Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230209

Research on medical image segmentation technology based on deep learning

Yiwen Zhang

College of Electronic and Information, Soochow University, Suzhou, China

1928401009@stu.suda.edu.cn

Abstract. With economy expanding quickly, people's demand for medical services has become higher and higher, and medical image as an important basis for medical diagnosis has naturally received widespread attention. However, traditional image segmentation methods are easily affected by noise and unable to meet the complex and changing practical clinical applications. The increasing utilization of deep learning technology enables effective resolution of these problems. In this paper, we will first introduce the traditional image segmentation techniques, and describe the main methods to realize traditional image segmentation and its limitations. Immediately after that, it is proposed that deep learning methods can solve the challenges of medical image segmentation with traditional methods, and then the structure, algorithms and applications of several of the most commonly used deep learning methods are introduced. This paper proposes that medical image segmentation based on deep learning can segment the image more robustly and with high accuracy, and it can automatically obtain the most suitable features. The research in this paper will be of great value to the research and application of medical image segmentation technology based on deep learning.

Keywords: medical image segmentation, deep learning, convolutional neural network, full convolutional network, U-Net.

1. Introduction

Image segmentation refers to the division of an image into several disjoint regions based on the geometric shape, spatial texture, grayscale, color and other relevant features of the image. The segmentation of the various parts are independent of each other, and the segmentation of the image features in the same region shows consistency and similarity. In contrast, the features of different regions are different.

The medical image segmentation technology studied in this paper involves statistical algorithms, traditional image segmentation methods, deep learning technology and artificial neural network segmentation algorithms, and the need for pathology analysis, clinical diagnosis and other aspects of professional medical knowledge. In addition medical images, compared to other natural images, revolve around complexity and variability, while the various ways the human body is represented also lead to different characteristics. There are also several major differences between medical imaging and general images as follows:

1. the human body has a huge number of tissues and organs, and it has no fixed shape, complexity and diversity;

2. the individual differences between people are large;

3. the overall image contrast is small, and each tissue and organ has a close connection, so the edges of each tissue and organ are blurred;

4. more noise on medical images.

Based on these differences, automatic segmentation of medical images becomes very difficult. At the same time, the scenarios in which medical image segmentation technology is applied are also very wide. Medical image segmentation technology makes the anatomical or pathological structural changes in the image clearer. It can accurately extract the lesion area from the medical image, which greatly improves the diagnostic efficiency and accuracy [1]. Medical image segmentation technology is usually applied to cell segmentation [2], brain and brain tumor segmentation [3], and cardiac image segmentation [4], etc., which contributes significantly to intelligent medical treatment and computer-aided diagnosis. In addition, medical image segmentation techniques can be used for image-guided surgery, where surgery guided by images optimized for visualization and contrast can greatly improve edge detection and reduce unnecessary resection of healthy tissue. Medical image segmentation technology can also make the original large medical image is divided to retain only the effective information, greatly saving the storage space to more efficiently complete the image data compression and transmission and other operations. The further construction of the Internet hospital provides great help.

Although medical image segmentation technology has a deep theoretical foundation and practical application value, two weaknesses cannot be ignored: its high sensitivity to noise and the continuity of the edge pixel values. These problems lead to the fact that in practice, it is easy to appear in the image of the target object there is part of the edge blurring or edge discontinuity and other phenomena, which will greatly affect the processing of the image, thus giving the realization of medical image segmentation has caused significant limitations. As big date is used widespread and the machine processing capabilities have substantial improvement, applying deep learning methods seems to solve the above problems.

This paper will introduce the principle and structure of deep learning based medical image segmentation technique and how to solve the above problems with it.

2. Traditional image segmentation technology

1. Traditional image segmentation techniques employ several methods for segmentation, including threshold-based approaches, region growing, classifier-based approaches, clustering algorithms and so on.

