2022 16th ICME International Conference on Complex Medical Engineering (CME) | 978-1-6654-9699-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/CME55444.2022.10063259

GI Tract Image Segmentation with U-Net 3D

Feiya Xiang[1],
Johns Hopkins University,
Baltimore, United States,
fxiang2@jhu.edu,

Mingda Huo[2]
Guangzhou China
Jinan University
md.huo@outlook.com

Yunjian Li [1],
Jiangmen, China,
Wuyi University,
2523322908@qq.com

Abstract-- Automatic segmentation of medical images can reduce the workload of clinicians, improve work efficiency, and reduce the probability of misdiagnosis and missed diagnosis. However, due to the difficulty of image feature representation, image segmentation is still one of the most challenging topics in the field of computer vision. Especially in the task of medical image segmentation, it is difficult to extract features with recognition ability because medical images in natural scenes are usually more fuzzy, noisy, and have low contrast than RGB images. In our paper, we utilize Unet3D for stomach and intestines' segmentation, which could assist radiation oncologists to deliver high doses of radiation using X-ray beams pointed to tumors while avoiding the stomach and intestines. We compare our model with other models using the metrics of dice similarity efficiency. Our Unet3D has the highest dice similarity efficiency of 0.868 among all compared models, which shows our model performs best.

Keywords: Index Terms: segmentation, medical images, U-Net 3D, dice similarity efficiency

I. Introduction

With the development and popularization of medical imaging equipment, X-ray, ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) have become four important image aids for clinicians to diagnose diseases, evaluate prognosis and plan operations in medical institutions. In clinical applications, different imaging methods are suitable for the diagnosis of different organs of the human body. The automatic segmentation of these medical images can reduce the workload of clinicians, improve work efficiency, and reduce the probability of misdiagnosis and missed diagnosis. Although many segmentation methods have been known, and they have succeeded in some cases, image segmentation is still one of the most challenging topics in the field of computer vision due to the difficulty of image feature representation. Especially in the task of medical image segmentation, because medical images are generally more fuzzy, noisy, and have low contrast than RGB images in natural scenes, it is difficult to extract features with identification ability. Many segmentation models applied in the field of natural images are not applicable in the field of medical images, and the traditional segmentation methods are relatively simple, and there is no improvement for the medical image scene. Until the emergence of the neural network, the automatic segmentation method of medical images has taken a big step forward. In recent years, with the development of computer technology and hardware equipment, a 3D neural network with a more complex structure and larger parameters has gradually attracted the attention of researchers and is increasingly applied to the field of medical image segmentation.

In our paper, we utilize Unet3D for stomach and intestines' segmentation, which could assist radiation oncologists to deliver high doses of radiation using X-ray beams pointed to tumors while avoiding the stomach and intestines, which could make cancer patients' daily treatments faster and allow them to get more effective treatment with fewer side effects and better long-term cancer control. Figure 1 shows the segmentation process. The tumor (pick thick line) is close to the stomach (red thick line). Thanks to the segmentation, the high dose of radiation can be directed to the tumor while avoiding the stomach.

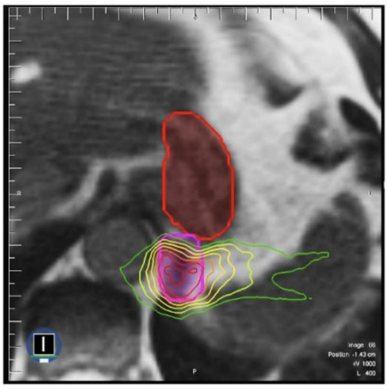

Figure 1: segmentation example

We use the dataset provided by the Kaggle platform. Related work is described in section II, and we introduce our methodology and experiment in sections III and IV.

