Brain Lesion Segmentation from MRI Images based on Attention Mechanism and U-Net

Haoyang Pei
Tondon School of Engineering
New York University
NY, USA
hp2173@nyu.edu

Yixuan Lyu
Tondon School of Engineering
New York University
NY, USA
yixuan.lyu@nyu.edu

***Abstract*—We present a multiple sclerosis lesion segmentation model based on attention mechanism and U-net, from 4 types of MRI images(T1 weighted, T2 weighted, PD-weighted and T2-weighted FLAIR). Our model leverages the different importance in a MRI image to give more weight to lesion part which leads to more precise segmentation result. Specifically, we first reproduce 3D U-net and Attention gate U-net from 2 papers. Then we propose and construct 3 types of network with placing attention layer in encoder process, decoder process and both side. We submitted our results to official grader and analysis the result in the end. Our work exploring the influence of attention mechanism to U-net and make comparison with the 5 models we constructed. As the output, we have a well-performed model based on encoder-decoder self-attention u-net.**

***Index Terms*—attention mechanism, U-net, lesion segmentation, MRI**

I. Introduction

Multiple sclerosis (MS) is a disease with great clinical significance that affects the central nervous system (CNS). But because MRI image is 3D, it usually takes too much time for medical participants to recognize the comprehensive lesion part. Also, the blur from MRI images between lesion and healthy part make doctors' position worse, even lead to misjudgement. Based on this, our goal is to propose a deep network which can accomplish segmentation task as precise as possible. To handle this, the most general and matured method is using U-net [1] which can generate preferable result by making decision across both local features and global view. We also learnt that attention mechanism can impose different weight to the input which is highly consist with our goal: we want our model give more attention to the lesion part. In this case, we enhance 3D U-net with mechanism using multi-head self-attention layer and have a persuasive result.

The contribution of this work include:

- Set up, train and compare 5 types of model. Considering the relationship between model structure and performance.
- Offering a persuasive model and parameter that work well.
- Open source our code, notebook and model parameter.

II. Datasets Description

In this report, we use dataset from ISBI 2015 Longitudinal MS Lesion Segmentation Challenge [3]. The description of dataset is shown in Table I below:

TABLE I
Dataset structure

Daraset	Number of subjects	Training set Testing set	MRI type	MRI generator
ISBI 2015	19	5/14	T1-w, T2-w, PD-w, FLAIR-w	3T Philips

This dataset includes T1-weighted, T2-weighted, PD-weighted, and T2-weighted FLAIR MRI with 3-5 time points acquired on a 3T MR scanner. Because MS lesion is a chronic disease, the acquisition time intervals were approximately 1 year to obtain the development history of lesion. And only 5 patients' data are recorded for training process.

III. Subtasks

The project has five main subtasks:

- Data Preprocessing.
- Implement the Attention U-Net in the Paper [1].
- Build the General 3D U-Net in the paper [2].
- Design the Multi-head Self-Attention 3D U-net.
- Train and Compare the performance of different models.

A. Data Preprocessing

a) Concatenate different MRI images: Because there are four types of MRI image, we first concatenate these four images into one bag as input to fully use the information. The concatenate process is shown in Fig 1.

b) Training and Validation Dataset Partition: Because there are connections between one patient's MRI image even in different time point.To prevent model remember patient character rather than lesion feature, in this project we divided training and validation set by patient(Patient 1-4 for training and Patient 5 for validation).

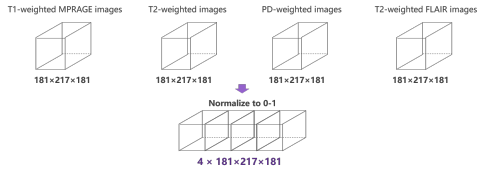

Fig. 1. Concatenate process

B. Implement the Attention U-Net in the Paper [1]

We accomplished the attention U-net [1] and use our datset for testing. In this model, attention mechanism is achieved by using attention gate, which is show in Fig 2

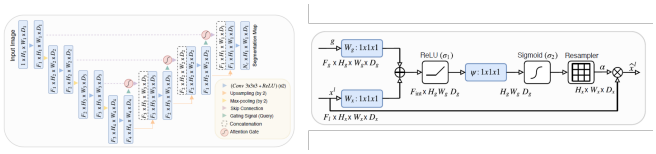

Fig. 2. Attention gate and model structure

The left figure shows the structure of model which adding attention gate to 3D U-net. While the right figure shows the internal structure of attention gate(GA).

