SSNet: Synergistic Segmentation of Brain MRI Scans using nnUNetv2 and SAM-track

Yong Eun Jang[1,2†], Kwang-Ju Kim[2†], In-su Jang[2] and Gwang Lee[1*]

[1]*Department of Molecular Science and Technology University of Ajou, Suwon, S.Korea*
[2]*Artificial intelligence application Electronics and Telecommunications Research Institute, Daegu, S.Korea*
{jye120, glee}@ajou.ac.kr
{jef1015, kwangju}@etri.re.kr

***Abstract*— Segmenting lesions using brain MRI (Magnetic Resonance Imaging) images is a method that can help doctors determine the NIHSS (National Institutes of Health Stroke Scale) score. Therefore, we proposed the SSNet model to segment brain MRI images using a foundation model. Our SSNet model was constructed using nnUNetv2, a model with good performance in medical image segmentation, and SAM-Tracker, a foundation model. Our model's IoU (Intersection over Union) is 0.463, and Dice is 0.611, and when only nnUNetv2 is used, the IoU is 0.569, and Dice is 0.702. Although the performance is currently low, the foundation model is evolving, so it is expected that the model we proposed will be able to utilize it in brain MRI images end-to-end.**

Keywords—nnUNetv2, SAM-track, foundation model, artificial intelligence, deep learning, segmentation, MRI

I. INTRODUCTION

Semantic segmentation in medical image data greatly contributes to scientific discovery by converting it into meaningful and structured information. Recently, with the development of the foundation model, we wanted to use the foundation model to build a model that automatically sets the prompt using the existing segmentation model without the user having to set the prompt directly [1].

In particular, segmenting lesions using brain MRI images is a method that can help doctors determine the NIHSS score. NIHSS is a standardized assessment tool that evaluates the symptoms of stroke patients. It consists of 11 items that evaluate specific abilities and range from 0 to 42 points, with higher scores indicating a more severe stroke. However, since the score measured by each doctor is different, it is expected that segmentation of the lesion with brain MRI images will measure more accurate scores.

nnUNetv2 is a U-Net-based segmentation model specialized for medical image data. This model is a model that learns by automatically optimizing various parameters according to the input data and has superior performance for medical image data. Therefore, we used it as a base model in our model. SAM (Segmentation anything model)-track is a foundation model and is a SAM-based tracker. Since SAM-track tracks all objects in the image, we used SAM-track by using the mask image predicted by nnUNetv2 as a prompt [2-4]

--

Co-first Authors: jye120@ajou.ac.kr, kwangju@etri.re.kr

Corresponding Author: glee@ajou.ac.kr

II. METHOD

A. Data-set

A total of 267 brain MRI images and mask pair data consisting of 257 train sets and 10 test sets were used. Train set was learned by performing 5 cross-validation. The brain MRI data shape is 218 x 182 x 182 (width, height, layer). The types of brain data are adc_INM_MNI, b0_BC_INM_MNI, b1000_BC_INM_MNI, and the format is NIfTI format (.nii.gz).

B. Backward and forward method

The backward and forward method is the method used in SAM-track. As shown in Figure 1 of this paper, create brain image mp4 files in the backward and forward directions centered on the middle layer, which is the area with the largest lesion in the lesion mask image predicted by nnUNetv2. The lesion mask image, which is the area with the largest lesion, is used as a prompt for sam-track to track the mp4 file in the backward and forward directions. When multiple lesions were predicted in one brain MRI image, each lesion was separated and the backward and forward method was applied [5].

C. Model structure and training method

In this paper, a model was constructed using U-Net-based nnUNetv2 and SAM-based SAM-track to learn brain MRI images. NnUNetv2 is the best model for medical image segmentation. Therefore, we used nnUNetv2 as the base model. NnUNetv2 uses the nii.gz file, a brain MRI data format, and is trained through 5 cross-validation. As shown in Figure 2, learn the train set with nnUNetv2 and predict the test set. Before running SAM-track, perform backward and forward methods to create a brain MRI image mp4 file. Perform the SAM-track model using the Brain MRI image mp4 file and the lesion mask image predicted by nnUNetv2 as prompts. SAM-track is a tracker based on the SAM model and uses pre-trained weights of the SAM model without fine-tuning. This method is performed on all brain data formats: adc, b0, and b1000. When SAM-track is performed, several objects are tracked as the frame passes, and only the objects initially specified are post-processed. Additionally, since the result is a 2D image, it is finally merged into 3D to complete the final task.

