2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)

TransUNet with Attention Mechanism for Brain Tumor Segmentation on MR Images

Enhao Wang
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
1809853di011008@student.must.edu.mo

Yaqi Hu
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
1809853di011003@student.must.edu.mo

Xiaoniu Yang
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
2009853gii20009@student.must.edu.mo

Xiaolin Tian
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
xltian@must.edu.mo

***Abstract*—Malignant brain tumor is a serious disease to human, causing cancer even death. Surgery is the main method for treatment of brain tumor. It's a key issue to detect the existence and position of brain tumor, for doctors to analyze the brain image and make a plan for treatment. This task, brain tumor segmentation, is mainly performed on MRI (Magnetic Resonance Image). In recent years, machine learning and deep learning received extensive attention and developed rapidly. This passage introduces a new deep learning model, TransNUNet, to perform tumor segmentation task on brain MRI dataset. This model introduces attention mechanism, Cbam, to TransUNet, and refines loss function. Research shows that TransNUNet has a higher Dice score than other two image segmentation nets, U-Net and TransUNet, which proves its potential on this task.**

***Index Terms*—brain tumor segmentation, magnetic resonance image, Transformer, TransNUNet**

I. Introduction

Cancer is a serious and essential disease for human, for its high death rate and difficulty to cure. According to World Health Organization [1], there are 10 million deaths from cancer in 2020, 1/6 of the total deaths of human. Cancer cells in brain, also called malignant brain tumor, grow abnormally and extend their area to unwanted brain parts, dealing damage to brain tissue. Although there is another type of tumor, benign tumor, that majority of them have no harm at the beginning, there is also a probability of them to deteriorate and become malignant tumor. Therefore, it's important to find and cure tumor as early as possible. In medical field, magnetic resonance image is mainly used to picture the inside structure and status of brain, for doctors to analyze brain disease. Formed using nuclear magnetic resonance technology, MRI has advantage of containing better resolution on soft tissue, and causing no ionizing radiation damage to human body. It is also widely used in tumor removal surgery.

This work is supported by Science and Technology Development Fund of Macao (FDCT-18-032-FI-001A/0009/2018/A).

Automated detection and segmentation of brain tumor by machine has growing importance. The accuracy and efficiency of tumor segmentation task, has a great impact on the success rate of brain tumor treatment. Deep learning, as its strong performance in image processing, is also used for medical image segmentation tasks. U-Net [2] model shows a great performance in image segmentation. This U-shaped structure has drawn great concern after it was proposed, lots of researchers made modifications and improvements on it. Until now, U-Net is still widely discussed, and its variants is so plenty, like two-stage U-Net [3], UNet++ [4] etc.

Transformer [5], first designed for natural language processing, is powerful in its structure that uses attention mechanism. In later research, it also shows great potential in the area of image processing. Matsoukas et al. [6] discussed the potential and probability that vision Transformer will replace convolutional neural network (CNN) on medical image processing. However, transformer-based architecture for visual application requires large datasets, but medical image dataset is usually limited. Comparing to CNN, Transformer is also weaker in capturing localized feature. For solution, J. Chen et al. combined U-Net and Transformer to create TransUNet [7], which gained from the advantage of both two structures.

Inspired by Transformer, we propose a new approach, TransNUNet [8]. Our method works on the structure of TransUNet, adds attention mechanism on it, and modifies the loss function. Our model achieved a good performance in brain MRI segmentation, and our Dice score is superior to U-Net and TransUNet. In the following parts of this passage, we will then illustrate proposed methodology, experiment results and conclusion.

II. Methodology

The steps in proposed methodology include implementing TransNUNet and refining loss function.