C. Build the General 3D U-Net in the paper [2]

According to the paper [2], we build the 3D U-net Model from scratch. The shape of the 3D U-Net is similar with that of the 2D U-Net. However, the 3D U-Net uses 3D convolution to extract features from the 3D MRI images.

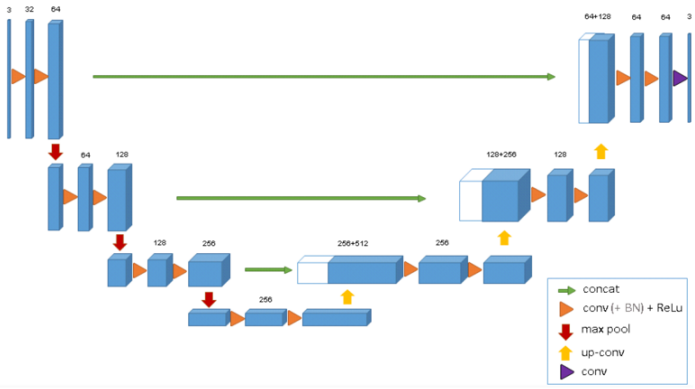

Fig. 3. 3D U-Net Architecture. Blue boxes represent feature maps. The number of channels is denoted above each feature map. [2]

D. Design the Multi-head Self-Attention 3D U-net

a) Method to Utilize the Multi-head Self-Attention in the U-Net: According to the paper [4], the formula of the Multi-head Self-Attention is shown in (1). Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions. With a single attention head, averaging inhibits this.

$$\begin{aligned} &\text{MultiHead}(Q,K,V) = \text{Concat}(\text{head}_1,...,\text{head}_\text{h})W^O \\ &\text{where head}_\text{i} = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \tag{1}$$

Where the projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$,$W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$,$W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$ and ,$W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$. In our project, $h = 3$. The paper [4] also explains the process of the Multi-head Self-Attention in the Fig.4.

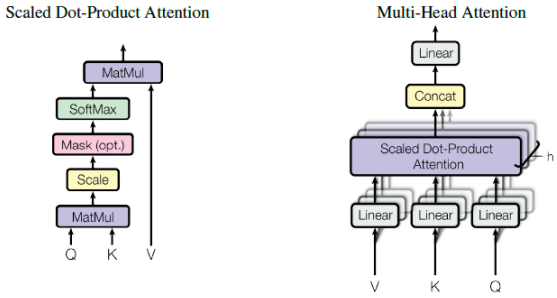

Fig. 4. (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel. [4]

In our project, the goal is to combine the Multi-head Self-Attention with 3D U-Net. However, the Multi-head Self-Attention cannot be directly applied in the 3D MRI images with 4 channels. Therefore, we design a 3D Multi-head Self-Attention implementation method to handle the 3D MRI images with 4 channels before combining it with the 3D U-Net. There are 6 steps:

1) Input Transformation
The goal of the input transformation is to convert the 5 dimensional input (including minibatch dimensional) to a 3 dimensional tensor so that it can be fed into the attention layer.The details are shown in Fig.5

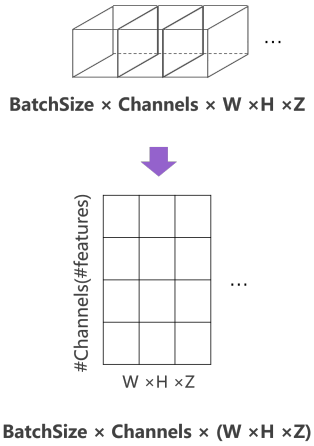

Fig. 5. Input Transformation. Transform the 3D image to 2D images

2) Linear Projection
This step can get multiple keys, queries and values of the original 3D images and significantly reduce the size of the input. Here we use $d_k = d_q = d_v = 10$, which means that it can reduce the size of the input to BatchSize × Channels × 10. The details are shown in Fig.6