II. Related Work

- Traditional segmentation methods

Early medical image segmentation methods usually rely on threshold segmentation, region growth, boundary tracking, and so on, The idea of threshold segmentation is to divide the pixels in the image into several categories according to the specific reading value. The gray values of pixels in the same category are similar, while the values of pixels in different categories are quite different. [1] A multi-threshold liver segmentation model is proposed. The selected threshold is determined by the histogram slope difference distribution of liver CT image: region growth refers to growing the image from a seed point or a region to a larger region and ensuring that all pixels in the current region have similar attributes in the process of growth. However, there are some problems in the application of traditional methods, such as a large amount of calculation, slow running speed, low segmentation accuracy, and so on. Nowadays, the method based on deep learning not only has better efficiency than the traditional method but also far exceeds the traditional method in accuracy.

- 2D deep learning segmentation methods

With the development of modern computer equipment and artificial intelligence technology, deep learning methods are gradually emerging, have better performance than traditional algorithms, and have also been widely used in medical image task scenarios. The proposal of a full winder neural network (FCN) [2] has greatly contributed to the medical image segmentation field. It successfully realizes the hierarchical feature representation of images. The shallow network is used to extract local features of images, and the deep network pays more attention to global features. Moreover, FCN is not easily affected by noise, contrast, and other information, so it has higher accuracy than general models in the field of medical images. As an improved network of FCN, u-net5 has brought a major breakthrough in the field of image segmentation. At present, the most widely used deep learning segmentation method mainly includes two architectures, one is to superimpose convolutional blocks like FCN, and the other is an encoding and decoding structure similar to u-net. However, no matter what kind of architecture, the mainstream deep learning methods are mostly limited to the 2D level, and cannot effectively use the spatial information of the third dimension of data. [3] A two-stage framework for 2D liver and tumor segmentation is proposed. In the first stage, the liver is roughly segmented. In the second stage, an edge enhancement network is designed to enhance the edge information of the liver and tumor, and the edge distance map between the liver and tumor boundary is designed as an additional monitoring signal to train the edge enhancement network, which significantly improves the segmentation performance of the liver and tumor. [4] An improved u-net network is proposed and optimized by using three mixed loss functions: cross entropy, intersection union ratio, and structural similarity. [5] It is improved in the encoder and decoder of the u-net network. The pre-trained ResNet-34 network is used in the encoder to extract features, and the bidirectional convolution long-term and short-term memory module fusion features are introduced into the decoder, which proves the effectiveness of the network on the lung CT public data set.

- 3D deep learning segmentation methods

3D convolution can better mine the spatial correlation of data and can make up for the deficiency of 2D segmentation spatial information extraction ability. As one of the most classic segmentation networks in the field of medical image segmentation, the proposed 3d-CNN architecture has reached the most advanced level in the field of brain injury automatic segmentation. [6] etc. Sichuan applied it in the field of brain tumor segmentation, and designed two different resolution channels, integrating low-resolution features and high-resolution features. [7] A 3D full convolution neural network with a deep supervision mechanism is proposed, which can accelerate the optimization process of the network and improve the segmentation efficiency of the network. [8] A multi-scale 3D full convolution network is proposed, and the complementary information of multimodal MR data is extracted for training. As an extension of u-net in 3D, v-nel14 introduces a short-distance jump connection in each stage and can carry out end-to-end training on 3D data such as CT and MRI.

However, 3D segmentation also has shortcomings, that is, it takes up too many computer resources. Therefore, the common 3D segmentation algorithms are down sampling the input data [8], which can increase the width of the network to extract more information. Even some Pseudo-3D algorithms [9][10] only take a very small number of slices as the model input, and these methods have more or less lost spatial information. In this paper, the complete 3D data block is used as the input of the neural network to complete the segmentation of liver and liver tumors, and the influence of different size input data on the effect of the model is discussed, which provides a certain reference for the design and optimization of 3D convolution neural network.

- Our Contribution

✓ We utilize U-Net3D as our model to segment the stomach and intestines.

✓ We introduce our dataset and do some analysis.

✓ In the process of the experiment, we do the comparing experiments and the result shows that our model performed better than the other models.