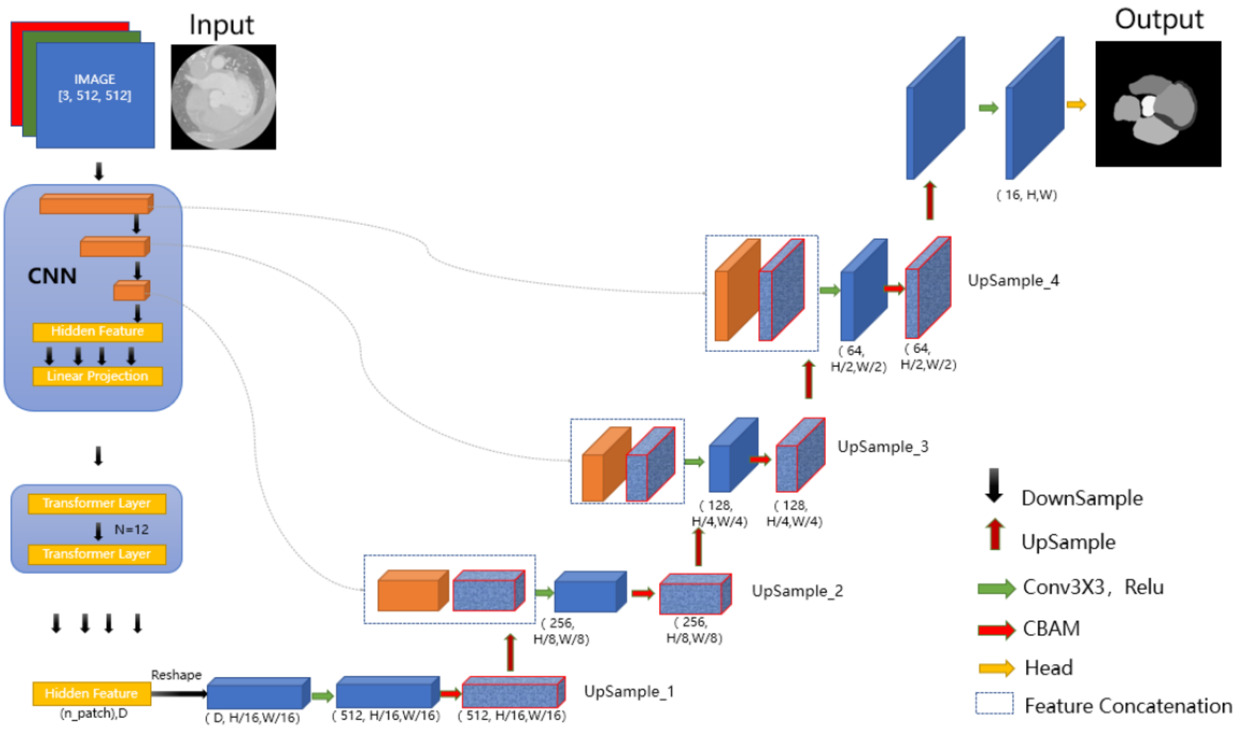


Fig. 1. Network Structure of TransNUNet [8]

A. TranaNUNet

As a deep learning model for image segmentation, TransNUNet combines CNN and Transformer together. The whole structure of network is shown as Fig. 1. The structure of TransNUNet is also U-shaped like U-Net, which divides into two stages, the first is down-sampling, going down from left, then up-sampling, going up to the right.

In the down-sampling stage, first is using CNN to preliminarily extract feature and obtain the feature map. In order to keep as much information of the original image as possible, the feature map gained is set as 1/8 of the size of the original image. Next, through sequentially stacked transformer layers, finer-grained feature extraction is implemented, and the down-sampling process is finished.

In the up-sampling stage, CNN as also used for decoder. We introduce an attention mechanism model Cbam [9] to enhance each model of different image sizes in the stage. These Cbam models enable the whole network to quickly locate the area of interest in the feature map, then analyze the region in detail. As shown in Fig. 1, the feature map created by down-sampling stage will go through Cbam and result in UpSample_1. Then feature concatenation is operated, and still through Cbam, UpSample_2 is obtained. Similar steps, UpSample_3 and UpSample_4 is created, through the two Cbam models that fuse 1/4 and 1/2 of the size of original image. These four attention mechanism parts can significantly enhance feature map and the classification ability of model.