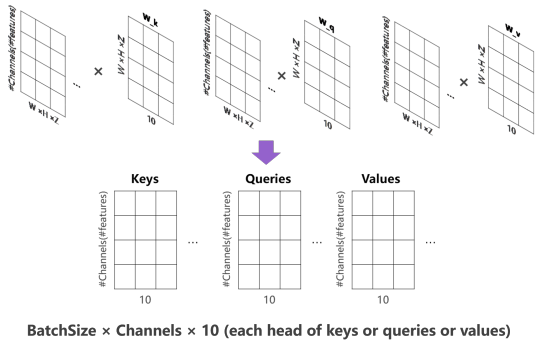

Fig. 6. Linear Projection. The figure only show the key, query, and value of one head. Each head has its own W_k, W_q, W_v

3) Compute Attention Map
This step is to compute attention map using multiple keys, queries and values. Each head of the Self-Attention layer has it own attention map and each attention map can attend different features. Through this process, the model can know which features or channels is more important to our results. The details are shown in Fig.7

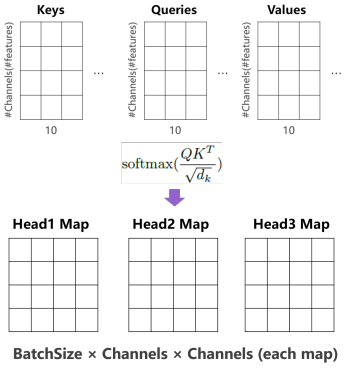

Fig. 7. Compute Attention Map.

4) Importance Weighting
This step gives each value of the output of the step different weights by multiplying attention map. The more important features can get a higher weights and the less important features can only get a lower weights.The details are shown in Fig.8
5) Concatenate and Linear Output
This step is to concatenate different weighted input and do linear transformation to get the result with original dimension.The details are shown in Fig.9
6) Inverse Transformation
The goal of the inverse transformation is to convert the 3 dimensional result into the original 5D input dimension.The details are shown in Fig.10

b) Encoder Multi-head Self-Attention 3D U-Net: Encoder Multi-head Self-Attention 3D U-Net means only adding the multi-head self-attention layer in the encoder process of the

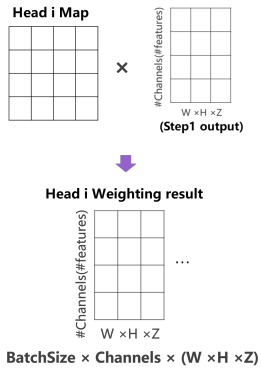

Fig. 8. Importance Weighting Process.

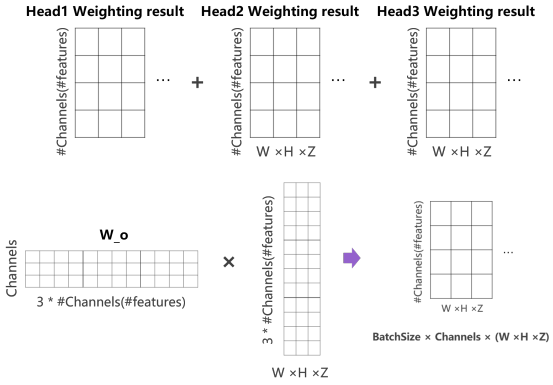

Fig. 9. Concatenate and Linear Output.

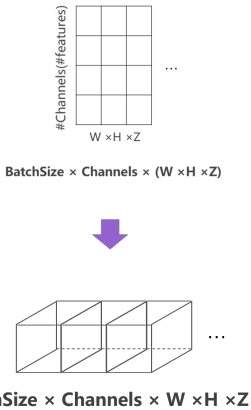

Fig. 10. Inverse Transformation.

3D U-Net. This type of architecture can extract the important features in the encoder process. The details are shown in Fig.11

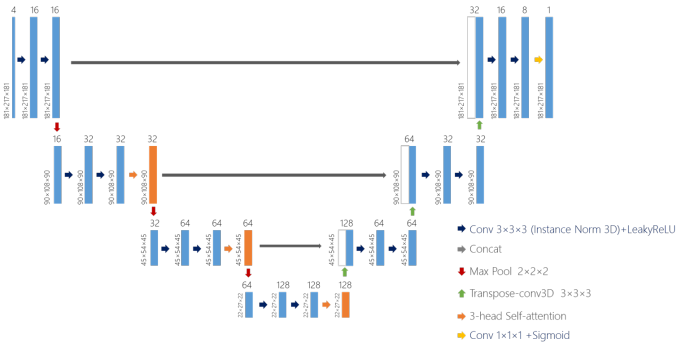

Fig. 11. The architecture of the Encoder Multi-head Self-Attention 3D U-Net.The orange block is the 3D Self-Attention layer with 3 head.

