Deep Learning-based Semantic Segmentation Methods in Medical Imaging

Jiaan Li [1, †, *], Peicheng Li [2, †], Houyuchen Li [3, †] and Tianheng Ying [4, †]

[1] Department of Mechanical Engineering, Shenyang Institute of Engineering, Shenyang, China

[2] Knowledge-First Empowerment Academy, USA

[3] School of Computer Science and Technology, Sichuan University, Sichuan, China

[4] School of Computer Science and Technology, Nanjing Tech University, Nanjing, China

* Corresponding author email: 631301120221@mails.cqjtu.edu.cn

[†]These authors contributed equally.

Abstract. Image semantic segmentation based on deep learning attracted much attention in the field of computer vision in recent years. In the past few decades, the tremendous developments of medical imaging technology have made it possess an increasingly important role in diagnosis. Therefore, the image semantic segmentation method based on the deep learning algorithm is widely considered by many researchers in the medical image processing field. In this paper, first of all, we introduced some semantic segmentation methods based on deep learning, including point cloud data-based methods, Recurrent Neural Network (RNN) based methods, attention mechanism-based methods, transformer-based methods, and Generative Adversarial Network (GAN) based methods. Then, the characteristics of each method and the problems improved by each method in the field of medical imaging are shown respectively, for example, point cloud data can better express the information of biological tissue structure to improve the capacity of processing global contextual dependencies, attention mechanism assists the model quickly discover key information of input images, transformer improve the effect of segmentation tasks, and GAN has superior data generation capacity that they have excellent performance in generating realistic-looking images. And we summarize the performance of each method in the medical dataset by distinct metrics. Finally, some problems of theory, architecture and application in future research were discussed.

Keywords: Deep Learning; Semantic Segmentation; Medical Imaging.

1. Introduction

In the past few decades, the tremendous developments of medical imaging technology have made it possess an increasingly important role in diagnosis. Through the application of medical imaging technologies, medical professionals have amassed a sizable amount of patient image data, and the continuously improved image resolution has laid a solid foundation for the study of brain tumors, lung diseases, cardiovascular diseases. Due to the huge amount of medical image data, manual screening is time-consuming and laborious, which cannot meet the development requirements of modern medical imaging. Therefore, accurate medical image segmentation method has attracted much attention from radiologist and researchers in recent years.

Traditional medical image segmentation methods include threshold method and graph division method which are mainly to segment the image data according to the low order information of the gray level, shape and other characteristics of the pixel contained in the image data [1, 2]. When encountering complex situations, e.g., the scene where the gray scale and shape of the image are changeable, the segmentation effect may encounter failure. Therefore, the image semantic segmentation method based on deep learning algorithm is widely considered by many researchers in the medical image processing field. Different from traditional segmentation methods, the image segmentation based on deep learning is a method of grouping according to the different semantics expressed by the pixels in the image, which divides and extracts the features in the medical image. With the introduction of Full Convolution Neural Network (FCN) in 2015, mainstream deep learning

algorithms marked by Convolutional Neural Network (CNN) [3], Recurrent Neural Network (RNN) [4], and Generative Adversarial Network (GAN) [5] began to substantially replace traditional algorithms based on threshold method, edge method. The FCN network is based on the CNN network, which converts the full connection layer at the end of the CNN into a convolution layer. At the same time, the FCN network can accept any size input of images, and can classify images at the pixel level [6], laying the foundation for the subsequent model development in the field of image semantic segmentation.

Currently, there are many kinds of image semantic segmentation methods based on deep learning algorithm, including point cloud data-based methods [7], RNN based methods, attention mechanism-based methods [8], transformer-based methods [9], and GAN based methods. Point cloud data can better express the information of biological tissue structure, and provide strong support for medically related lesion area information segmentation and accurate matching, medical diagnosis, telemedicine and other tasks. RNN uses the feature of circular connection to store the history input in the internal state of the network, which could be later combined with the current input to update the current state. So RNN have been used to further improve semantic segmentation with the capacity of processing global contextual dependencies. In order to better solve the problem of parallel computing capability defects of RNN network, attention mechanism and Transformer models based on self-attention mechanism have been proposed successively. Attention mechanism assist the model quickly discover key information of input images. Transformer requires a lot of encoder stacks to expand the receptive field and get high-level features, which maintains the global semantic dependency throughout, improving the effect of segmentation tasks. GAN are proposed by Goodfellow, and then GAN showed great potentialities and became one of the most popular deep network architectures in recent years. GAN is composed of generating network and discriminant network. The discriminant network and the generating network are in conflict with one another because the discriminant network determines whether the input data is real or generated, while the generating network is in charge of producing simulated data. In order to comprehensively investigate deep learning-based medical image segmentation methods, this paper reviews recent progress of some mainstream algorithms in this field.

