Medical Image Segmentation Based on Deep Learning

Abstract: As a new biomedical image processing technology, medical image segmentation is making a significant contribution to sustainable medical care. It has become an important research direction in the field of computer vision. With the rapid development of deep learning, medical image processing based on complex and deep neural networks has become the focus of research. This article focuses primarily on deep learning-based medical image segmentation. First, I will introduce the basic concepts and characteristics of medical image segmentation based on deep learning. By building a research context, we will summarize three common methods of medical image segmentation and their limitations are summarized and expand the direction of future development. Despite the success in segmenting medical images in recent years, segmenting medical images based on deep learning still faces many research challenges. For example, the segmentation precision is not very accurate, the number of medical images in the dataset is small, and the resolution is low. Inaccurate segmentation results may not reflect actual clinical needs.

Introduction

Image segmentation is a classic problem in computer vision research, has become a focus in the field of image understanding and one of the most difficult problems in image processing since its complexity. It is affected by many aspects, including noise, low contrast, illumination, and irregularity of object boundaries. Image segmentation is usually used to locate objects and boundaries (lines, curves, etc.) in an image. More precisely, image segmentation is the process of assigning a label to each pixel in the image so that pixels with the same label have certain characteristics. Currently, methods of image segmentation are evolving towards a faster and more precise direction.

With the advancement of medical technology, new medical imaging devices are becoming more and more popular. The main types of medical imaging widely used in the clinic are computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), X-ray, and ultrasound image (UI). In addition, it also includes some common RGB images, such as microscope images and fundus retina images. There is some very useful information in medical imaging. Doctors use CT scans and other medical images to do justice to the patient's condition, which gradually becoming the mainstay for the clinical diagnosis of D-cutter. As a result, the study of medical image processing has become a research axis in the field of computer vision.

With the rapid development of artificial intelligence, in special deep learning (DL), image segmentation methods based on learning and learning have achieved good results in the field of image segments. Compared to traditional methods of machine learning and computer vision, deep learning has some advantages in terms of segmentation accuracy and speed. Therefore, learning to deepen medical images segments can help physicians determine the size of a diseased tumor, quantitatively

evaluate efficacy before and after treatment, and significantly reduce workload of the doctor.

To summarize the different methods, we search for the keywords "Medical Image Processing" or "Deep Learning" in Google Scholar and Archive for the latest publications. The paper I chose is mainly based on the deep learning method. We guarantee that the results of all papers are reviewed. To date, there is no universal absolute image segmentation method, but there is a basic consensus on the general law of image segmentation, which produces a number of research results and methods. This article focuses on the classic and the latest methods of the past. First, it focuses on the application of deep learning techniques in the medical image department. A more in-depth study of its network structure and methods. Also analyze its advantages and disadvantages. Second, I share many medical image segmentation metrics for evaluation and networking.

Applications of Deep Learning in Image Segmentation

Deep learning is the driving force behind the development of the imaging field, such as image classification and segmentation. Image division is different from image classification. The image classification method only shows the categories to which the entire image belongs, but splitting an image requires identification information for each pixel in the image.

The Semantic Segmentation of Complete Convolutional Networks Study is the first paper to apply deep learning to image segmentation and has achieved excellent results. After that, many of the segmentation models were taken from FCN. This network is affected by the structure of the VGG network. It no need to enter the dimensions of the image. All classes are fully integrated, which is a new perspective. However, the results obtained after the FCN segmentation are still not smooth, vague enough. It is not sensitive to details of the image. Subsequently, Ronneberger et al. [2] proposed U-Net due to the lack of training images in biomedical imaging. This network has two advantages: First, the network can determine the location of the target category. Second, the input training data is a patch, equivalent data augmentation, which solves the problem of low number of biomedical images. Most medical images are usually three-dimensional, such as CT scans and MRIs. Although the CT image we usually see is a two-dimensional image, it is only a fragment of the two-dimensional image. Therefore, if you want to segment some diseased tissue, you need to use a three-dimensional conjugate nucleus. For example, the convolution used by the 3D U-Net segmentation network is 3D. Two-dimensional convolution in U-Net is converted into three-dimensional convolution, which is suitable for segmenting three-dimensional medical images.

Medical Image Segmentation Based on Deep Learning

When image segmentation operations are performed, the convoluted neural network has better feature extraction and feature expression functions. Manual extraction of image features or excessive image pre-processing is not necessary. Therefore, in

recent years, CNN has been used to split medical images. He has had great success in field and auxiliary diagnosis. This section summarizes the results of current classical research and divides existing methods of medical image segmentation into three categories based on medical education: FCN, U-Net and 3D-U-Net. Each class is presented separately. The advantages and disadvantages of each method are compared.