c) Decoder Multi-head Self-Attention 3D U-Net: Decoder Multi-head Self-Attention 3D U-Net means only adding the multi-head self-attention layer in the decoder process of the 3D U-Net. This type of architecture can extract the important features in the decoder process. The details are shown in Fig.12

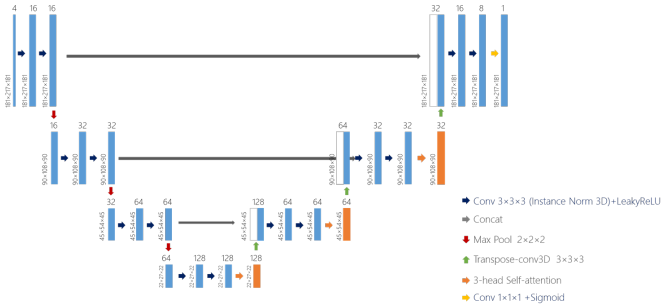

Fig. 12. The architecture of the Decoder Multi-head Self-Attention 3D U-Net.The orange block is the 3D Self-Attention layer with 3 head.

d) Encoder-Decoder Multi-head Self-Attention 3D U-Net: Encoder-Decoder Multi-head Self-Attention 3D U-Net means adding the multi-head self-attention layer both in the decoder and decoder process of the 3D U-Net. This type of architecture can extract the important features both in the encoder and decoder process. The details are shown in Fig.13

E. Train and Compare the performance of different models

a) Hyperparameters: We train 5 model using the same hyperparameters, including learning rate, training epochs, batch size, loss function, and training algorithm. The specific information for training is shown in Table.II.

b) Training and Compare the performance of different models: Our models are trained on the New York University High Performance Computing Greene using RTX8000 GPU, 8 cores CPU, and 32GB RAM. It costs roughly 4 hours for each model. After training, we compare the performance of different models in terms of different metrics. The results are shown in the Part.IV.

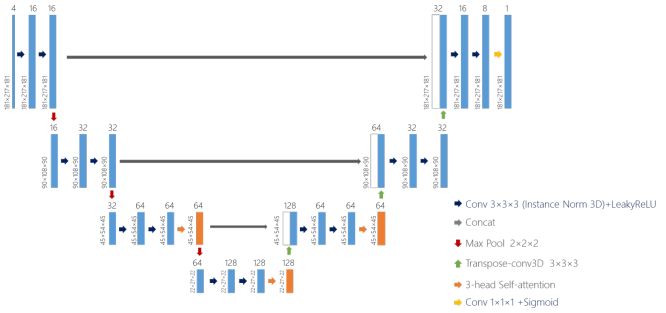

Fig. 13. The architecture of the Encoder-Decoder Multi-head Self-Attention 3D U-Net.The orange block is the 3D Self-Attention layer with 3 head.

IV. Experiments And Results

A. Training Curves

From the Fig.14, we see that:

- After adding Multi-head 3D Self-Attention in the U-Net, the speed of the convergence become faster than general U-Net in the paper [2] and the Attention U-Net in the paper [1].
- Encoder-Decoder Multi-head Self-Attention 3D U-Net has the lowest validation loss and highest dice score among 3 kinds of Self-Attention U-Net we designed. The Decoder Multi-head Self-Attention 3D U-Net has the worst performance in terms of the dice score.

B. Segmentation Results

From the Fig.15, the manual volume for mask 1 given by rater 1 is 4331 and the manual volume for mask 2 given by rater 2 is 4013. According to the Fig.16, except for the Decoder Multi-head Self-Attention 3D U-Net, the other four models has a relatively accurate segmentation volume.

C. Metrics Summary

According to the Table.III, we see that attention U-net in the paper [1] and Encoder-Decoder Multi-head Self-Attention U-Net have higher validation dice score and lower validation loss than other models. The results show that adding attention mechanism into the U-Net improve the segmentation accuracy.

D. Official Scores

We test those 5 models on the 14 official unlabeled test data and store the segmentation results as .nii files. Then we upload results to the official website and get the official test scores in Table.IV. From the results, We see that the Encoder-Decoder Self-Attention U-Net has the best performance in terms of Volume Correlation score, which proves that the Encoder-Decoder Self-Attention U-Net designed by ourselves improve the segmentation accuracy on dataset from ISBI 2015 Longitudinal MS Lesion Segmentation Challenge [3].

