American Journal of Biomedical Science & Research

www.biomedgrid.com
ISSN: 2642-1747

Opinion

Copy Right@ Yong Wu

Application of Transformer in Medical Image Segmentation

Wenyin Zhang[1], Weijie Hao[1], Yuan Qi[2] and Yong Wu[2]*

[1]*School of Information Science and Engineering, Linyi University, China*

[2]*College of chemistry and chemical Engineering, Linyi University, China*

***Corresponding author:** Yong Wu, School of Information Science and Engineering, Linyi University, Linyi 276000.

To Cite This Article: *Wenyin Zhang, Weijie Hao, Yuan Qi, Yong Wu. Application of Transformer in Medical Image Segmentation. Am J Biomed Sci & Res. 2021 - 14(4). AJBSR.MS.ID.002014.* DOI: 10.34297/AJBSR.2021.14.002014.

Received: October 20, 2021; **Published:** October 25, 2021

Abstract

Accurate lesion segmentation is of great significance to the diagnosis of patients' illness and the formulation of treatment plan, and the segmentation results can provide doctors with important reference information. However, the focus of medical images is highly similar to the surrounding normal tissues, so the segmentation of medical images is a very challenging task. Transformer is gradually applied to the field of medical image segmentation by researchers with its strong learning ability.

Keywords: Medical Image; Segmentation; Transformer

Introduction

Medical image segmentation, as one of the most challenging tasks in clinical diagnosis, needs to identify and segment lesions from background medical images. However, this process is time-consuming and laborious, and the error rate is high. Traditional segmentation relies heavily on manual labeling information and clinical experience of clinicians. In addition, the complex changes presented by various imaging modes are also a great challenge to image segmentation. Therefore, developing a new algorithm for automatic segmentation of medical images is of great significance to improve the accuracy and efficiency of clinical diagnosis. Transformer can model and extract features from the global information of self-attention mechanism [1], but for medical image segmentation task, local and global features are very important for dense prediction, so there are many segmentation models that combine Transformer with traditional deep learning methods in the field of medical influence segmentation [2]. This paper first explains the background of the application of Transformer in medical image segmentation, then introduces in detail the recent research on image segmentation using Transformer, and finally looks forward to the future development.

Medical Image Segmentation based on Transformer

With the development of deep learning, a large number of coding-decoding based architectures and large-scale annotated medical data sets have emerged, and the accuracy of automatic segmentation of medical images has been continuously improved. However, due to the superposition of convolution layers and continuous sampling operations, the existing models have many problems of representation and information decline, resulting in that the models cannot completely model the global contextual feature dependencies. With the extensive success of Transformer in the field of machine translation and Natural Language Processing (NLP) [3] and the high accuracy in the task of image recognition [4], researchers have also taken bold attempts in the field of medical image segmentation.

Chen et al. [5] put forward a framework called TransAttUnet, which combines multi-level attention mechanism with multi-scale jump connection, and adds a Self-Aware Attention (SAA) module to the network, which also has Transformer Self-Attention (TSA) and Global Spatial Attention (GSA). this method can effectively learn the non-local interaction between encoder functions, and achieves good experimental results on several medical image segmentation datasets. Gao et al. [6] proposed a self-attention mechanism and UTNet for relative position coding, and proposed a new self-attention decoder for recovering fine-grained details from skipped connections in the encoder. Experiments show that this method can improve segmentation accuracy and has good robustness. Li et al. [7] proposed a Segtran framework based on Transformer, which mainly includes a compression module and an extension module, the compression module makes the transformer's self-attention regular, and the extension module is used to learn diversified representations. This method shows high accuracy in fundus images, colonoscopy images and brain tumor images, and has strong generalization ability. Hatamizadeh et al. [8] proposed an UNet Transformers(UNETR) framework, which uses a Transformer as an encoder to learn input information, and can effectively capture global multi-scale information. The encoder of Transformer is directly connected to the decoder through jumping connections with different resolutions [9], which increases the connection between codecs and codecs. This method has achieved good results in image segmentation tasks such as brain tumors and spleen. Karimi D et al. [10] proposed a method based entirely on self-attention between adjacent image blocks. The method divides 3D images into n3 3D slices, and then predicts the segmentation map of the central slice of the image through self-attention between these slices. This model can get good experimental results under the condition of small training data. Valanarasu J M J et al. [11] proposed a gated axial attention model, which extended the existing architecture by introducing additional control mechanisms into the self-attention module. At the same time, this method also proposed a Local-Global training strategy (LoGo), which further improved the segmentation performance. Chen et al [12] proposed a TransUNet which combined the transformer Decoder with UNet, and boosted more details by recovering local spatial information, the method achieved better performance on both multi-organ and heart segmentation tasks. Chang Y et al [13] proposed a TransClaw U-Net structure, which consists of an encoder part that combines convolutional operation and linear transform, and a decoder part to retain the bottom up sampled structure for better detail segmentation performance, and the convolutional part is used to extract shallow spatial features to recover the resolution of the up-sampled image, the Transformer is used without the coding part, the self-attention mechanism is used to obtain the global information of the image, the results indicated that the method achieved a great improvement in both accuracy and generalization ability.

