Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/17/20240702

Application of deep learning in medical imaging segmentation

Keke Huang

HD school, Shanghai, China

rallen82707@student.napavalley.edu

Abstract. The increasing demand for segmentation of lesions in medical images necessitates research on automatic segmentation. Manual segmentation is inefficient due to training time and energy constraints. Deep learning-based image segmentation technology can improve efficiency and aid in diagnosing conditions. This technology provides accurate and detailed data support for clinical medicine, making it a crucial field in medical image processing. This essay introduces image segmentation and its classification, and explains the differences between two segmentation methods, semantic segmentation and instance segmentation and their respective application fields. Additionally, it introduces several well-known deep neural networks for segmenting medical images using deep learning. Regarding the models, this article introduces the structure and characteristics of each model as well as their respective advantages and disadvantages. This essay also introduces examples of deep learning using different deep neural network models to segment specific medical images, including image segmentation based on FCN for the heart and U-Net for the kidneys.

Keywords: deep learning, medical imaging segmentation, deep neural networks

1. Introduction

When clinicians or researchers use medical imaging to do quantitative analysis, real-time monitoring, and treatment planning on a particular internal tissue or organ, they often need to have thorough knowledge about this tissue and organ in order to make the best treatment decisions. As a result, biomedical imaging is playing an ever-more-important role in the detection and treatment of illness. The use of medical image analysis has greatly helped screening, diagnosis, classification, decision-making and treatment consultation, evaluating the effectiveness of benign tumors and malignant diseases, brain function and mental disorders, cardiovascular and cerebrovascular diseases as well as other serious diseases. The first stage of medical image analysis is medical image segmentation technology, which is based on digital image processing. Image segmentation is the process of dividing a picture into a number of random segments based on elements such spatial texture, colour, grayscale, and geometric forms, in order to make these characteristics seem comparable or consistent both inside and across areas. Make a clear distinction. In the area of medical image processing, it can promote clinical diagnosis and therapy by assisting clinicians in precisely and rapidly detecting lesions. The two primary categories of medical image segmentation technologies are deep learning algorithms and conventional algorithms.

Edge detection-based image segmentation techniques, region-based image segmentation techniques, threshold-based image segmentation techniques, and others are examples of traditional image segmentation techniques. Traditional algorithms frequently struggle to solve complicated medical picture segmentation tasks satisfactorily because of their restricted feature extraction and processing capabilities. The threshold segmentation approach's core idea is to provide a number of feature thresholds and categorise the image pixels into target regions and background areas with different shades of grey [1]. The threshold segmentation approach has the benefits of quick image segmentation, easy calculation, and high effectiveness. However, because this method just evaluates the properties of the pixel grey value and does not consider spatial factors, it is susceptible to noise [1]. The region-based segmentation approach is a segmentation technology that finds regions directly. Its key drawbacks include sluggish segmentation speed and noise sensitivity [1]. Edge-based segmentation techniques address the segmentation issue by identifying edges connecting various regions [1]. Usually, the changes in pixel gray value on the edges between different areas are often drastic, which is one of the main assumptions for implementing edge detection methods. To judge edge points, edge detection systems often use the greatest value of the image's first derivative or the zero-crossing point information of the second derivative [1]. However, this technology has significant limits because many medical images cannot ensure edge closure and continuity. Due to their superior feature extraction and processing skills, deep learning algorithms have become widely used in the segmentation of medical images.

2. Imaging segmentation

To make image analysis simpler, imaging segmentation involves segmenting a visible input [2]. A fragment is an object or a piece of an object that includes groups of super pixels. It is no longer required to consider individual pixels as observational units thanks to picture segmentation. Imaging segmentation is splitting an input image into discrete segments strongly associated to regions of interest (Rols) in a given image [3]. In order to provide a solid foundation for clinical diagnosis, treatment, and pathological research as well as to help doctors make more accurate diagnoses, such as measuring tissue volume to measure tumour size and assist with drug dosage, medical image segmentation separates parts of medical images that have specific special meanings [3]. Semantic segmentation and instance segmentation are two broad categories of image segmentation techniques.