TABLE II
HYPERPARAMETERS OF TRAINING FOR 5 MODELS

Model	Learning Rate	Training Epochs	Batch Size	Loss Function	Training Algorithm
General 3D U-Net	5e-3	100	4	Soft Dice	Adam
Attention U-Net in the paper [1]	5e-3	100	4	Soft Dice	Adam
Encoder Multi-head Self-Attention 3D U-Net	5e-3	100	4	Soft Dice	Adam
Decoder Multi-head Self-Attention 3D U-Net	5e-3	100	4	Soft Dice	Adam
Encoder-Decoder Multi-head Self-Attention 3D U-Net	5e-3	100	4	Soft Dice	Adam

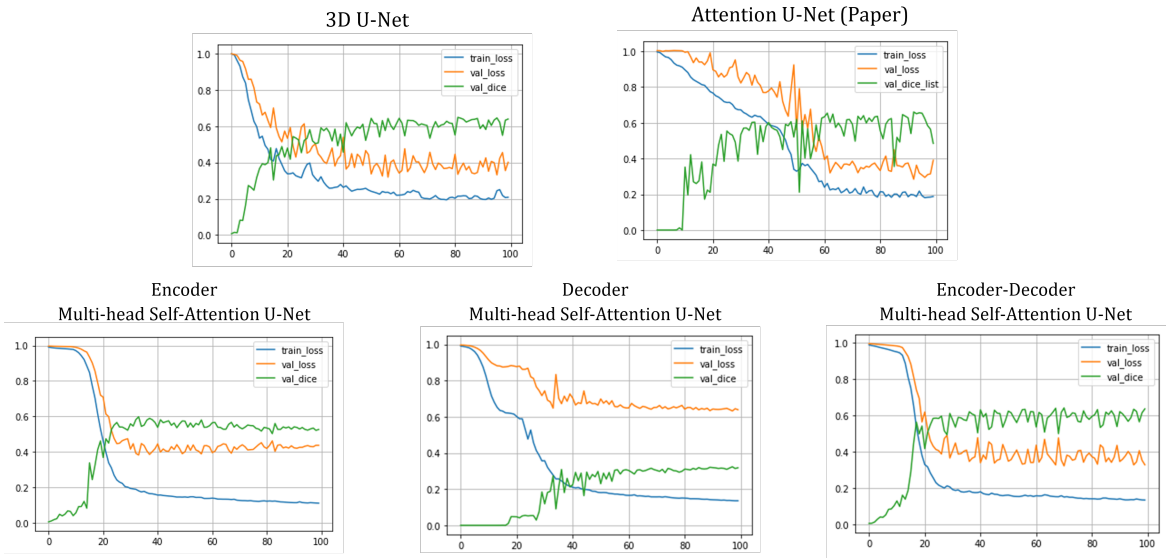

Fig. 14. The Training, Validation Loss Curves, and the Validation dice score curves for 5 models.The Blue line is the training loss curve. The orange line is the validation loss curve. The Green line is the validation dice score curves

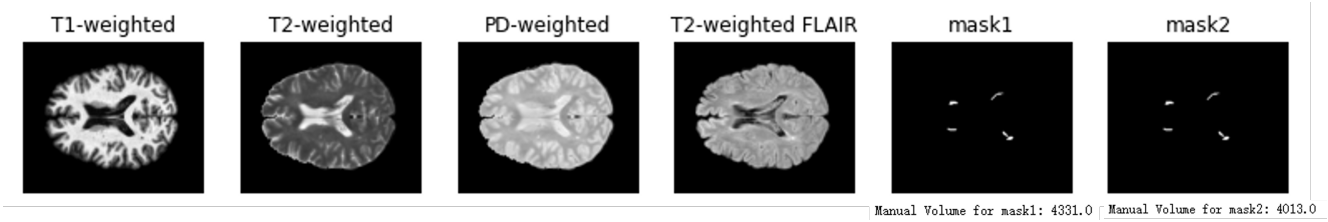

Fig. 15. The ground truth of the Training 05, Timepoint 01. The training 05 is the subject we used for validation in our project. It contains 4 timepoints and here we use timepoint 01 as demo to display the result.