2. Thresholding-based method achieves segmentation of an image by selecting a threshold for dividing the categories based on the grayscale features of a scalar image, and then a thresholding program compares the grayscale value of each pixel in the image with the selected threshold, and then according to the comparison result, pixels in the image are classified as one category when their intensities are greater than the threshold, and the other pixels are classified as another category. When segmenting pictures with various structures that have some quantitative characteristics or contrasting intensities, thresholding is a straightforward yet often successful technique. The thresholding approach has two key drawbacks: it can only create two categories in its most basic form, and it cannot be used to segment multi-channel pictures when the image includes numerous extraction targets. In addition, the selection of the threshold value only takes into account the pixel value characteristics of the pixel point itself and ignores the spatial characteristics of the image, resulting in the threshold segmentation method being very sensitive to the intensity of the noise appearing in the image and the situation where each object in the image has a large number of overlapping gray values and uneven intensity, making it difficult to obtain the desired segmentation results.

3. Region growing is a computational methodology utilized to extract interconnected sections inside an image, employing predetermined criteria as the basis for segmentation, which requires the operator to manually select a seed point and then extract all pixels with the same intensity value that are connected to the initial seed point. Unlike the thresholding method, the region growing method takes

into account the spatial information of the image. However, region growing method obtains the seed point by manual human-computer interaction, so each region to be extracted must have a seed point, which makes the operation more complicated and less flexible. In addition, the region growing method is more sensitive to noise, which may lead to holes or even breaks in the extracted regions.

4. Classifier method is a pattern recognition approach that divides the feature space in an image using data with predetermined labels. Classifiers are relatively computationally efficient since they don't require iteration. Classifiers, as opposed to thresholding techniques, may be used on multi-channel pictures. Classifiers' drawback is that they need user input to get training data. Each picture that has to be segmented can have the training set gathered, but doing so takes a lot of time and effort. However, employing the same training set over a large number of scans would provide biased findings and ignore patient-specific anatomical and physiological variations.

5. Clustering techniques facilitate the process of segmentation largely by using pre-existing statistical information, effectively carrying out similar tasks as classifier methods but without relying on training data. Due to this rationale, clustering algorithms are commonly denoted as unsupervised techniques. In order to address the limited availability of training data, clustering techniques employ an iterative process that involves both picture segmentation and attribute description for each group. Clustering algorithms provide the ability to autonomously train themselves by utilizing the data that is readily accessible. Clustering techniques, by virtue of not necessitating training data and without direct integration of spatial modeling, clustering algorithms have a significant advantage for fast computation and play a great role in processing robust tasks with uneven intensities such as MRI images. However, at the same time the lack of spatial modeling also leads to the sensitivity of clustering algorithms to noise and intensity inhomogeneity.

6. Traditional segmentation methods have the advantage of simplicity and ease of implementation. Still, they are susceptible to noise, but medical images often have more noise points, which leads to the limitations of traditional segmentation methods in clinical applications. In addition, the traditional segmentation method is based on certain set of artificial features to segment the target image, if the selected artificial features can not represent the distribution of the target image well, then the subsequent segmentation results based on this feature will be poor.

3. Deep learning based medical image segmentation technology

Unlike traditional segmentation methods that extract fixed features set manually, deep learning methods can learn the most suitable features for the distribution of the sample data, which are not affected by noise as traditional segmentation methods are generally susceptible to. The most suitable features are automatically obtained by learning the target data to meet the complex and changing clinical application requirements.

Deep learning is good at dealing with unstructured data on which it enables computational models to learn features incrementally from multi-level data. Figure 1 shows a Venn diagram about deep learning [5].