B. Loss Function

In TransNUNet, we use cross-entropy as loss function for evaluation and optimization of classification part of model, and Dice loss as evaluation criteria for the performance of model's prediction. Equation (1) shows the formula of Dice function. In this function, A and B represent ground truth and prediction result by model. We add 1 to both the numerator and denominator, in order to avoid denominator becoming 0 and reduce overfitting. The more similar and closer prediction result and ground truth, the smaller the Dice loss, and we can say that the segmentation performance of model is better.

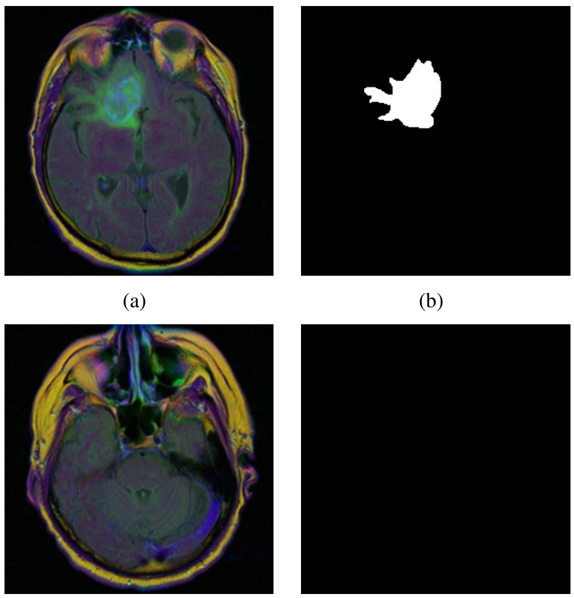
$$DiceLoss = 1 - \frac{2|A \cap B| + 1}{|A| + |B| + 1} \quad (1)$$

For result feature map, we implement bilinear interpolation resampling to use it as an auxiliary branch. Then for both this branch and original results put into optimizer, we use the two loss functions mentioned above to calculate for back propagation. We consider this method can improve the predict accuracy of pixels on the edge, thus the prediction performance of model can also be improved.

III. Experiment and Result

A. Dataset and Experiment Setup

The dataset used for training is from website www.kaggle.com. The size of all the MR images in dataset is 256*256*3, 3-channel pseudo-colored images. The dataset contains 3929 cases of 110 patients. Each case includes two images, one MR image with or without low-grade gliomas (LGG), and one manual FLAIR abnormality segmentation mask, as shown in Fig. 2.



(a) (b) (c) (d)

Fig. 2. Two example cases from dataset. Case 1: (a) brain image with tumor and (b) its mask; Case 2: (c) brain image with no tumor and (d) its mask.

For data processing, firstly, the dataset was split into train and test sets, with same portion of cases of brain tumor. For better performance of model, to enlarge the size of dataset, data augmentation is implemented. The brain MR image in each case is randomly processed to create augmented image. Random image processing method includes rotating 90°, flipping and adjusting image's brightness and contrast, each method has a probability of 0.5 to occur. In the experiment, we randomly select 1/10 of the train set as validation set, to observe the performance of model after each epoch in the training process.

We trained the model of TransNUNet, as shown in Fig. 1. Our GPU device is Nvidia Tesla V100 32G. At the beginning of training step, the batch size is 8 and the learning rate is 0.0001. Through some preliminary experiments, we conclude that the suitable number of epoch is 100, and using dataset after data augmentation has better performance than using original dataset.