The remainder of the paper is structured as follows. In Section 2 we introduced five semantic segmentation methods based on deep learning algorithms which stress the problems in medical imaging area. Section 3 summarize the performance of each method in medical dataset by distinct metrics. Section 4 concludes this paper and discusses the prospective challenges of algorithms mentioned in this paper.

2. Method

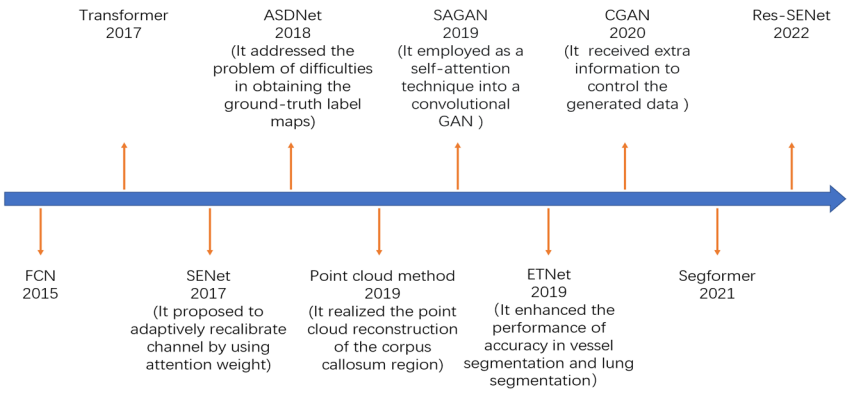


Fig 1. Brief summary of key deep learning-based methods mentioned in this paper

Figure 1 shows the summary of dominant deep learning algorithms which later are used in medical imaging analysis. The methods include cloud point-based algorithms, recurrent neural network, attention mechanism, transformer and generative adversarial networks.

2.1 Cloud Point-based Algorithm

In 1992, Besl and McKay pioneered the integration of the still popular Iterative closest point (ICP) registration algorithm into 3D data for registration [10]. Among them, the ICP algorithm calculates the Euclidean distance of each point in two-point clouds to minimize the mean square error to achieve registration, but this approach has the drawback of taking a long time to calculate and being susceptible to local optimal solutions. Later, a large number of scholars optimized the ICP algorithm for this shortcoming. For example, the Deep Closet Point (DCP) network model proposed by Wang and Solomon in 2019 improves the above shortcomings of ICP [11], fuses DGCNN and the features extracted by an attention module, and significantly improves the performance of rigid registration algorithms. Nevertheless, this method is currently only suitable for small-scale object-level point clouds.

In 2018, Schaffert et al. based on the PointNet model, according to the point to plane correspondence registration standard [12], from global to local correspondence, the single vertebrae 3D CT volume model was registered with the 2D X-ray image, which significantly affected the registration of 2D X-ray images in minimally invasive medical procedures. Subsequently, Hansen et al. conceived an end-to-end registration method for variable point set in 2019 [13], which can learn the geometric features of irregular points through DGCNN and regularized and fully distinguishable high-dimensional coherent point drift models, and register the set of close points extracted by lung CT in large in-and exhaled states. This processing idea provides a scheme for optimizing the high-dimensional coherent point drift model when processing point clouds with relatively unrelated information.

In 2019, the point cloud reconstruction method based on Gaussian mixture model was proposed by Yang and Chakraborty [14], which realized the point cloud reconstruction of the corpus callosum region of the brain MRI (magnetic resonance imaging) image, and can compare and analyze the difference of point cloud data in the corpus callosum region of the brain of patients with dementia and normal people, which is of great significance in the clinical research of dementia.

2.2 Recurrent Neural Network

Based on the improved method by RNN image classification model ReNet [15], the literature has proposed the ReSeg network [16]. First, the network feeds the input image into a pre-trained VGG-16 network for extracting high-level representations [17], then sends the generated feature map to multiple ReNet layers. Finally, several up-sampling layers composed of deconvolution networks are employed to restore the resolutions of the feature map.

The main advantage of RNN is to process sequence information, so it has great application in language processing and its advantage in image segmentation is not obvious.