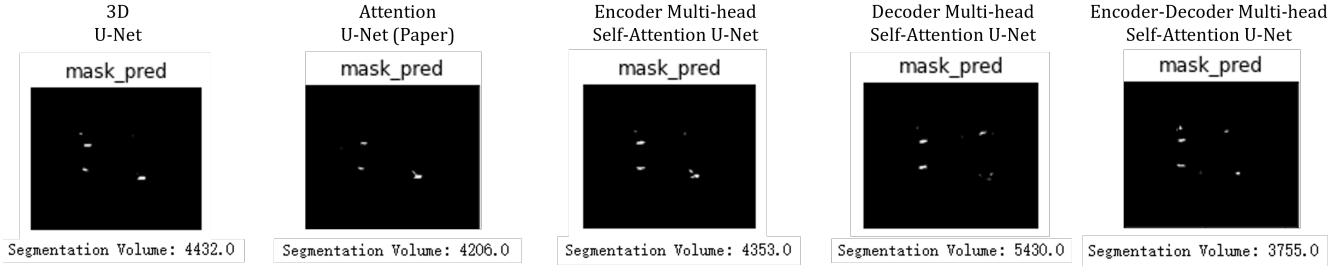

Fig. 16. The segmentation result of the Training 05, Timepoint 01 of 5 models

TABLE III
METRICS SUMMARY FOR 5 MODELS BASED ON VALIDATION DATASET(TRAINING 05)

Model	Val Loss (Rater1)	Val Loss (Rater2)	Val Dice (Rater1)	Val Dice (Rater2)	Average Val Loss	Average Val Dice
3D U-Net	0.371	0.430	0.663	0.624	0.401	**0.643**
Attention U-Net in the paper	0.277	0.338	0.699	0.646	**0.307**	**0.672**
Encoder Multi-head Self-Attention 3D U-Net	0.350	0.416	0.627	0.570	0.383	0.598
Decoder Multi-head Self-Attention 3D U-Net	0.616	0.662	0.336	0.301	0.639	0.318
Encoder-Decoder Multi-head Self-Attention 3D U-Net	0.294	0.366	0.670	0.616	**0.330**	**0.643**

TABLE IV
OFFICIAL SCORES FOR 14 UNLABELED TEST DATA

Model	Dice	Jaccard	PPV	TPR	LFPR	LTPR	VD	SD	Manual Volume	Segmentation Volume	Volume Correlation
3D U-Net	0.51056	0.354511	0.796488	0.397386	0.356993	0.341846	0.5061	4.508016	15648.29	6881.426	0.85233
Attention U-Net in the paper	0.55320	0.399631	0.671	0.567192	0.480162	0.410893	0.921791	4.831307	15648.29	17422.48	0.47763
Encoder Multi-head Self-Attention 3D U-Net	0.47354	0.321337	0.720005	0.378348	0.454582	0.297764	0.47773	4.9131	15648.29	7111.279	0.80549
Decoder Multi-head Self-Attention 3D U-Net	0.24424	0.142242	0.347998	0.222723	0.732975	0.253302	0.546265	5.688551	15648.29	9445	0.61711
Encoder-Decoder Multi-head Self-Attention 3D U-Net	0.53042	0.372394	0.724793	0.44887	0.540298	0.403205	0.428661	4.083102	15648.29	8925	**0.85772**

V. SUMMARY

We have finished 2 models in the paper [1] [2] and 3 models designed by ourselves for MS lesion segmentation task:

- 3D U-Net
- Attention Gate + U-Net
- Encoder Multi-head Self-Attention 3D U-Net
- Decoder Multi-head Self-Attention 3D U-Net
- Encoder-Decoder Multi-head Self-Attention 3D U-Net

Among all of the models, the Encoder-Decoder Multi-head Self-Attention 3D U-Net designed by ourselves has the best segmentation performance in term of the segmentation correlation score on dataset from ISBI 2015 Longitudinal MS Lesion Segmentation Challenge [3].

However, we still need to improve the performance of our model by adding RNN, LSTM, and Transformers to link the connection between each time point and also fine-tune our attention architecture using k-fold validation method.

REFERENCES

[1] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.

[2] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016: 424-432.

[3] Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. Neuroimage: 2017, 148;77-102

[4] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv preprint arXiv:1706.03762, 2017.

DEPOSITORY

https://github.com/HaoyangPei/Attention-Unet