III. RESULT

A. Predictive image analysis

Figure 3 shows the result of predicting nnUNetv2,

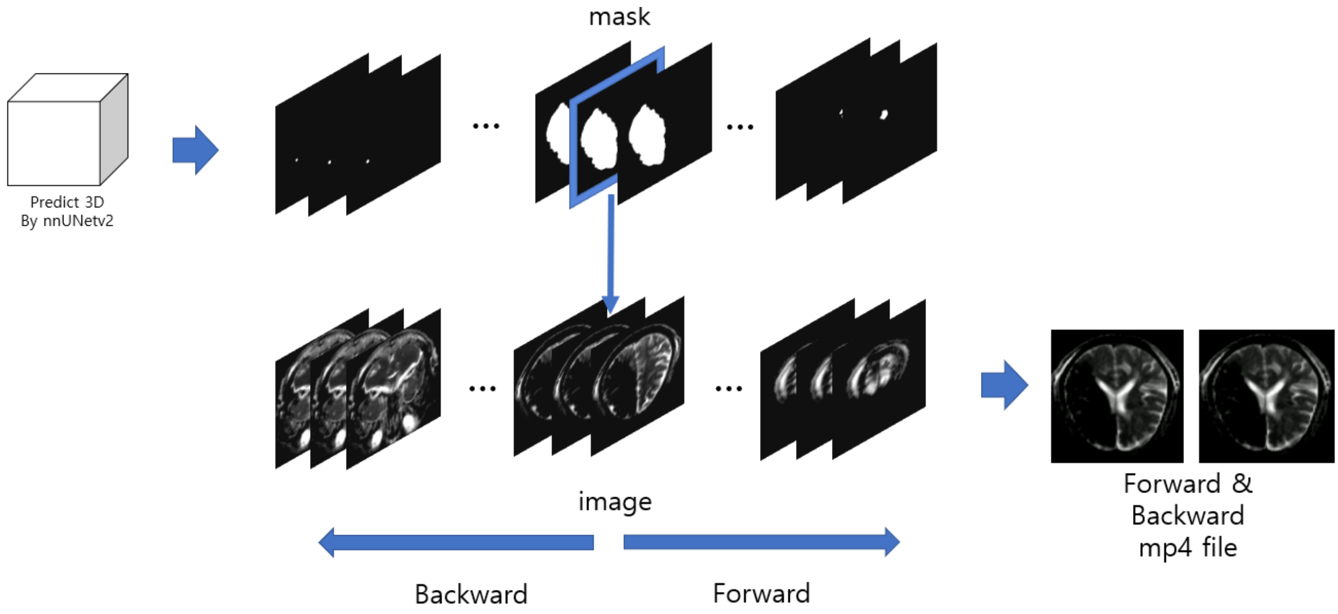

Fig. 1. The method of performing tracking by creating an mp4 file in the backward and forward directions, focusing on the central part with a large area in the mask image of the lesion predicted by nnUNetv2, is called the backward and forward method.

nnUNetv2 +SAM-track, and ensemble. When looking at a brain MRI image, if an area that should originally appear white is black, there is a lesion in that area. It can be seen that there is no significant difference in the lesions predicted by nnUNetv2, nnUNetv2 +SAM-track and ensemble. However, areas with large lesions are generally similar, but in areas with small lesions, the performance of the model combining nnUNetv2+SAM-track is somewhat inferior. Nevertheless, as the foundation model has recently developed, it is considered meaningful to create an end-to-end model by combining the existing model and the foundation model.

B. Metric and result

IoU and Dice were used to evaluate the performance of segmentation.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \qquad (1)$$

IoU compares the similarity between two arbitrary masks, as shown in Equation (1) [6].

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \qquad (2)$$

Dice measures the overlap of the area between two segmentations, A and B target mask as shown in Equation (2) [7].

As shown in Table 1, nnUNetv2's IoU value was 0.569, and the Dice value was 0.702. In the adc image using nnUNetv2 + sam-track, the IoU value was 0.417, and the Dice value was 0.566. In the b1000 image using nnUNetv2 + sam-track, the IoU value was 0.460, and the Dice value was 0.614. The IoU value of the ensemble of Adc, b0, and b1000 was0.463 and the Dice value was 0.611. Although the performance is not as good as the task using only nnUNetv2,