DOI: 10.54254/2755-2721/32/20230209

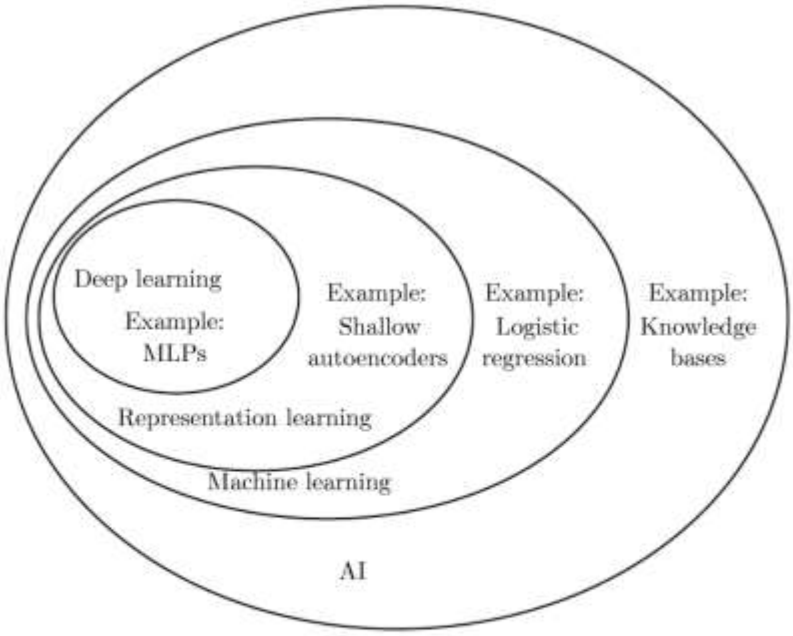

Figure 1. A Venn diagram about deep learning [5].

3.1. Convolutional neural network

Convolutional Neural Network (CNN) is the core of the whole deep learning application in computer vision. Figure 2 depicts a simple CNN architecture designed for the purpose of classifying the MNIST dataset [6].

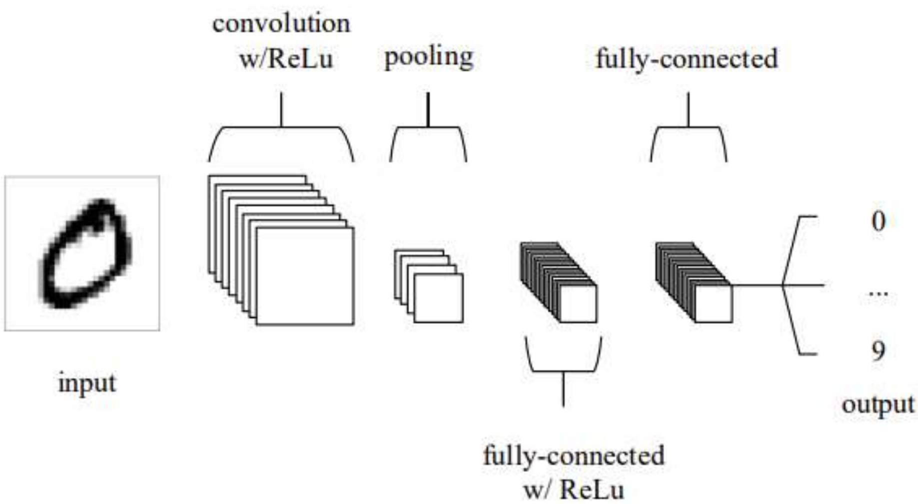

Figure 2. A simple CNN architecture [6].

Convolutional layers are present for feature extraction from the image. Each pixel is convolved with trainable weight filters to produce a new feature map, which is then fed to the activation function. The activation function receives the output from the upper layers and passes it to the math function. It contains two commonly used activation functions: Sigmoid and Relu. The main role of the activation function is to help the artificial neural network to learn and understand the very complex and nonlinear data.

Pooling layers employ a filter to systematically examine the entirety of the input, but the filter utilized in the pooling operation does not include any weights. The essence of the pooling operation is a kind of downsampling, which completes the features' compression and downsampling. Pooling operations are also performed on the input image in a sliding fashion, where representative values are selected and output according to some rule for the neuron activation values in the region covered by each pooling kernel. In most CNNs, the pooling layer is maximum, which means that the maximum activation values in the region of the pooling kernel are selected as the output values for downsampling. Maximum pooling can increases the invariance of the local movement of objects in the input image.