III. METHODOLOGY

U-net is widely used in the field of medical image segmentation [11]. It innovatively proposes to combine shallow and deep features by jumping connection, which solves the problems of the low contrast of the medical image and fuzzy boundary of the segmentation target. Although u-net appeared as early as 2015, it is still the most popular network framework in the field of medical image segmentation. At present, most of the data directly output from CT or MRI are stored in 3D formats such as nifti. Using 3D convolution operation, the voxel correlation features in the whole 3D space can be extracted, and the end-to-end mapping of 3D volume data from input to output can be realized. Extend u-net to 3D data for application, that is, 3D u-net network[12], as shown in Figure 2.

The original 3D u-net is mainly divided into two parts: encoder and decoder. The encoder is used to extract the features of the input data and store them temporarily in the form of a multi-channel feature map, while the decoder is used to reconstruct the extracted features and restore the size of the original data. The encoder is composed of multi-layer convolution and pooling operations, in which the convolution layer is used to extract data features, and the pooling layer is used to reduce the data dimension and the storage space occupied by features, so that more features can be extracted in parallel. The decoder consists of convolution and deconvolution operations, in which the function of the convolution layer is the same as that of the encoder, and the function of the deconvolution layer is opposite to that of the pooling operation. It is used to restore the original dimension of the image and ensure that the data size of the same depth of the neural network is the same, which is convenient for jumping and connecting to fuse the characteristics of different levels of data. Finally, the category probability of each voxel is output through the activation function.

To sum up, the encoding and decoding structure of 3D u-net can well capture the context information of 3D data, integrate different levels of features, and improve the segmentation effect of medical images from a global perspective.

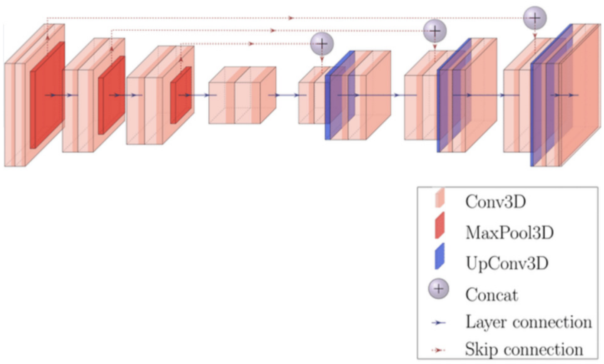

Figure 2: our model's Architecture

IV. EXPERIMENTS

- Experiments data

The training annotations are provided as RLE-encoded masks, and the images are in 16-bit grayscale PNG format. Figure 2 shows the input figure. Figure 4 shows the percent training images mask with masks, where large bowel accounts for 36.6%, the highest percentage among the three types.

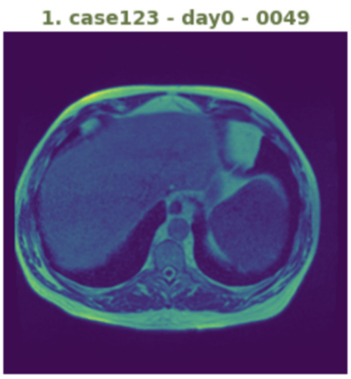

Figure 3: input MRI scan

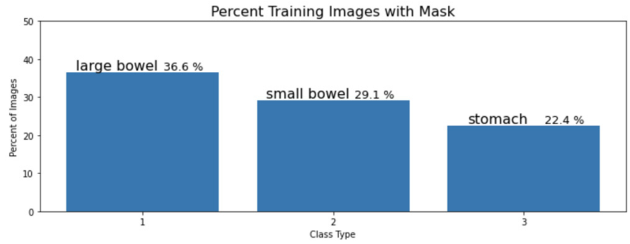

Figure 4.the percent training images mask with masks

- Experiments metrics

In this paper, dice similarity efficiency (DSC) [13] is used to calculate the accuracy of segmentation results, and to evaluate the performance of the segmentation algorithm. The calculation formulas are as follows:

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

X represents the real sample label area; Y indicates the predicted output result area $|X|$ represents the area (or volume) of the real sample label area, $|Y|$ represents the area (or volume) of the predicted output result area. DSC$\in$ [0, 1] indicates the phase degree between the output result and the sample label. The larger the value, the higher the accuracy of the segmentation result.