B. Prediction Result and Analysis

After prediction process on the test set, we use dice similarity coefficient (DSC) [10] as standard for evaluation. DSC will calculate the similarity between prediction result and ground truth for images in the test set. The formula of DSC is shown as (2). We tried different batch sizes to evaluate the best batch size for the performance of our method. Fig. 3 shows the DSC scores of different batch sizes on TransNUNet. It can be seen that the best batch size falls on 12.

$$DSCscore = \frac{2|A \cap B|}{|A| + |B|} \tag{2}$$

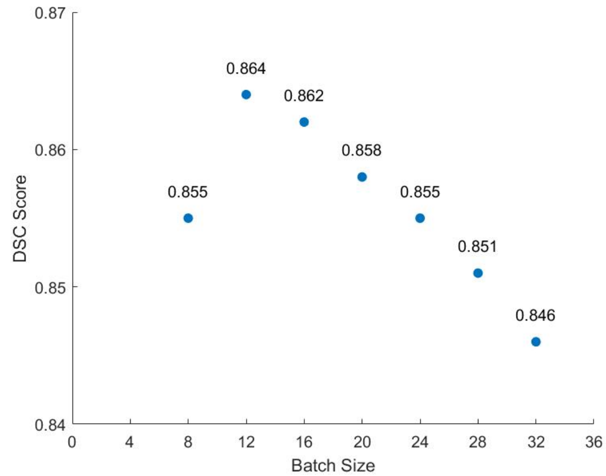


Fig. 3. DSC Score of TransNUNet under Different Batch Sizes

We also trained U-Net and TransUNet on the same brain MRI dataset. Fig. 4 shows the comparison of prediction result of three models. The five images in Fig. 4 are from test set. It can be seen that the overall prediction result of our method is more similar to ground truth than the other two, so that our method has better segmentation effect.

Using the same function, we also calculated DSC scores of U-Net and TransUNet. Table I shows the average DSC scores of three nets. The experimental result proves that our method achieves best result. We have better average score over the other two nets on brain MRI dataset, which indicates that our method has the ability and potential to do the brain tumor segmentation task. From our DSC score of 0.864, we can say that the proposed model, TransNUNet, has a strong learning ability, due to the attention mechanism we added to the network.

TABLE I
DSC Score of Different Nets on Brain MRI Dataset

Net	Score
U-Net [2]	0.811
TransUNet [7]	0.653
Proposed Method [8]	0.864

IV. Conclusion

In this experiment, we propose TransNUNet as a deep learning model for image segmentation. We introduce an attention mechanism into the net structure to improve its learning ability, and refine loss function for evaluation. We realize our work on brain MRI dataset for brain tumor segmentation, and achieve a better score on the performance comparing to two