A fully connected layer, which is in general found at the end of the CNN, contains neurons with directly connections to neurons in both neighboring layers, not connected to either of them, and its role is to purify and synthesize the multidimensional features and pass them to the subsequent regression analysis layer or classifier to accomplish the corresponding tasks.

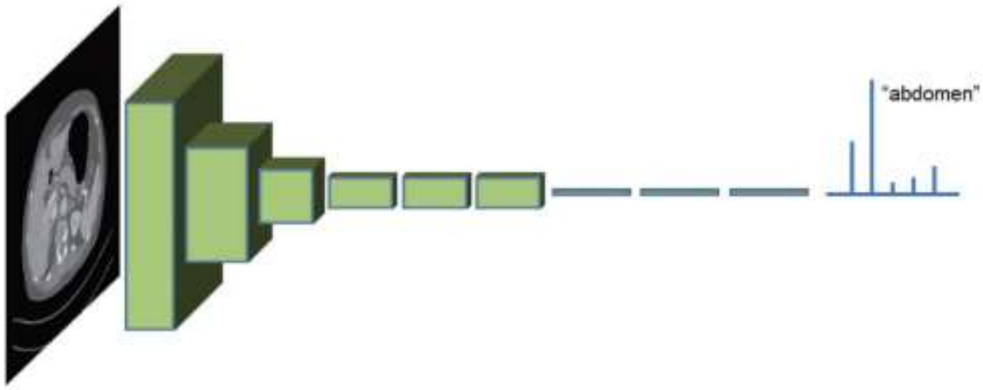

Figure 3. Convolutional neural network (CNN) [7].

Figure 3 shows an overview schematic of a representative CNN architecture which is capable of producing predictions for individual images through the utilization of softmax outputs for the purpose of multi-class categorization [7]. CNNs have been successfully applied to many medical image segmentation tasks. Kayalibay et al. demonstrated a CNN-based medical image segmentation method for bone and tumor segmentation tasks for hand and brain MRI, respectively [8].

However, the utilization of the CNN model for image classification leads to the compression of the 2D matrix information in the original image by the fully connected layer. Consequently, this compression results in the loss of spatial information inside. Given the significance of spatial information in semantic segmentation tasks, its influence on the use of CNN models for picture segmentation is noteworthy. Deep learning image segmentation algorithms were mainly realized by sliding image blocks in the early days, i.e., a fixed-size image block is intercepted around the target pixel and fed into the CNN. The classification result obtained is the category to which the current pixel belongs. The image block sliding method has many repetitive computation operations and low efficiency, and the segmentation accuracy is directly limited by the image block size, which has certain limitations.

3.2. Full convolutional network

Long et al. [9] proposed a Full Convolutional Network (FCN) to overcome the limitations of CNN. FCN designs an end-to-end, pixel-to-pixel encoding and decoding structure that solves the semantic segmentation problem of CNN and reduces the loss of spatial information. Meanwhile, FCN can take inputs of any size and, using effective inference and learning, generate outputs of the same size, eliminating the limitation of CNN that has limitations on image size. Fig. 4 depicts the standard configuration of an FCN for the semantic segmentation of image slices, acquired by using computed tomography (CT) with ConvNets.

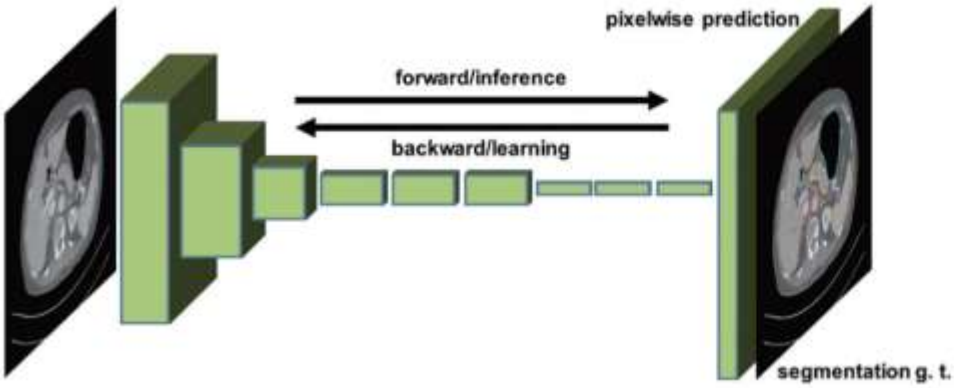

Figure 4. Fully convolutional network (FCN) [7].