- Experiment setting

Table I shows what hyperparameters we choose in our training process.

TABLE I. experiment setting

Learning rate	1e-4
Batch size	12
channels	(32, 64, 128, 256, 512)
strides	(2, 2, 2, 2)
kernel_size	3

- Experiment results

We do compare experiments using the same metrics and the same dataset. The experimental results of competing models and our model are shown in table II. We compared other models with Unet3D. The higher DSC is, the more optimal the model will be

TABLE II. compared experiment and results

Models	**DSC**
Unet(ResNet50)	0.754
Unet(EfficientNetB0)	0.748
Unet2.5D(EfficientNetB1)	0.841
Unet3D	0.868

We can see the result that our Unet3D owns the highest DSC 0.868 among these models, which shows our model is the most excellent.

V. CONCLUSION

In our paper, we utilize U-Net3D for stomach and intestines' segmentation, which could assist radiation oncologists to deliver high doses of radiation using X-ray beams pointed to tumors while avoiding the stomach and intestines. The dataset is provided by the Kaggle platform. Related work is described in section II, and we introduce our methodology and experiment in sections III and IV. In the future, we will improve our model.

REFERENCE

[1] SIRI S K, KUMAR S P, V. LATTE M. Threshold-Based New Segmentation Model to Separate the Liver from cT Scan Images [J]. IETE Journal of Research, 2020: 1-8.

[2] RONNEBERGER o, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]. Medical Image Computing and Computer-Assisted Intervention, MICCAI, 2015: 234-241.

[3] TANG Y, TANG Y, ZHU Y, et al, E2 Net: An Edge Enhanced Network for Accurate Liver and Tumor Segmentation on CT Scans [c], Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2020: 512-522.

[4] JALALI Y, FATEH M, REZVANI M, et al. ResBCDU-Net: A Deep Learning Framework for Lung CT Image Segmentation. Sensors, 2021, 21(1): 268.

[5] WARDHANA G, NAGHIBI H, SIRMACEK B, et al. Toward reliable automatic liver and tumor segmentation using convolutional neural network based on 2.5D models [J]. International Journal of Computer Assisted Radiology and Surgery,

2021, 16(1):41-51.

[6] KAMNITSAS K, FERRANTE E, PARISOT S, et al. DeepMedie for Brain Tumor Segmentation [J]. Brainlesion: Glioma, Multiple Sclerosis, Stroke and

Traumatic Brain Injuries, 2016, 2016, 10154: 138- 149,

[7] DOU Q, YU L Q, CHEN H, et al. 3D deeply supervised network for automated segmentation of volumetrie medical images [J]. Medical Image Analysis, 2017, 41: 40-54.

[8] LI x M, DOU Q, CHEN H, et al. 3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images [J]. Medical Image Analysis, 2018, 45: 41-54.

[9] VALANARASU J M J, SINDAGI V A, HACIHALILOGLU I, et al. KiU-Net: Overcomplete Convolutional Architectures for Biomedical Image and Volumetric Segmentation [J]. IEEE Transactions on Medical Imaging, 202: l-1.

[10] LI X M, CHEN H, QI X J, et al. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes [J]. Ieee Transactions on Medical Imaging, 2018, 37(12): 2663-2674.

[11] ZYUZIN V, CHUMARNAYA T. Comparison of Unet architectures for segmentation of the left ventricle endocardial border on two-dimensional

ultrasound images IC. Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Apr 25-26, 2019; 110-113.

[12] CICEK 0, ABDULKADIR A, LIENKAMP S S, et al, 3D U-net: Learning dense volumetric segmentation from sparse annotation [C], Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture

Notes in Bioinformatics), 2016: 424-432.