Fully Convolutional Neural (FCN) Network is the leading network of the most advanced and modern deep learning technology in the field of semantic segmentation. This section introduces the advantages and disadvantages of the FCN networks. Various forms of FCN and their applications have been presented.

For a general classification, CNN networks, such as VGG and ResNet, include fully connected layers at the end of the network. Category probabilistic information can be retrieved after the SoftMax layer, but probabilistic information is indirect. In other words, only the category of the whole image can be identified, but not the category of individual pixels. Therefore, this full join method is not suitable for image segmentation. Long et al. [3] proposed a fully integrated network for the above issues. In a typical CNN structure, the first five layers are composite layers. The sixth and seventh layers are fully connected layers with a length 4096 (one-way vector). The eighth layer is a fully connected layer with a length of 1000, which is likely to be 1000 categories. FCN converts the 3 layers from the 5th to the 7th layer into a convolutional layer, and its convolutional dimensions are 7 × 7, 1 × 1 and 1 × 1, to obtain a two-dimensional feature map of each pixel. Then through a softmax layer to get the classification information of each pixel. Segmentation problem solved. A fully integrated network can accept input images of any size. FCN uses the resolution layer to sample the feature map of the final convolutional layer and restore it to the same size as the input image. In this way, a prediction can be made for each of these pixels while preserving the spatial information in the original input image. Finally, pixels are arranged by pixel to complete the map of over-sampled objects According to the oversampling magnification, it is divided into FCN-32, FCN-16 and FCN-8.

The structure of the FCN network is shown in fig.1

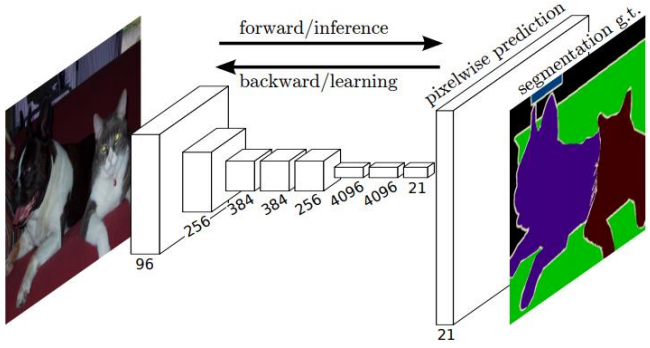

Figure 1. the Structure of the Fully convolutional networks

However, the drawbacks of FCN are also very important. First, the sampling results are relatively blurry and insensitive to the details of the image, resulting in poor segmentation. Second, the idea of segmentation is to classify each pixel without complete consideration. The pixel-to-pixel relationship is lack of spatially consistency.

2D U-Net In image segmentation tasks, especially medical image segmentation, U-Net is undoubtedly one of the most successful methods, which is our first chosen research paper: U-Net: Convolutional Networks for Biomedical Image Segmentation. In this article, the authors proposed a network and training strategy that relies on the powerful use of data augmentation to make more effective use of available annotated samples. This method was proposed at the MICCAI conference in 2015 and has now reached more than 25,000 references. The encoder (down-sampling)-decoder (up-sampling) structure and skip connection adopted are a very classic design method.

The structure of U-Net is shown in the fig.2. The left side can be regarded as an encoder, and the right side can be regarded as a decoder. The encoder has four sub-modules. Each sub-module contains two convolutional layers. After each sub-module, there is a down-sampling layer implemented by max pool. The resolution of the input image is 572x572, and the resolutions of modules 1-5 are 572x572, 284x284, 140x140, 68x68 and 32x32 respectively. Since the convolution uses the valid mode, the resolution of the next sub-module here is equal to (resolution of the previous sub-module -4)/2. The decoder contains four sub-modules, and the resolution is sequentially increased by up sampling until it is consistent with the resolution of the input image (because the convolution uses the valid mode, the actual output is smaller than the input image). The network also uses a skip connection to connect the up-sampling result with the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder.