FCN networks add jump-linking devices of different layer degrees to the part of the coding layer, which is implemented by summing the convolution with the corresponding anti-convolution for feature fusion. In FCN, a transposed convolutional layer replaces the final densely connected layer of the CNN in order to upsample the network's low-resolution feature maps. While performing semantic segmentation, this process restores the input image's original spatial dimension, thus solving the semantic segmentation problem of CNNs and reducing the loss of spatial information. FCNs only need to compute the softmax at each pixel of the final feature maps. In this way the shallow representational information is used to complement the spatial details of the deeper semantic information, thus achieving a more efficient and efficient feature fusion. information to complement the spatial details of the deeper semantic information to achieve more accurate results and ensure the network's robustness.

FCN can effectively learn to make dense predictions for per-pixel tasks. The proposed FCN eliminates the limitation of input image size, simplifies the preprocessing process, and reduces the loss of spatial information. However, the results obtained by a simple upsampling operation are still not fine enough, and the segmented output maps are still blurry, smooth, and insensitive to the details in the image.

3.3. U-Net

Based on FCN, U-Net has made further improvements. Figure 5 shows the structure of U-Net [10]. The difference between U-Net and FCN is that the jump link operation of U-Net is a superposition operation carried out together with the deconvolution simultaneously, and the shallow representational information gives the deep semantic information. U-Net also uses mirroring operation for the edge processing of convolution to ensure that the corresponding encoder and decoder have the same size. Gordienko et al. performed lung segmentation experiments on chest X-ray images using the U-Net network, and the results show that U-Net network can perform medical image segmentation quickly and accurately.

DOI: 10.54254/2755-2721/32/20230209

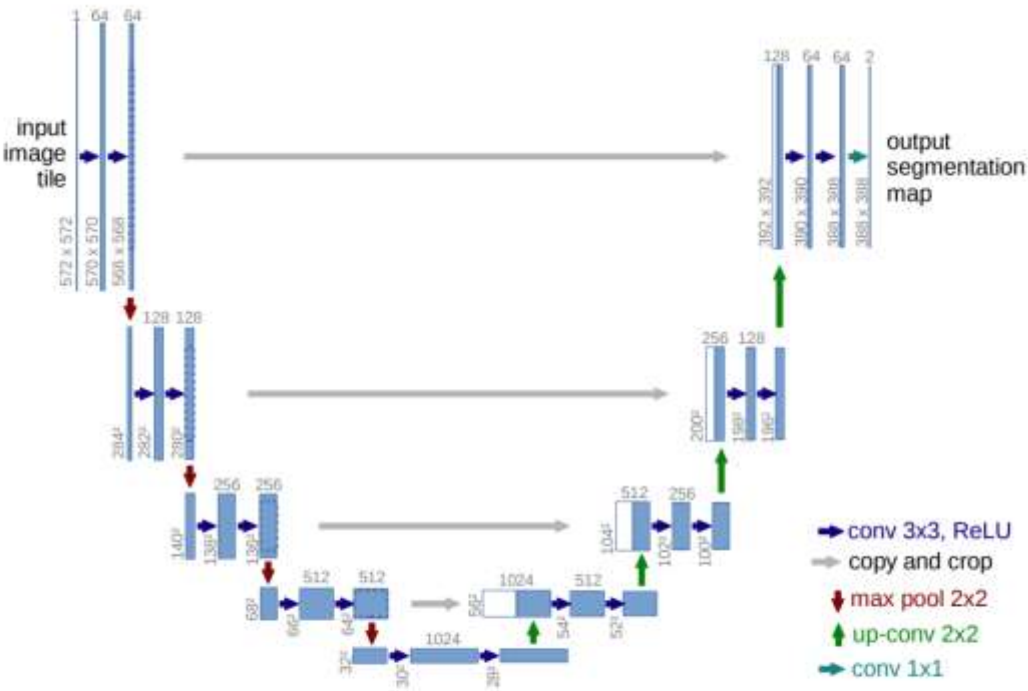

Figure 5. U-Net architecture [10].

Since U-Net is proposed for medical image segmentation, it has attracted extensive attention from research scholars as soon as it was proposed. In 2016, Cicek et al. extended the original U-Net network architecture so as to establish a 3D U-Net network architecture [11]. The authors proposed that U-Net was originally designed for cellular segmentation of 2D images, whereas much of the medical image data is actually volumetric data in 3D. Although it is possible to split the volumetric data into 2D image sequences for processing, this approach ignores the positional relationships between the different layers and often the images at different positions differ significantly, which is not conducive to the network learning generic features. The authors found that biomedical images are very rich in volumetric data. The computer screen can only show 2D slices, making it challenging to mark the segmentation labels immediately on the 3D level. However, neighboring 2D slices often contain approximate picture information, so 3D U-Net can learn to generate high-density volumetric segmentation by simply training on sparsely labeled 2D images. 3D U-Net input 3D volume and processes it by replacing U-Net's original 2D operations with corresponding 3D operations with relatively good experimental results.

In addition, Zhou et al. further optimized U-Net in 2018 and proposed UNet++ to meet the demand for more accurate segmentation [12]. UNet++ adds more jump connection paths and up-sampled convolutional blocks to the original UNet network architecture, and the intermediate hidden layer deep supervision is used, that spans the semantic divide between encoder and decoder, and solves the challenges of gradient vanishing during UNet++ network training and reduces the inference time of the model. The experimental results in Figure 6 demonstrate that UNet++ performs better than U-Net.