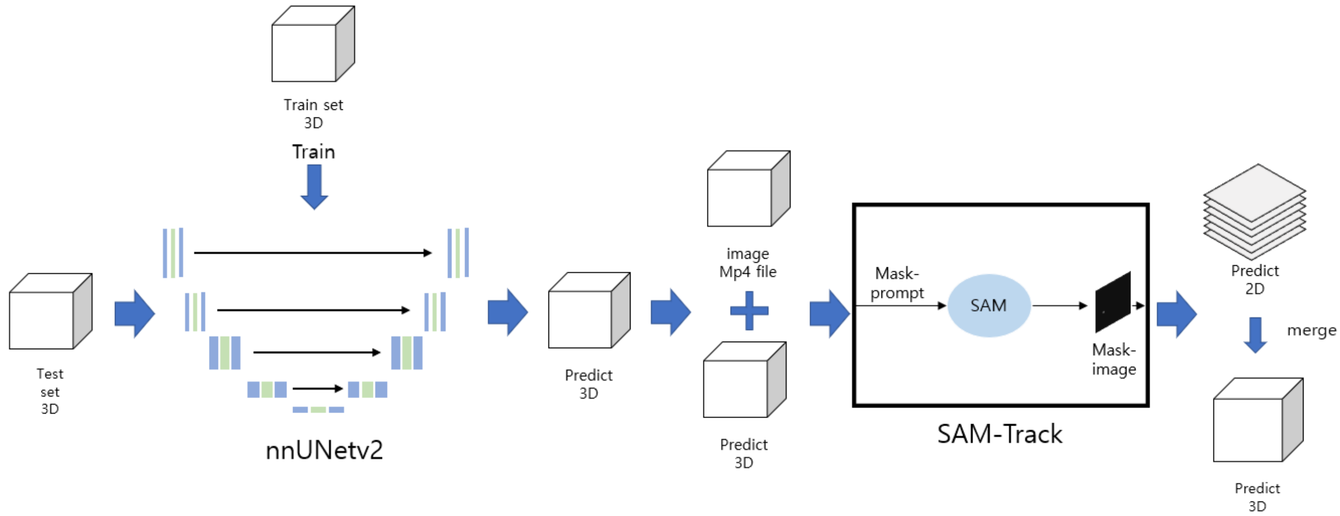

Fig. 2. The model network was built using nnUNetv2 and SAM-track. Learn nnunetv2 with the train set and predict the test set Prediction is made by performing a sam-track applying the backward and forward method using the image mp4 file and the mask image predicted by nnUNetv2 as a prompt. Finally, merge the predicted 2D image into a 3D image

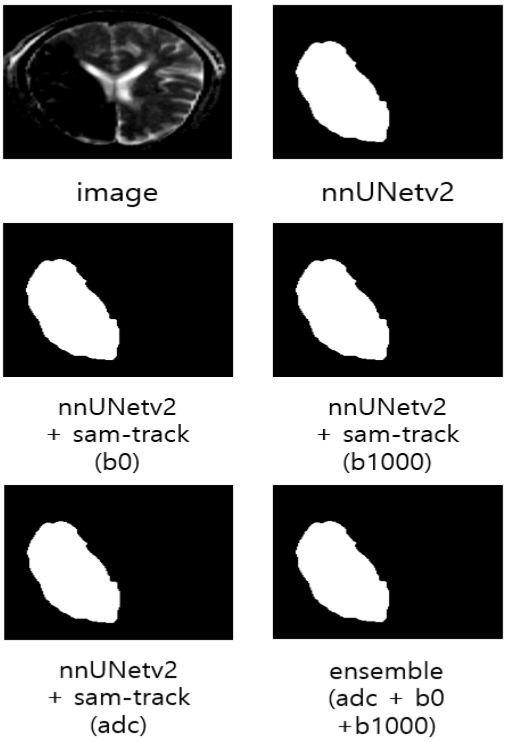

Fig. 3. The brain MRI image, the mask image predicted with nnUNetv2, the mask image predicted in the form of b0, b1000, and adc brain MRI data using nnUNetv2 + SAM-track, and finally the mask image obtained by ensemble of b0, b1000 and adc.

TABLE I. EXPERIMENTAL RESUTLS ON THE TEST-SET

Model	IoU (mean)	Dice (mean)
nnUNetv2	0.569	0.702
nnUNetv2 + SAM-track (adc)	0.417	0.566
nnUNetv2 + SAM-track (b0)	0.420	0.567
nnUNetv2 + SAM-track (b1000)	0.460	0.614
Ensemble (adc + b0 + b1000)	0.463	0.611

performance is not as good as the task using only nnUNetv2, it is at a similar level and is meaningful in that it builds the model end-to-end without the user directly setting the prompt by combining the recently developed foundation model.

IV. CONCLUSION

In this paper, we proposed a model to segment lesions in brain MRI images by combining the existing medical segmentation model, nnUnetv2, and the foundation model, SAM-track. We trained the trainset with nnUNetv2 through 5cross validation and predicted the lesion mask image for the test set. Using the backward and forward method, we created a brain image mp4 file in the backward and forward directions and performed a sam-track using the center area of the predicted lesion mask image as a prompt. Finally, the task was performed by merging the predicted 2D lesion image into 3D.

Compared to using only the existing nnUNetv2, we found that the IoU and Dice values of our proposed model were slightly lower. Therefore, it is meaningful in that it combines the recently developed foundation model and the existing segmentation model. In addition, as a future study, we plan to build a model to regress the NIHSS score using brain lesion segmentation.

ACKNOWLEDGMENT

This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean Government (Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry) under Grant 24ZD1100.

REFERENCES

[1] Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods 18.2 (2021): 203-211.

[2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[3] Cheng, Yangming, et al. "Segment and track anything." arXiv preprint arXiv:2305.06558 (2023).

[4] Isensee, Fabian, et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." Nature methods 18.2 (2021): 203-211.

[5] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[6] Cheng, Yangming, et al. "Segment and track anything." arXiv preprint arXiv:2305.06558 (2023).

[7] Kirillov, Alexander, et al. "Segment anything." arXiv preprint arXiv:2304.02643 (2023).