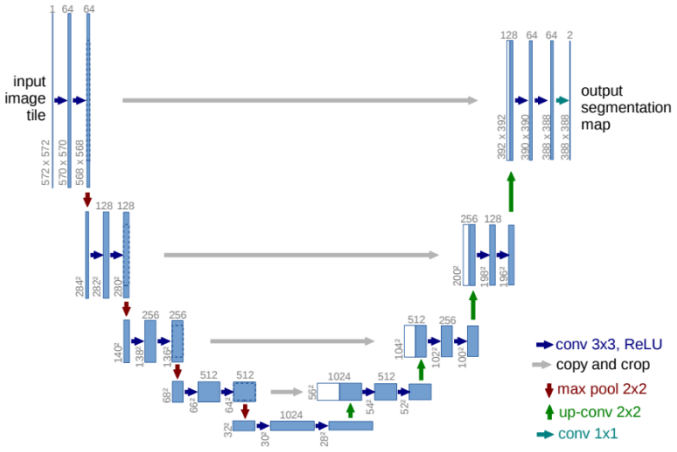

Figure 2. U-net architecture

U-Net is suitable for segmenting medical images because its structure can combine low-level and high-level information at the same time. Basic information improves accuracy. Advanced information helps to extract complex functionality.

3D U-Net: The improvement of U-Net has become a hotbed of medical image segmentation research. Many variations have been developed on this basis. Çiçek et al. [1] proposed a 3D model of U-Net. This model is intended to ensure that the U-Net structure has richer spatial information. 3D CNN can extract more robust volumetric representations in all three axes X, Y and z axes. Using 3D information in segmentation fully exploits

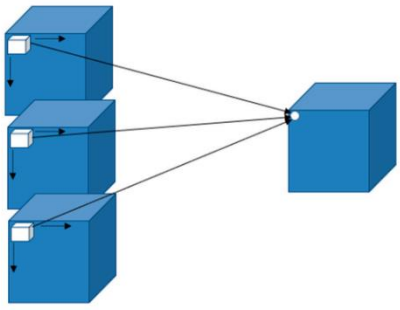
Figure 3. 3D CNN convolution

the advantages of spatial information. The 3D convolution has more depth than the 2D convolution, that is, 2D slices of medical images. For a 3D image, C × N × H × W, where C, N, H, W represent the number of channels, the number of slice layers, and the height and width of the convolutional kernel. As in the case of 2D convolution, the resulting value equals the height and width of the sliding window and the number of layers per channel. CNN's 3D convolution process is shown in fig.3.

Compared with U-Net, this network only uses three down sampling operations, and batch normalization is used after each convolutional layer, but both 3D U-Net and U-Net do not use dropout. Its network structure is shown in fig.4. Through a shortcut, a single resolution layer in the encoding path is passed to the decryption path so that it has the original high-resolution properties. The matrix maintains the segmentation of the 3D image by importing a continuous 2D slice sequence into the 3D image. Not only can the network train and tick other locations that are not labeled in the weakly labeled data sets, but also the ability to train and tick new data in many weakly labeled data sets. Compared with the U-Net input, the input is 132 × 132 × 116 stereoscopic images with three channels. The output image size is 44 × 44 × 28 image size. 3D U-Net retains the best basic features of FCN and U-Net. Its appearance is a great help for stereo images.

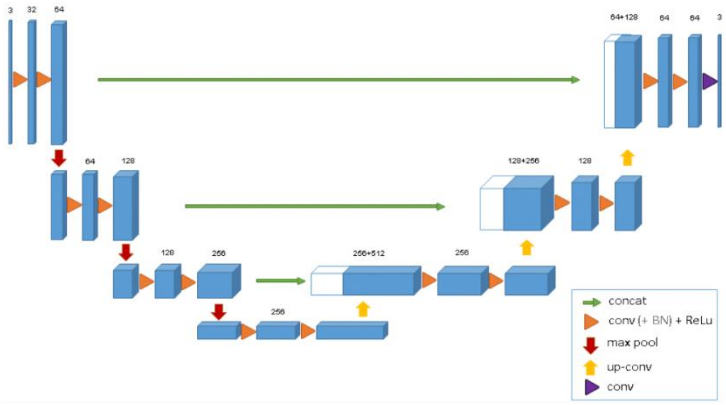

Figure 4. The structure of the 3D U-Net

Segmentation Evaluation Metrics

A precise objective indicator is needed to evaluate the quality of an algorithm. In medical segmentation algorithms, hand-drawn annotations by physicians are often used as the gold standard (basic truth, GT for short). Other consequences of segmentation the algorithm are the prediction results (abbreviated as Rseg, SEG). Segmental evaluation of medical images is divided into two methods: pixel-based and overlay-based.