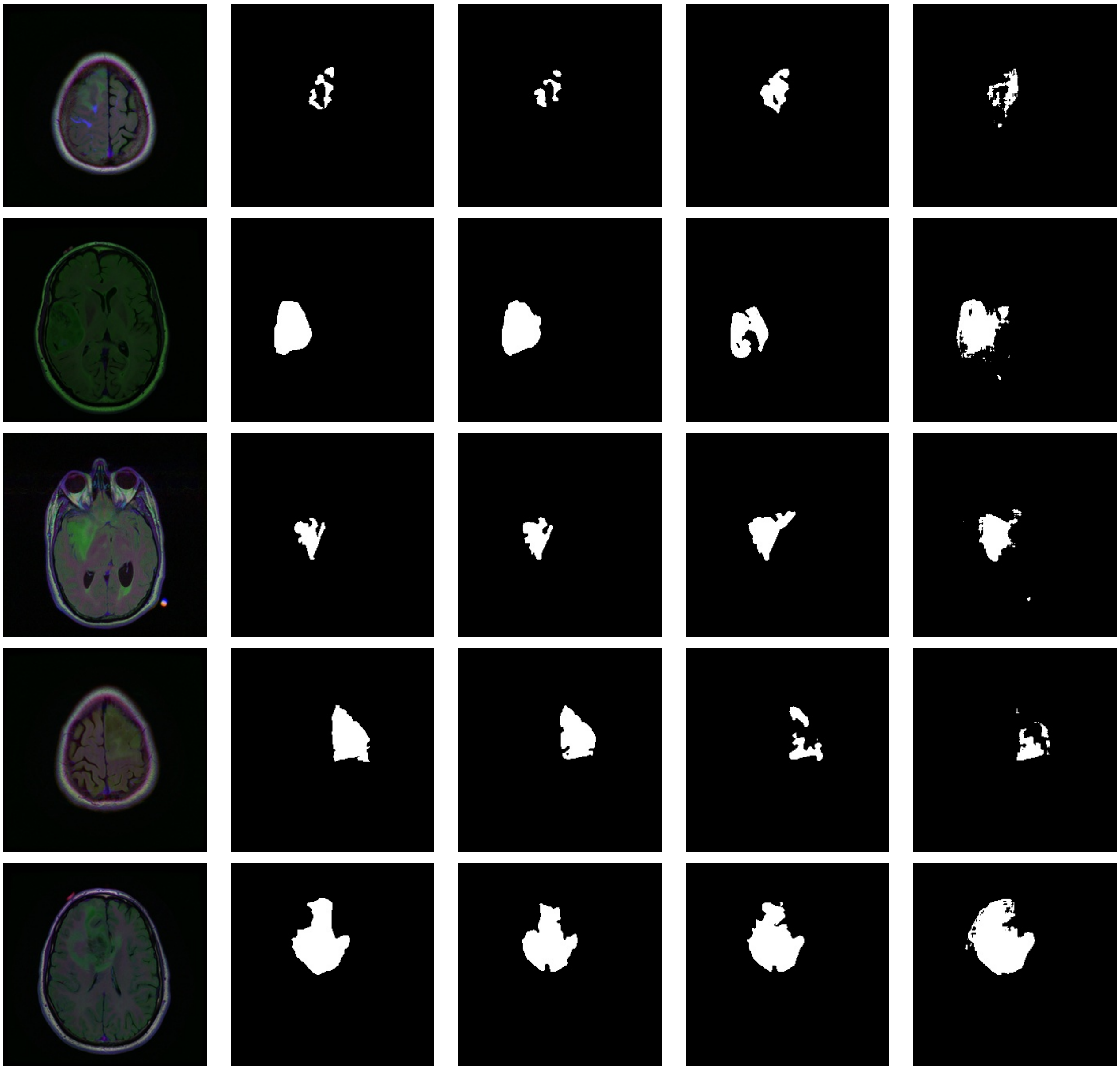


Fig. 4. (a) Brain MR Image. (b) Ground Truth. (c) Segmentation Result of Our Method. (d) Segmentation Result of U-Net. (e) Segmentation Result of TransUNet.

former image segmentation nets, which proves the effect of our work. We point out that attention mechanism has improving achievement on the network, enhancing the performance of handling image pixels in detail. We predict that our proposed method also has the potential to do other medical image segmentation tasks, not only brain MRI segmentation.

REFERENCES

[1] "Cancer", Who.int, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed: 19-April-2022].

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[3] Tao Liu, Yun Tian, Shifeng Zhao, Xiaoying Huang, and Qingjun Wang. Automatic whole heart segmentation using a two-stage u-netframework and an adaptive threshold window. IEEE Access, 7:83628–83636, 2019.

[4] Z.W. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, and J.M. Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Anylysis and Multimodal Learning for Clinical Decision Support*, pages 3-11, 2018.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion-Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008, 2017.

[6] Christos Matsoukas, Johan Fredin Haslum, Magnus Söderberg, and

Kevin Smith. Is it Time to Replace CNNs with Transformers for Medical Images? arXiv preprint arXiv:2108.09038, 2021.

[7] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan-Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

[8] Yang, Xiaoniu and Tian, Xiaolin. TransNUNet: Using Attention Mechanism for Whole Heart Segmentation. In *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA) Power, Electronics and Computer Applications (ICPECA), 2022 IEEE 2nd International Conference*, pages 553-556 Jan, 2022.

[9] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.

[10] Lee R Dice. Measures of the amount of ecologic association between species. Ecology, 26(3):297–302, 1945.