2.1. Semantic segmentation

By using semantic segmentation, this technique may give each pixel of a digital image a class name, such as "tree," "sign," "pedestrian," "road," "structure," etc [4]. It is also considered to be a pixel-level image classification problem since it requires the ability to distinguish between various items in an image [4]. Before employing features to create various categories in an image, semantic segmentation seeks to extract features from the picture. Three phases can be used to broadly split semantic segmentation. To categorise certain items in the image, the training data must first be examined. After that, objects are found, and bounding boxes are generated around them using a semantic segmentation network. To train a semantic segmentation network and group the pixels in the image locally, a segmentation mask is created. CT scans, X-rays, geosensing, and other imaging techniques can all involve semantic segmentation.

2.2. Instance segmentation

The goal of instance segmentation, a form of image segmentation, is to identify and describe each unique instance of an item in an image [4]. Individual instances of any segmented class can be segmented via instance segmentation, which can detect all instances of a class [4]. Therefore, it is also known as combining the capabilities of object detection and semantic segmentation. There are significant 2 steps of instance segmentation. Firstly, it performs object detection to determine each object's bounding box. Second, inside each enclosing box, a semantic segmentation model is used. All instances within each class are just divided by instance segmentation, which divides each person into a distinct class. Instance segmentation can be used in medical domain, robotics, etc.

DOI: 10.54254/2753-8818/17/20240702

3. Deep learning

Deep learning (DL) is a research trend that has evolved as machine learning and artificial intelligence have increased in popularity [5]. It employs a deep neural networks model to extract features from various data types such as sounds, text, and images for the learning process. A neural network comprises many neurons and can think of each neuron as a little information processing unit [5]. The entire deep neural network is built by connecting the neurons in a specific fashion.

Many forms of deep neural networks, including convolutional neural networks, fully convolutional networks, U-shaped networks and recurrent networks are effective at handling medical image segmentation. This section provides an overview of several deep learning neural networks used for picture segmentation.

3.1. Deep neural networks

Deep learning techniques construct artificial neural networks using individual layers. In the network between the input and output layers, there are hidden layers that perform computations. The network's output layer decides how to use the signals the input layer has received [3]. Numerous hidden layers exist between deep neural networks' input and output layers [3]. The general neural network architecture is shown in figure1.

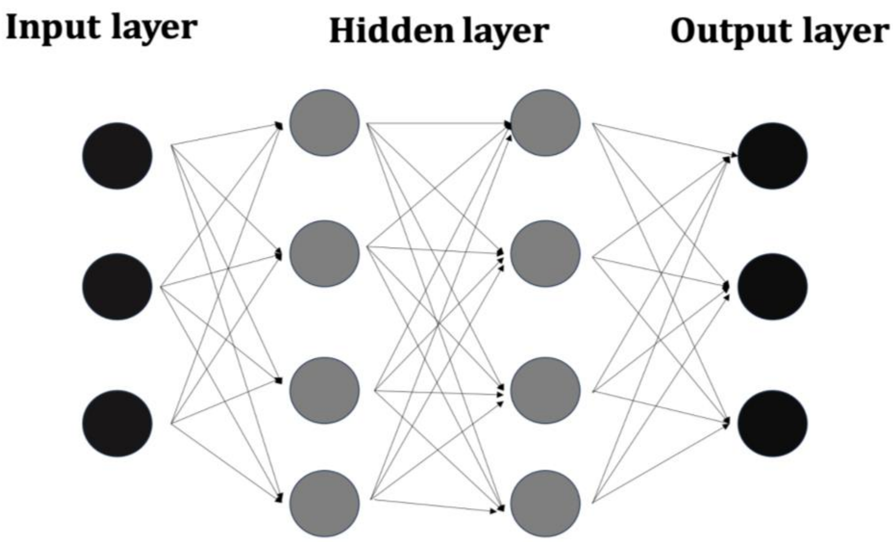

Figure 1. The structure of the general neural network (original).

3.1.1. Convolutional neural work. One of the most important networks in deep learning is the convolutional neural network (CNN) [6]. It has grown to be one of the deep learning field's most representative neural networks. Convolution structures are used by the CNN, a kind of feedforward neural network, to extract features from data. [6]. Unlike traditional feature extraction approaches, CNN does not need manual feature extraction [6]. CNN's architecture is motivated by visual perception.

CNN excels at image processing. There are two main features of CNN. One is efficiently reducing massive data volumes' dimensionality into smaller data volumes. The other is that it complies with image processing fundamentals while still efficiently retaining image attributes. Pixels, the building blocks of images, are made up of colours. To represent colour information, each pixel contains RGB 3 parameters. As a result, processing a photo requires processing millions of parameters. Processing such a vast volume of data takes a lot of resources. The first issue that CNN addresses is to "simplify complex problems" and turn a high number of parameters' dimensionality into a small number of parameters. Furthermore, it has no impact on the recognition outcomes. Additionally, traditional approaches will yield drastically different characteristics when the image's objects move, but CNN will not. When an image is flipped, rotated, or moved, it still keeps its qualities in a way that is similar to seeing the image as it was originally [6]. It is also good at recognizing similar images.

The three primary elements of CNN are the convolution layer, the pooling layer, and the fully connected layer [3]. Extracting regional characteristics from the picture is the responsibility of the convolutional layer. The convolutional layer performs the mathematical operation of multiplying the local neighbours of the image pixels with the kernel. CNN generates its feature map by convolving the provided image using several convolution kernels. The pooling layer greatly lowers the parameter magnitude (dimensionality reduction) but does not affect the data depth. Comparable to the traditional neural network component is the fully connected layer, which is used to output the desired results.

Building a CNN model is the key to medical image segmentation. Carry out model training on the constructed model, prepare a set of labelled image data as a training set, and specify the loss function and optimization algorithm. The trained CNN model can perform image segmentation. The trained CNN model receives a fresh medical image. Based on previous data sets, the model segments the picture and the lesion's position.

3.1.2. Fully convolutional networks. Fully convolutional networks (FCN) use convolutional layers rather than completely connected ones, allowing for the input of any image size [3]. It is the forerunner for end-to-end semantic segmentation. Instead of making patch-wise predictions from the full-size input image, the model may produce a spatial segmentation map and make dense predictions at the pixel level. The model uses skip connections, which combine feature maps from older layers with up-sampled feature maps from the most recent layer [7]. Thus, the model generates a thorough segmentation in a single pass. There is a model which is an advanced FCN is called ParseNet, it uses the global average pooling to attain global context [3]. The FCN network structure is seen in Figure 2.

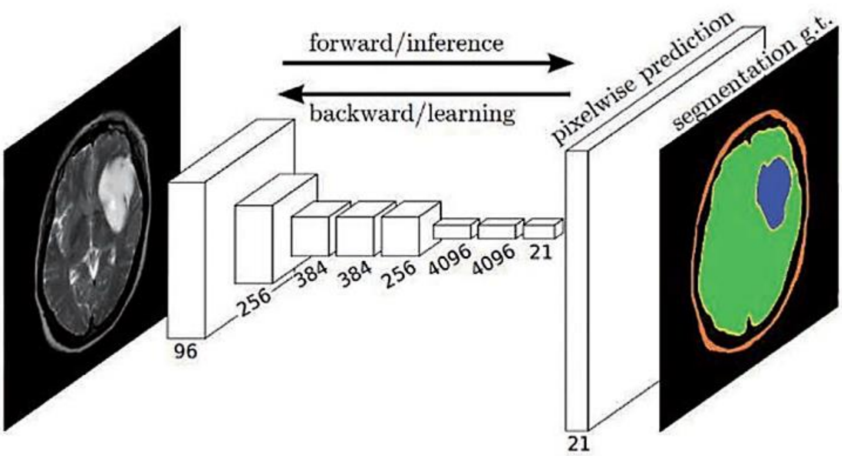

Figure 2. The network structure of FCN [8].