Summary

Recently, most of the medical image segmentation tasks based on transformer are combined with traditional methods, such as UNet or CNN, the network model with Transformer often achieves state-of-art performance. Compared with traditional volume, Transformer makes full use of GPU resources by parallel operations, and measures performance in the minimum number of sequence operations required. In image segmentation tasks, it is a key challenge that learn dependencies between remote points, and the shorter paths between any combination of positions in the input and output sequences, easier to learn remote dependencies. In CNN, the number of operations required to compute the association between two positions by convolution grows with distance, but the transformer-based self-attention mechanism is independent of the number of operations required to compute the association between two positions and the distance; Meanwhile the self-attention mechanism used in Transformer is able to produce the self-attentive mechanism in Transformer to produce interpretable models, and each attention head learns to perform different tasks. Although the medical image processing based on transformer has shown excellent performance in all aspects, there are still many problems:

High computational consumption: compared with the lightweight CNN model, the image processing based on transformer does not significantly improve the running speed, and the medical image processing based on transformer will produce a large number of parameters, which will cause great pressure on the calculation. At the same time, it is difficult to realize quickly due to the huge cost of calculation; High requirements for data: when the amount of data is small, it is difficult to give full play to the excellent performance of transformer;

Future Research Directions

Based on the performance of transformer in medical image segmentation and the existing problems, this paper looks forward to the future development as follows:

Operation efficiency and economic benefits: in view of the problems of transformer in operation speed, large amount of data and high cost, in the next period of time, it will be a research hotspot to reduce the operation speed, parameters and cost of medical image segmentation based on transformer; Reduce the dependence of the model on the number of data: to solve the problem that transformer requires a high amount of data, it is an important research direction to find a model that passes through less data sets or labels less data sets without reducing the segmentation accuracy; Security issues: because the ways and conditions of medical image acquisition are different, and medical image processing involves the personal safety of patients, medical image segmentation based on transformer has great research value in terms of security, and the research on how to improve the robustness method will be concerned by researchers.

Acknowledge

This work was supported by the National Natural Science Foundation of China (61572230, 61173078, 61771230), and the Key R &amp; D plan of Shandong (2017CXZC1206, 2019GNC106027, 2019JZZY0101), Shandong Provincial Natural Science Foundation (ZR202102240817), and the Youth program of Shandong Natural Science Foundation (ZR2020QF011, ZR2020QF014).

References

1. Fu J, Liu J, Tian H, Li Y, Bao Y, et al. (2019) Dual attention network for scene segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 3146-3154.
2. Sinha A, Dolz J (2020) Multi-scale self-guided attention for medical image segmentation. IEEE journal of biomedical and health informatics. IEEE J Biomed Health Inform 25(1): 121-130.
3. Devlin J, Chang MW, Lee K, Toutanova K (2021) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
4. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, et al. (2021) An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR.
5. Chen B, Liu Y, Zhang Z, Guangming Lu, David Zhang (2021) TransAttUnet: Multi-level Attention-guided U-Net with Transformer for Medical Image Segmentation. arXiv preprint arXiv:2107.05274.
6. Gao Y, Zhou M, Metaxas DN (2021) UTNet: a hybrid transformer architecture for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham pp. 61-71.
7. Li S, Sui X, Luo X, Xinxing Xu, Yong Liu, et al. (2021) Medical Image Segmentation using Squeeze-and-Expansion Transformers. arXiv preprint arXiv:2105.09511.
8. Hatamizadeh A, Yang D, Roth H, Daguang Xu (2021) Unetr: Transformers for 3d medical image segmentation. arXiv preprint arXiv:2103.10504.
9. Petit O, Thome N, Rambour C, Loic Themyr, Toby Collins, et al. (2021) U-net transformer: self and cross attention for medical image segmentation. International Workshop on Machine Learning in Medical Imaging. Springer, Cham pp. 267.
10. Karimi D, Vasylechko S, Gholipour A (2021) Convolution-Free Medical Image Segmentation using Transformers. arXiv preprint arXiv:2102.13645. p. 78-88.
11. Valanarasu JMJ, Oza P, Hacihaliloglu I, Vishal M. Patel (2021) Medical transformer: Gated axial-attention for medical image segmentation. arXiv preprint arXiv:2102.10662.
12. Chen J, Lu Y, Yu Q, Xiangde Luo, Ehsan Adeli, et al. (2021) Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.
13. Chang Y, Menghan H, Guangtao Z, Zhang Xiao-Ping (2021) TransClaw U-Net: Claw U-Net with Transformers for Medical Image Segmentation. arXiv preprint arXiv:2107.05188.