2.3 Attention Mechanism

Attention mechanism is a method which makes the network model focus on the important part of the picture while ignoring its irrelevant information. The attention score is obtained by updating the score matrix through the back propagation algorithm of the neural network. Attention mechanism can be divided into six categories e.g., channel attention and spatial attention [18]. In medical image analysis area, tremendous work involving attention mechanism has been done. In 2018, Nie et al. proposed a semi-supervised neural network which is named Attention based Semi-supervised Deep Networks (ASDNet) to address the problem of difficulties in obtaining the ground-truth label maps which requires massive labor wok and professional knowledge [19]. SENet was proposed to construct links among feature maps in different channels [20]. Zhao et al. [21] combined the feature of SENet with ResNet as Res-SENet to enhance the performance of CNN and suppress irrelevant feature information in image analysis of patient's abdominal arteries. In ET-Net [22], edge-attention representations are acquired by the application of edge-attention module, which are later transferred to decoding layers. This procedure enhances the performance of accuracy in vessel segmentation and lung segmentation bimages.

2.4 Transformer

As image classification is closely related to semantic segmentation, many segmentation models have been born based on Vision Transformer (VIT) [23, 24]. Among them, SETR Segmentation Transformer (SETR) divides an image into several patches of equal size based on ViT [25], adds location information, makes full use of the attention mechanism to form an encoder, and finally decides to the original image size. Also based on ViT, the Segmenter model introduces category embedding in the decoding part to improve the accuracy of segmentation [26]. In order to avoid the problem that the size of the training set image is different from that of the input set image, which leads to the reduction of the segmentation ability, the SegFormer model reduces the output image features layer by layer to obtain the feature information of different sizes and uses the MLP decoder (lightweight all MLP decoder) to aggregate the features of different sizes to achieve efficient segmentation [27].

Because medical images have the characteristics of fuzzy boundary, low contrast, and inconsistent target size, and the traditional convolutional neural network has a limited receptive field, which cannot capture the global features very well. Transformer model has the advantage of capturing the global dependency in the feature map and has high segmentation accuracy. A series of models based on transformer are applied to medical image segmentation.

2.5 Generative Adversarial Networks

GAN have superior data generation capacity gained wide attention by the medical yield. Because of their excellent performance in producing realistic-looking images, their capacity is used in real image processing to address the issue of the persistent lack of annotated data in the medical output. Therefore, GAN and their extensions have been used for many tasks about medical image segmentation. The work of segmenting medical images is crucial since it tries to precisely identify the precise object boundaries as organs and lesion locations.

The following summarizes several used GAN architectures in the medical image segmentation. Proposed by Majurski et al. [28], the GAN and U-net are used in this transfer learning-based method in order to improve the accuracy of cell border identification. First, the GAN was trained to learn an abstract representation of the data from raw data in an unsupervised manner. The U-net encoder served as the discriminator and the U-net decoder served as the generator for the GAN design. So, the encoder component of U-net received the discriminator optimized weight [29]. Conditional Generative Adversarial Network(cGAN)is a kind of conditional model whose generator and discriminator can receive extra information to control the generated data such as class labels proposed by Mirza et al [30]. Based on the model, based on contextual information-aware cGAN, Singh et al. reported a method for segmenting breast cancers in ultrasound images [31]. Traditional convolutional GAN captures struggle mightily to pick up on some image classes, and it only records spatial data, missing out on images with geometric patterns. Self-Attention Generative Adversarial Network (SAGAN) proposed by Zhang et al [32]., it incorporates a self-attention technique into a convolutional GAN and generates images using global, long-range dependencies.

3. Result and Discussion

3.1 Result

According to the main models based on deep learning algorithms from four fields (i.e., Attention mechanism, point cloud, Transformer and GAN), we can get diverse data by researches and experiments in the previous decade. Most of the models can be measured by Dice score and mloU (%). To begin with, it is evident that attention mechanism includes consequences in both Dice score and mloU (%), where ETNet(contain data set REFUGE and DRIVE), ASDNet and Res-SENet were proposed. Moreover, the mloU of the ETNet were 86.7%(RETUGE) and 77.44%(DRIVE), while the Dice score of ETNet(REFUGE), ASDNet and Res-SENet were 0.89, 0.91 and 0.878 correspondingly.

Besides, the mloU of Segformer-B5(Cityscapes) in Transformer occupied the leading position, contributing 84%, while the mloU of SETR-PUP and Segmenter-MASK (Cityscapes data set) clustered between 81% and 84%. In addition, Warxick-QU and DRIVE were two data set used in GAN-based transfer learning for a UNet segmentation and UNet GAN in GAN, whose Dice score were 0.63 and 0.82.

However, the models about Point Cloud need to be evaluated by data on disparate parts of the body, e.g., chest CT and MTRE (mm). Come up with the model based on DGCNN by Wang et al. In 2019, the data of Chest CT(Sen) was 0.859. Moving onto the models adopted with PPC, PPC-R, PPC-RM and PPC-L by Schaffert et al., the data as single vertebrae 3D CT/2D X-Ray were 0.75±0.21mm, 0.79±0.22mm, 1.18±0.42mm and 0.74±0.26 mm separately.

3.2 Discussion

Table 1. Performance of various semantic segmentation method on different datasets

Method	Algorithm	Datasets	mIoU (%)	Dice score
Attention mechanism	ETNet	REFUGE	86.7	0.89
Attention mechanism	ETNet	DRIVE	77.44	_
Attention mechanism	FCN	DRIVE	74.55	_
Attention mechanism	Unet	DRIVE	75.46	_
Attention mechanism	ASDNet	Pelvic	_	0.91
Attention mechanism	Res-SENet	3D-IRCADb	_	0.878
Point cloud method	DGCNN	Intra(DSC)	_	_
Point cloud method	PPC,PPC-R, PPC-RM,PPC-L	Intra(DSC)	_	_
Point cloud method	DGCNN+CPD	Intra(DSC)	_	_
Transformer	SETR-MLA	ADE20K	50.28	_
Transformer	SETR-MLA	PASCAL Context	55.83	–
Transformer	SETR-PUP	Cityscapes	81.64	_
Transformer	Segmenter-MASK	ADE20K	53.63	_
Transformer	Segmenter-MASK	PASCAL Context	59	–
Transformer	Segmenter-MASK	Cityscapes	81.3	_
Transformer	Segformer-B5	ADE20K	51.8	_
Transformer	Segformer-B5	Cityscapes	84	_
GAN	GAN-based transfer learning for a U-Net segmentation.	Warwick-QU	_	0.63
GAN	cGAN	N\A	_	
GAN	cGAN	TCIA	_	0.93
GAN	FCN	TCIA	_	0.79
GAN	UNet	TCIA	_	0.85
GAN	UNet GAN	DRIVE	_	0.82

Compared with other models, attention-based ET-NET has a very high ranking in both REFUGE and DRIVE dataset, and has been ranked first in REFUGE so far. Compared with contemporaneous and early models, attention-based models have higher segmentation accuracy in data sets. The

attention-based model places the end point of feature extraction in an important part of the image and ignores irrelevant content, so it has better feature extraction ability than the traditional convolutional neural network. In ADE20K, Cityscapes, and other data sets, many transformer models almost occupy the top position in the list. Among them, SETR, Segmenter, and Segformer, which have been put forward successively, have achieved excellent results in the dataset, and the segmentation accuracy has been continuously improved, far exceeding the traditional convolutional neural networks such as FCN.

Transformer based model can establish the connection of global context information, which is different from other methods and can well mitigate the impact of noise. In data sets such as DRIVE, the GAN model is among the best, especially the U-net model, which has a high score and recall rate in the test. Through the combination of U-net and GAN, the GAN model with high accuracy has well performance in medical data sets lacking accurate annotation and can segment images well.

4. Conclusion

In this work, we reviewed the recent applications of deep learning-based semantic segmentation methods in medical imaging through a comprehensive investigation of medical image segmentation methods. This paper reviews five mainstream algorithms in this field, which are cloud point-based algorithms, recurrent neural network, attention mechanism, transformer and generative adversarial networks. Extensive experiments from previous studies are conducted to evaluate the methods included in this paper. Results on various tasks e.g., lung imaging, vessel imaging and rectum imaging show that network which combines attention-based module with CNN or GAN have significantly improved its performance in comparison with the basic neural network of FCN and UNet. The main challenges of deep learning-based medical image segmentation methods at present are how to improve the segmentation accuracy and how to find the related information among image pixels in a better way. Therefore, point cloud-based methods with less computing cost, more compatible general attention module and unsupervised semantic segmentation methods are expected to be the feasible advancements in the upcoming years.

References

[1] Zhang, R, et al. Human Brain MR Image Segmentation Based on Level Set Method. Beijing Gongye Daxue Xuebao / Journal of Beijing University of Technology 43.2,2017:244-250.

[2] Wadhwa, Anjali, Anuj Bhardwaj, and Vivek Singh Verma. A review on brain tumor segmentation of MRI images. Magnetic resonance imaging 61,2019: 247-259.

[3] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[4] Lipton Z C, Berkowitz J, Elkan C. A Critical Review of Recurrent Neural Networks for Sequence Learning:, 10.48550/arXiv.1506.00019. 2015.

[5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Nets. Neural Information Processing Systems. MIT Press, 2014.

[6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE,2015: 3431-3440.

[7] Nguyen A, Le B. 3D point cloud segmentation: A survey. 2013 6th IEEE conference on robotics, automation and mechatronics(RAM). IEEE, 2013:225-230.

[8] Mnih V, Heess N, Graves A. Recurrent models of visual attention. Advances in neural information processing systems, 2014, 27.

[9] VASWANI A,SHAZEER N,PARMAR N,et al.Attention is all you need. Advances in Neural Information Processing Systems,2017:5998-6008.

[10] Besl P J, McKay N D. Method for registration of 3-D shapes. Sensor fusion IV: control paradigms and data structures. Spie, 1992, 1611: 586-606.

[11] Schaffert R, Wang J, Fischer P, et al. Metric-driven learning of correspondence weighting for 2-D/3-D image registration. German Conference on Pattern Recognition. Springer, Cham, 2018: 140-152.

[12] Wang Y, Solomon J M. Deep closest point: Learning representations for point cloud registration. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 3523-3532.

[13] Yang L, Chakraborty R. A GMM based algorithm to generate point-cloud and its application to neuroimaging. 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops). IEEE, 2020: 1-4.

[14] Hansen L, Dittmer D, Heinrich M P. Learning deformable point set registration with regularized dynamic graph cnns for large lung motion in copd patients. /International Workshop on Graph Learning in Medical Imaging. Springer, Cham, 2019: 53-61.

[15] Visin F, Kastner K, Cho K, et al. ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks:, 10.1109/TIP.2016.2548241. 2015.

[16] Visin F, Ciccone M, Romero A, et al. ReSeg: A Recurrent Neural Network-Based Model for Semantic Segmentation. Computer Vision & Pattern Recognition Workshops. IEEE, 2016.

[17] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, 2014.

[18] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: A survey. Computational Visual Media, 2022: 1-38.

[19] Nie D, Gao Y, Wang L, et al. ASDNet: attention based semi-supervised deep networks for medical image segmentation. International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 370-378.

[20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[21] Zhao Jie, Li Xu, Shen Tong. Abdominal artery segmentation based on SENet attention mechanism and deep residual network. Science Technology and Engineering, 2022, 22(22): 9529-9536.

[22] Zhang Z, Fu H, Dai H, et al. Et-net: A generic edge-attention guidance network for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 442-450.

[23] Vaswani A,Shazeer N, Parmar N,et al. Attention is all you need. Advances in Neural Information Processing Systems,2017:5998-6008.

[24] Dosovitskiy A,Beyer L, Kolesnikov A,et al.An image is worth 16×16 words:transformers for image recognition at scale. arXiv:2010.11929,2020.

[25] Zheng S, Lu J, Zhao H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. arXiv:2012.15840,2020.

[26] Strudel R,Garcia R, Laptev I,et al.Segmenter:transformer for semantic segmentation. arXiv: 2105. 05633, 2021.

[27] Xie E,Wang W,Yu Z,et al.SegFormer:simple and efficient design for semantic segmentation with transformers. arXiv:2105.15203,2021.

[28] Majurski M, Manescu P, Padi S, et al. Cell Image Segmentation Using Generative Adversarial Networks, Transfer Learning, and Augmentations. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2019.

[29] Eicken T V, Basu A, Buch V , et al. U-Net: a user-level network interface for parallel and distributed computing (includes URL). Acm Sigops Operating Systems Review, 1995, 29(5):40-53.

[30] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.

[31] Vks A, Ana C , Fa B , et al. Breast tumor segmentation in ultrasound images using contextual-information-aware deep adversarial learning framework. Expert Systems with Applications, 162. (2020).

[32] Zhang H, Goodfellow I, Metaxas D, et al. Self-Attention Generative Adversarial Networks: 10. 48550/ arXiv.1805.08318. 2018.