3.1.3. U-shaped networks. Medical imagine semantic segmentation using U-Net. U-Net is an end-to-end neural network architecture. The network initially executes four sets of convolutions and downsampling operations on the input image in order to extract information about the image's features. The network then does four sets of deconvolutions and upsampling operations to expand the image [9]. U-Net offers skip connections between downsampling and upsampling operations with the same number of channels to aid the decoder in fixing target data more successfully [9]. Due to the U-Net network's superior structural performance, it is frequently employed in segmenting medical images. It produces positive results in the clinical auxiliary diagnosis of serious diseases like breast cancer, lung cancer, liver tumours, and brain tumours.

The U-Net approach has a lot of benefits. First, it can efficiently segment a limited set of labelled training images. Second, the U-Net architecture combines the context data obtained from the upsampling path with the position information of the path obtained from the downsampling path to predict an appropriate segmentation map [3]. Only 572x572 pictures can be detected by this model, though. The depiction of segmentation using the U-Net model is shown in Figure 3.

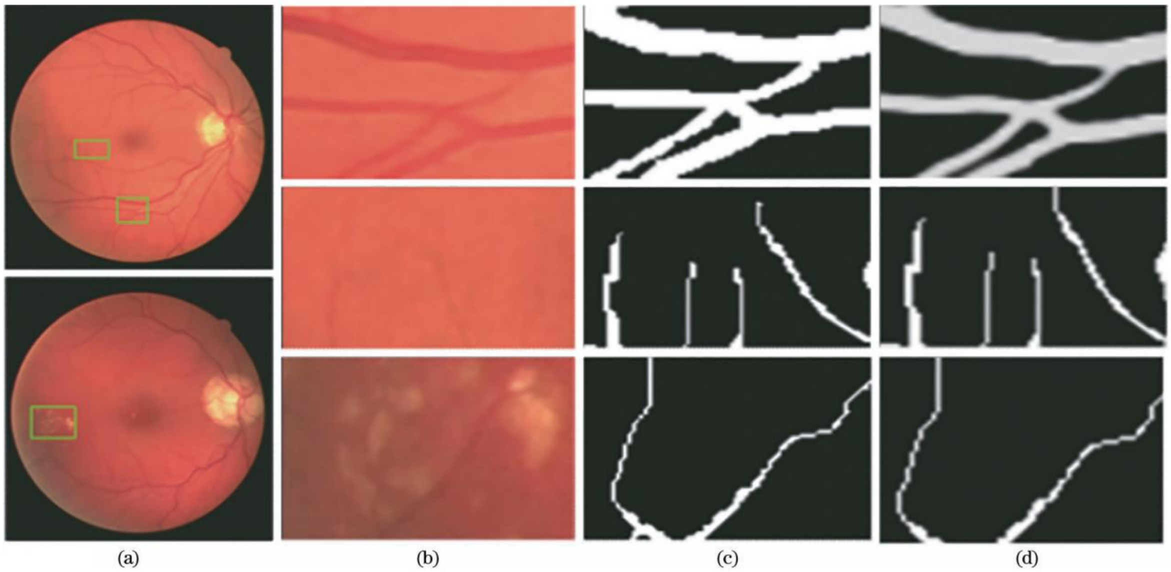

Figure 3. Partial blood vessel region segmentation diagram based on U-Net model [10]. (a) original fundus retinal pictures in colour; (b) images of local fundus retina; (c) local standard retinal segmentation images; (d) local retinal segmentation result images.