Dice index: The dice coefficient is a function for uniformity assessment. It is often used to calculate the similarities or overlaps between two samples. It is also the most commonly used. Values ranges from 0 to 1. The closer the value is to 1, the better the segmentation effect. Given two sets A and B, the index is defined as:

$$Dice(A,B) = 2\frac{|A \cap B|}{|A| + |B|}$$

Jaccard index: The Jaccard index is similar to the coefficient of the dice. Given two sets A and B, the index is defined as:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Segmentation Accuracy (SA): Percentage of properly segmented area compared to actual GT area. Of these, Rs is the reference region of the segmented image manually drawn by the expert. T is the actual field of the image resulting from the segmentation of the algorithm. | Rs - Ts |: Number of poorly segmented pixels

$$SA = \left(1 - \frac{|R_S - T_S|}{R_S}\right) \times 100\%$$

To evaluate U-net, a task for cell segmentation in light microscopic images. This segmentation task is part of the challenges of tracking ISBI units in 2014 and 2015. The first data set "PhC-U373" contains glioblastoma-astrocytoma U373 cells on a polyacrylamide substrate, recorded by a phase contrast microscope. It has 35 partially labeled training images. Here, the intersection union (IoU) achieved by Ronneberger et al. [2] averages 92%, which is much better than 83% of the sub-optimal algorithm. The second data set "DIC-HeLa" is HeLa cells recorded on flat glass by a differential interference contrast (DIC) microscope. It has 20 partially labeled training images. Here we have achieved an average number of IOUs of 77.5%, which is significantly better than the 46% of the sub-optimal algorithm.

To evaluate the quantitative performance in the semi-automatic setting of 3D U-net, Çiçek et al. [1] evenly divided all 77 manually annotated slices in all 3 samples into 3 subsets and performed batch standardization and non-batch. In the case of standardization, 3-fold cross-validation is performed. To do this, they removed the test pieces and kept them unmarked. This simulates an application where users provide more sparse annotations. To measure the effect of using a full 3D environment, they compared the results with a pure 2D implementation, which treats all marked slices as separate images. IoU is used as an accuracy measurement to compare the real slices of the ground that are dropped with the predicted 3D volume. IOU is defined as true positive/(true positive + false negative + false positive). The results show that their method has been able to induce very accurate 3D segmentation from a few annotation slices with very little annotation effort.

Conclusions and Future Directions

While research on segmentation of medical images has made great progress, the segmentation effect has not yet been able to meet the needs of practical applications. The main reason is that the current study of medical image segmentation still presents the following difficulties and challenges:

1. Medical image segmentation is the intersection between these two disciplines. Clinical pathology is complex and varied. However, artificial intelligence scientists do not understand the clinical need. Physicians do not understand the specific technique of artificial intelligence. Therefore, artificial intelligence may not respond well to specific clinical needs. To promote the application of artificial intelligence in the

medical field, it is necessary to strengthen the broad collaboration between practitioners and scientists in the field of machine learning. This collaboration will solve the problem that researchers in the field of machine learning cannot obtain medical data. This can help machine learning researchers develop better deep learning algorithms for clinical needs and apply them to computer aided diagnostic devices, thereby improving accuracy. And diagnostic efficiency.

2. Medical images are different from natural images. Different medical images are different. This difference also affects the adaptability of the learning and learning model during segmentation. Noise and artifacts in medical images are also a major issue when pre-processing data.

3. Limitations of existing medical imaging datasets. The current medical imaging data set is relatively small. Training of deep learning algorithms requires a large amount of supporting data, which leads to the problem of overfitting when training a learning model. Improving data such as changes in geometry and an increase in color space is one way to overcome the shortage of training data.

4. Deep learning models have their own drawbacks. The research focuses on three aspects: the design of the network structure, the design of the 3D data segmentation model, and the design of the loss function. The design of the network topology is worth investigating. The change in the network structure has significant effects and can easily be transferred to other tasks. Three-dimensional medical data can more accurately capture the target's geometrical information, and this information can be lost if the three-dimensional data is sliced. Therefore, designing a three-dimensional convolution model for processing three-dimensional medical image data is a reproducible direction. Loss function design has always been a difficult point in deep learning research.

To distinguish medical images, deep learning has performed excellent. More and more new methods are being used to improve the accuracy and robustness of segmentation. Diagnose various diseases with AI realizes the concept of sustainable medical treatment. It has become a powerful tool for clinicians. But this problem still needs to be solved and we can expect many innovations and research results in the coming years.

Reference

[1] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv:1606.06650 [cs]*. [online] Available at: https://arxiv.org/abs/1606.06650 [Accessed 4 May 2021].

[2] Ronneberger, O., Fischer, P. and Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. [online] arXiv.org. Available at: https://arxiv.org/abs/1505.04597.

[3] Long, J., Shelhamer, E. and Darrell, T. (n.d.). *Fully Convolutional Networks for Semantic Segmentation*. [online] . Available at: https://openaccess.thecvf.com/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf.