56, pp. 44–67, 2019.

[4] J. Yang, X. Huang, Y. He, J. Xu, C. Yang, G. Xu, and B. Ni, "Reinventing 2d convolutions for 3d images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3009–3018, 2021.

[5] K. Yan, J. Cai, Y. Zheng, A. P. Harrison, D. Jin, Y. Tang, Y. Tang, L. Huang, J. Xiao, and L. Lu, "Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in ct," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2759–2770, 2020.

[6] D. Zhang, B. Chen, J. Chong, and S. Li, "Weakly-supervised teacher-student network for liver tumor segmentation from non-enhanced images," *Medical Image Analysis*, vol. 70, p. 102005, 2021.

[7] F.-Z. Mokrane, L. Lu, A. Vavasseur, P. Otal, J.-M. Peron, L. Luk, H. Yang, S. Ammari, Y. Saenger, H. Rousseau, B. Zhao, L. H. Schwartz, and L. Dercle, "Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules," *European Radiology*, vol. 30, no. 1, pp. 558–570, 2020.

[8] F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," *Neurocomputing*, vol. 338, pp. 321–348, 2019.

[9] M. N. Cheema, A. Nazir, B. Sheng, P. Li, J. Qin, and D. D. Feng, "Liver extraction using residual convolution neural networks from low-dose ct images," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 9, pp. 2641–2650, 2019.

[10] Y. Tang, R. Gao, H. H. Lee, S. Han, and B. A. Landman, "High-resolution 3d abdominal segmentation with random patch network fusion," *Medical Image Analysis*, vol. 69, no. 3, p. 101894, 2021.

[11] Y. Xie, J. Zhang, H. Lu, C. Shen, and Y. Xia, "Sesv: Accurate medical image segmentation by predicting and correcting errors," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 286–296, 2020.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.

[13] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view ct," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[14] K. Men, P. Boimel, J. Janopaul-Naylor, H. Zhong, M. Huang, H. Geng, C. Cheng, Y. Fan, J. P. Plastaras, and E. Ben-Josef, "Cascaded atrous convolution and spatial pyramid pooling for more accurate tumor target segmentation for rectal cancer radiotherapy," *Physics in Medicine & Biology*, vol. 63, no. 18, p. 185016, 2018.

[15] 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*, vol. 37, no. 12, pp. 2663–2674, 2018.

[16] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1316–1325, 2020.

[17] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 1471, 2020.

[18] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, "Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 965–976, 2021.

[19] L. Song, H. Wang, and Z. J. Wang, "Bridging the gap between 2d and 3d contexts in ct volume for liver and tumor segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3450–3459, 2021.

[20] S. Di, Y. Zhao, M. Liao, Z. Yang, and Y. Zeng, "Automatic liver tumor segmentation from ct images using hierarchical iterative superpixels and local statistical features," *Expert Systems with Applications*, vol. 203, p. 117347, 2022.

[21] J. Li, G. Huang, J. He, Z. Chen, C.-M. Pun, Z. Yu, W.-K. Ling, L. Liu, J. Zhou, and J. Huang, "Shift-channel attention and weighted-region loss function for liver and dense tumor segmentation," *Medical Physics*, 2022.

[22] D. T. Kushnure and S. N. Talbar, "Hfru-net: High-level feature fusion and recalibration unet for automatic liver and tumor segmentation in ct images," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106501, 2022.

[23] Y. Wang, Y. Zhou, W. Shen, S. Park, E. K. Fishman, and A. L. Yuille, "Abdominal multi-organ segmentation with organ-attention networks and statistical fusion," *Medical Image Analysis*, vol. 55, pp. 88–102, 2019.

[24] H. Zaidi and B. M. W. Tsui, "Review of computational anthropomorphic anatomical and physiological models," *Proceedings of the IEEE*, vol. 97, no. 12, pp. 1938–1953, 2009.

[25] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, "3d deeply supervised network for automated segmentation of volumetric medical images," *Medical Image Analysis*, vol. 41, pp. 40–54, 2017.

[26] B. Zhou, Z. Augenfeld, J. Chapiro, S. K. Zhou, C. Liu, and J. S. Duncan, "Anatomy-guided multimodal registration by learning segmentation without ground truth: Application to intraprocedural cbct/mr liver segmentation and registration," *Medical Image Analysis*, vol. 71, p. 102041, 2021.

[27] N. Capobianco, M. Meignan, A.-S. Cottereau, L. Vercellino, L. Sibille, B. Spottiswoode, S. Zuehlsdorff, O. Casasnovas, C. Thieblemont, and I. Buvat, "Deep-learning 18f-fdg uptake classification enables total metabolic tumor volume estimation in diffuse large b-cell lymphoma," *Journal of Nuclear Medicine*, vol. 62, no. 1, pp. 30–36, 2021.

[28] M. Nagendran, Y. Chen, C. A. Lovejoy, A. C. Gordon, M. Komorowski, H. Harvey, E. J. Topol, J. P. A. Ioannidis, G. S. Collins, and M. Maruthappu, "Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies," *British Medical Journal*, vol. 368, 2020.

[29] R. Dey and Y. Hong, "Hybrid cascaded neural network for liver lesion segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1173–1177.

[30] M. Perslev, A. Pai, J. Runhaar, C. Igel, and E. B. Dam, "Cross-cohort automatic knee mri segmentation with multi-planar u-nets," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 6, pp. 1650–1663, 2022.

[31] M. Perslev, E. B. Dam, A. Pai, and C. Igel, "One network to segment them all: A general, lightweight system for accurate 3d medical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 30–38.

[32] S. Liu and X. Guo, "Improving brain tumor segmentation with multi-direction fusion and fine class prediction," in *International MICCAI Brainlesion Workshop*. Springer, 2019, pp. 349–358.

[33] Y. Xia, F. Liu, D. Yang, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, and H. Roth, "3d semi-supervised learning with uncertainty-aware multi-view co-training," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3646–3655.

[34] V. Sundaresan, G. Zamboni, P. M. Rothwell, M. Jenkinson, and L. Griffanti, "Triplanar ensemble u-net model for white matter hyperintensities segmentation on mr images," *Medical Image Analysis*, vol. 73, p. 102184, 2021.

[35] L. Zhou, X. Meng, Y. Huang, K. Kang, J. Zhou, Y. Chu, H. Li, D. Xie, J. Zhang, W. Yang *et al.*, "An interpretable deep learning workflow for discovering subvisual abnormalities in ct scans of covid-19 inpatients and survivors," *Nature Machine Intelligence*, vol. 4, no. 5, pp. 494–503, 2022.

[36] Y. Li, T. Ren, J. Li, X. Li, and A. Li, "Multi-perspective label based deep learning framework for cerebral vasculature segmentation in whole-brain fluorescence images," *Biomedical Optics Express*, vol. 13, no. 6, pp. 3657–3671, 2022.

[37] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820–838, 2021.

[38] Z. Wu, J. Wei, J. Wang, and R. Li, "Slice imputation: Multiple intermediate slices interpolation for anisotropic 3d medical image segmentation," *Computers in Biology and Medicine*, vol. 147, p. 105667, 2022.

[39] J. Minnema, J. Wolff, J. Koivisto, F. Lucka, K. J. Batenburg, T. Forouzanfar, and M. van Eijnatten, "Comparison of convolutional neural network training strategies for cone-beam ct image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 207, p. 106192, 2021.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 630–645.

[41] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, T. Cao, Y. Zhu, Z. Nie, and X. Yang, "Toward data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation," *Medical Physics*, vol. 48, no. 3, pp. 1197–1210, 2021.

[42] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Conference and Workshop on Neural Information Processing Systems*, 2017.

[43] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "Cpfnet: Context pyramid fusion network for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3008–3018, 2020.

[44] L. Liu, Z. Wen, S. Liu, H.-Y. Zhou, H. Zhu, W. Xie, L. Shen, K. Ma, and Y. Zheng, "Mixsearch: Searching for domain generalized medical image segmentation architectures," *arXiv Preprint arXiv:2102.13280*, 2021.

[45] R. Wang, S. Cao, K. Ma, Y. Zheng, and D. Meng, "Pairwise learning for medical image segmentation," *Medical Image Analysis*, vol. 67, p. 101876, 2021.

[46] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers *et al.*, "The medical segmentation decathlon," *Nature Communications*, vol. 13, no. 1, p. 4128, 2022.

[47] I. France, "3dircadb, 3d image reconstruction for comparison of algorithm database," 2016.

[48] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[49] K. Yasaka, H. Akai, O. Abe, and S. Kiryu, "Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: a preliminary study," *Radiology*, vol. 286, no. 3, pp. 887–896, 2018.

[50] Z. Wang, E. Wang, and Y. Zhu, "Image segmentation evaluation: A survey of methods," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5637–5674, 2020.