3.1.4. Recurrent neural networks. Recurrent neural networks (RNN) are compact, universally parameterized models that provide a range of conditional distributions. RNNs are good at solving challenging sequence problems, including character prediction, handwriting generation, and machine translation. Traditional neural networks cannot process inputs in a relationship between the previous and subsequent inputs; rather, they can only process inputs independently. However, some tasks call for better processing of sequence information, which refers to how one input relates to another. So, there is RNN, which uses hidden layers to solve this problem. The hidden state of RNN is its main and most important part. It will remember some information about the sequence and remember previous inputs. To perform the same task on all inputs or hidden layers and create output, RNN uses the same parameters for each input. Compared to other neural networks, this reduces parameter complexity. In RNN-based image semantic segmentation, the RNN layer is embedded into CNN, the local spatial characteristics of the picture are extracted in the convolution layer, and the pixel sequence features are extracted in the RNN layer. CNN first processes the input image to create a feature map; this feature map is then input into an RNN to create an image context map; this layer is then used to serialise the pixels and create global semantic features; finally, the deconvolution layer is used to process the upsampling; and then produces the segmentation results.

3.2. Comparison between different image segmentation algorithms

The various deep neural networks discussed in these sections have many applications. Each model has unique advantages and disadvantages. Table 1 compares different deep learning architectures.

Table 1. Comparison of different architectures of deep learning.

Type of network	Pros	Cons
Convolutional neural network (CNN)	It is simple. Model learns rapidly.	Fix size of output It requires number of labelled images for classification
Fully convolutional networks (FCN)	Output is a spatial segmentation map	Hard to train for a good performance
U-shaped networks (U-net)	Just a few tagged training photos are used	Input image size is limited to 572 × 572.
Recurrent neural networks (RNN)	It can learn sequential events	Need for big datasets

4. Application of deep learning in medical imaging segmentation

Deep learning collects data, analyses and processes the data, and then further divides the data into three data sets: training, verification, and testing. After selecting a deep learning model, you should analyse the entire system, test it, and assess how well it performed. Medical image segmentation aims to identify regions of interest (ROIs) such as lesions and tumours. Deep learning processes medical images to identify and screen organs and lesions and confirm the location and size of lesions. Due to the varying needs for various diseases and body parts, when using deep learning to medical image segmentation, the choice of deep neural network models is also variable.

Deep learning technology provides an effective way to help analyse the structure and function of blood vessels by segmenting specific tissues in different ways. The left and right ventricles' endocardium and epicardium are divided as a result of ventricular segmentation [3]. These segmentations are essential for obtaining clinical signals. The left, right, and myocardium may be distinguished right away on short-axis cardiac magnetic resonance (MR) images based on FCN. The end-to-end technique based on FCN produces efficient segmentation, which is significantly faster and more accurate than older methods. Most cardiac MR images have low resolution and motion objects, hence 2D FCN is often employed for segmentation rather than 3D FCN.

Segmenting the liver and liver cancers using three-dimensional volumetric images is one of the current study fields in medical image processing. The tasks of liver segmentation and tumour detection are challenging, in contrast to other organs like the heart, spleen, stomach, and kidneys, which have equivalent qualities in terms of form, texture, and intensity values. The U-net model is frequently used to automatically segment liver lesions, including intrahepatic cholangiocarcinoma, liver abscess, liver tuberculosis, etc. To accurately segment the liver area, Sun Mingjian and others proposed a fully convolutional network 3D U NetC [11]. This network is further improved based on U-net and uses a deeper network to extract deeper features. And convolution is performed on it to combine the feature maps more efficiently, which improves the accuracy to a certain extent [11]. The segmentation of liver tumour areas is mainly based on three-dimensional segmentation because this can fully display the spatial information of the image. However, three-dimensional segmentation and deep convolution will bring many parameters and increase the computational cost of the network.

There are many other uses of deep learning in medical image segmentation, including using 2D-CNN to segment brain tissue from multimodal MR images, using 3D-FCN to segment lung nodules from CT images, and using the U-Net model to segment eyes. retinal blood vessels in. Other applications will not be listed and discussed in this essay.

5. Conclusion

Since traditional segmentation has many limitations, research on automatic segmentation has become the latest research field. Automatic medical image segmentation based on deep learning is already the latest research field. This article introduces image segmentation and deep learning technology. For image segmentation, this article explains two segmentation methods, semantic segmentation and instance segmentation, their respective principles, and their application places and differences. For deep learning, this article also introduces several popular deep neural network models in image segmentation, focusing on the principles of convolutional neural networks, fully convolutional networks, U-shaped networks and recurrent networks, and compares the performance of these models. pros and cons. CNN is a significant network in deep learning, using convolution structures to extract features from data. It excels in image processing, reducing dimensionality and retaining image attributes. CNN's architecture focuses on visual perception and retaining image attributes. Its main components are the convolution, pooling, and fully linked layers, which extract local features and generate feature maps. FCN use convolutional layers for end-to-end semantic segmentation, generating spatial segmentation maps and dense pixel-wise predictions. U-Net is a neural network architecture used in medical image segmentation, combining convolution and downsampling to extract and expand feature information. It efficiently segments small training images and predicts fair segmentation graphs. RNNs are compact, universally parameterized models for solving sequence problems like character prediction and handwriting generation. To improve processing and minimise parameter complexity, they employ hidden layers. This article introduces the particular use of deep learning technology in medical picture segmentation in the last section. Different deep medical segmentation models are utilised in various contexts. It focuses on the segmentation technique for cardiac pictures based on the FCN model. Segmentation, for the segmentation of liver and liver lesions based on the U-net model and mentioned the application of other models to segment lesions in other organs. Deep learning-based medical image segmentation algorithms continue to encounter several difficulties, including 3D Challenges. It is hoped that these problems can be improved in the future.

References

[1] Anjna E, Kaur E R. 2017, Review of image segmentation technique[J]. *International Journal of Advanced Research in Computer Science*, 8(4): 36-39.

[2] S. Prince Mary et al 2020 J. *Phys.: Conf. Ser.* 1712 012016

[3] Malhotra, P. et al. (2022) *Deep Neural Networks for medical image segmentation, Journal of Healthcare Engineering.* Available at: https://www.hindawi.com/journals/jhe/2022/9580991/ (Accessed: 09 October 2023).

[4] "Mrinal Walia." Oct 13, 2022, Roboflow Blog. https://blog.roboflow.com/difference-semantic-segmentation-instance-segmentation/

[5] Aljabri, M. and AlGhamdi, M. (2022) 'A review on the use of Deep Learning for medical images segmentation', *Neurocomputing*, 506, pp. 311–335.

[6] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, Dec. 2022,"A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE *Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999-7019.

[7] Shelhamer, Evan, Jonathan Long, and Trevor Darrell, 2017, "Fully convolutional networks for semantic segmentation." IEEE *Trans. Pattern Anal. Mach*. Intell. 39.4: 640-651.

[8] Long J, Shelhamer E, Darrell T. June 7-12, 2015, Fully convolutional networks for semantic segmentation [C] ,2015IEEE *Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA. New York : IEEE Press , 2015 : 3431-3440.

[9] Zhou Tao, Hou Senbao, Lu Huiling, Zhao Yanan, Dang Pei, Dong Yali. 2022, *J Biomedical Eng*, 39(4): 806-825.

[10] Li D X, Zhang Z. 2020, Improved U-net segmentation algorithm for the retinal blood vessel images[J]. Acta Optica Sinica, 40(10): 1010001.

[11] Huan Zhang, Dawei Qiu, Yibo Feng, Jing Liu. 2022, Improved U-Net Models and Its Applications in Medical Image Segmentation: A Review[J]. *Laser & Optoelectronics Progress*, 59(2): 0200005.