DOI: 10.54254/2755-2721/32/20230209

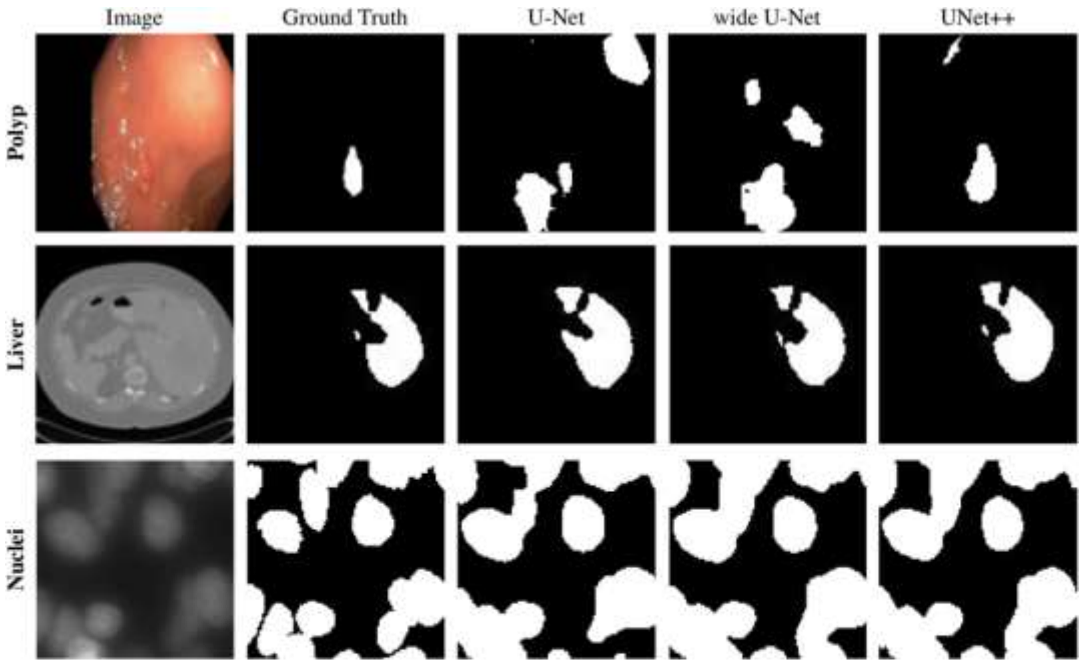

Figure 6. Segmentation results based on U-Net and UNet++ [12].

In recent years, more and more improvements and variants based on U-Net have been continuously proposed. Huang et al. proposed Unet 3+ in 2020, which further proposes a hybrid loss function to improve the borders and receive better segmentation outcomes; Cao et al. proposed Swein-unet, which extract context features using hierarchical SwinTransformer with shifted windows as the encoder. These researches have promoted the development of medical image segmentation technology and also expanded new ideas for future development.

4. Conclusion

This paper provides an overview of the importance and research background of medical image segmentation technology and explains the task of it. This paper introduces the principles of traditional image segmentation methods and their limitations. This paper reviews deep learning based medical image segmentation technology, introduce CNN, FCN, U-Net and its variants, and their respective principles, structures and applications are described.

Based on the research of this paper, it is known that the traditional medical image technology is easily interfered by noise and less flexible, while deep learning-based one is able to segment the image with more robustness and high accuracy, and can automatically obtain the most suitable features, which makes up for the shortcomings of the traditional ones. Deep learning-based medical image segmentation has brought a qualitative leap for medical image processing compared with traditional image segmentation techniques, and still has great potential for future development. For example, it can realize real-time medical image segmentation by compressing the model while ensuring the accuracy and stability; for example, uncertainty analysis algorithms can be added to allow the model to give the segmentation results while pointing out uncertain segmentation, so that the doctor can intervene to correct the segmentation, and ensure the quality of the segmentation and the results in the actual clinical application; for example, it can leave the support of large-scale high-quality labeled datasets, improve the weakly supervised learning under sparse labeling and data enhancement under small datasets, and expand the training set for the deep model, and so on, all of which are the directions for future development and research. Using the deep learning method, we can accomplish many complex medical image segmentation that are difficult to realize now, with great research significance and practical application value.

References

[1] Wang R, Lei T, Cui R, Zhang B, Meng H and Nandi A K 2022 Medical image segmentation using deep learning: A survey *IET Image Process.* **16** 1243–67
[2] Wang A, Zhang Q, Han Y, Megason S, Hormoz S, Mosaliganti K R, Lam J C K and Li V O K 2022 A novel deep learning-based 3D cell segmentation framework for future image-based disease detection *Sci. Rep.* **12** 342
[3] Zheng P, Zhu X and Guo W 2022 Brain tumour segmentation based on an improved U-Net *BMC Med. Imaging* **22** 199
[4] Fu Z, Zhang J, Luo R, Sun Y, Deng D and Xia L 2022 TF-Unet:An automatic cardiac MRI image segmentation method *Math. Biosci. Eng.* **19** 5207–22
[5] Bengio Y, Goodfellow I, Courville A 2017 Deep learning *Cambridge,MA, USA: MIT press*
[6] O'Shea K and Nash R 2015 An Introduction to Convolutional Neural Networks
[7] Roth H R, Shen C, Oda H, Oda M, Hayashi Y, Misawa K and Mori K 2018 Deep Learning and Its Application to Medical Image Segmentation **36**
[8] Kayalibay B, Jensen G and van der Smagt P 2017 CNN-based Segmentation of Medical Imaging Data
[9] Long J, Shelhamer E and Darrell T Fully Convolutional Networks for Semantic Segmentation
[10] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation
[11] Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O 2016 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation
[12] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N and Liang J 2018 UNet++: A Nested U-Net Architecture for Medical Image Segmentation *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 3–11