3D Abdominal CT Image Segmentation with nnUNet

Wenyuan Sun, Bowen Li, Yixin Yang, Jianxin Lei

Team name: MISEU

Abstract – Medical image segmentation via deep learning has been widely applied in computer aided diagnosis and medical treatment planning, while the prediction accuracy and resource consumption as well as time cost are still great challenges during the application. In this study, we introduce nnUNet in 3D abdominal CT image segmentation tasks and accurate segmentation results are generated through prediction.

***Index Terms* – computed tomography, abdominal organs, image segmentation, deep learning, nnUnet, FLARE21**

1. Introduction

Medical image segmentation based on deep learning have shown its great value on computer aided diagnosis and medical treatment planning. Segmentation methods with effective deep learning models have been implemented to achieve accurate region labeling and clear organ representation, which have been proved to be a great modality in many clinical scenarios. For example, Mehta *et al.* proposed a conceptually simple network for generating discriminative tissue-level segmentation masks for breast cancer diagnosis. [1] Shvets et al. proposed a deep learning method for automatic instrument segmentation in robot assisted surgery.[2]

However, for most medical image segmentation tasks, two main challenges limit the effectiveness and efficiency in real applications. (1) Segmentation accuracy: For most medical images such as CT and MRI, the image contrast is often low for accurate boundary detection and noises increase the difficulty of feature extraction and region segmentation. (2) High resource cost: for some 3D medical images, especially abdominal organ imaging results, the memory and time cost required for computation limits flexible applications such as real-time segmentation. Therefore, it is significant to propose a fast and low memory cost segmentation method to improve the efficiency of deep learning technologies.

In late 2020, a deep learning framework called nnUNet [3] was proposed based on UNet to achieve automated configuration for any datasets. Meanwhile, it is fast and data efficient, which makes it a suitable solution for the problem above. Thus, for FLARE21 challenge, we seek to apply nnUNet to the given abdomen CT image datasets, trying to achieve a superior segmentation performance.

2. Method

Figure 1 illustrates a basic U-Net architecture [4], on which nnUNet is built. Figure 2 shows how nnUNet automatically configures for the given input dataset [1]. No more new structures are involved in this model, while it builds dependencies between "dataset

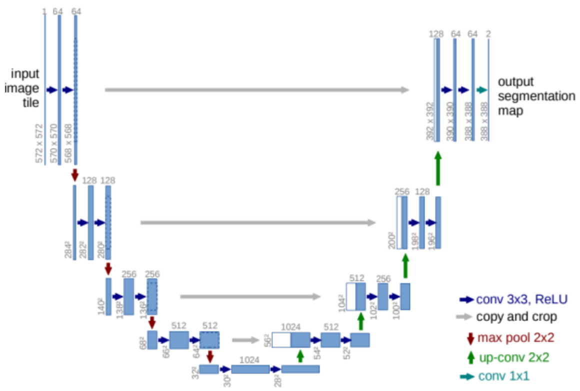

Figure 1. Basic UNet architecture [2]

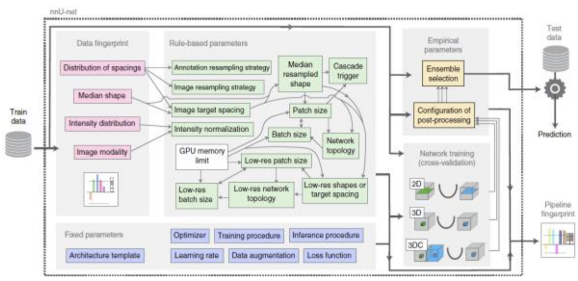

Figure 2. Schematic diagram for the nnUNet automated method configuration [1]

Table 1. Data splits of FLARE2021. [3-10]

Data split	Center	Phase	# Num.
Training (361 cases)	The National Institutes of Health Clinical Center	portal venous phase	80
	Memorial Sloan Kettering Cancer Center	portal venous phase	281
Validation (50 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
Testing (100 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
	Nanjing University	various phases	50

fingerprint" (i.e., key properties of the dataset) and "pipeline fingerprint" (i.e., key parameters of the method), which gives it ability to generate a suitable solution for the given input data.

In this work, we develop our method using nnU-Net method and 3d_fullres configuration. We use adam optimizer instead of SGD in the FLARE21 baseline method. In all, there are 30.78M trainable parameters among 115 layers.

3. Dataset and Evaluation Metrics

3.1. Dataset

The dataset used of FLARE2021 is adapted from MSD [5] (Liver [6], Spleen, Pancreas), NIH Pancreas [7-9], KiTS [10,11], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [12].

The total number of cases is 511. An approximate training/validation/testing split of 70%/10%/20% is employed resulting in 361 training cases, 50 validation cases, and 100 testing cases. The detail information is presented in Table 1.

3.2. Evaluation Metrics

The main evaluation metrics of this study includes: Dice Similarity Coefficient (DSC) and Normalized Surface Distance (NSD) for segmentation accuracy evaluation, and Running time as well as Maximum used GPU memory (when the inference is stable) for resource consumption measurement.

4. Implementation Details

Table 2. Environments and requirements

Ubuntu version	Ubuntu16.04
CPU	ES-1650
RAM	32GB
GPU	GTX-1080 Ti
CUDA version	10.0
Programming language	Python 3.7
Deep learning framework	nnUNet[3]

4.1. Environments and requirements

The environments and requirements of the baseline model used in our study are shown in Table 2.

4.2. Training protocols

The training protocols of our method is shown in Table 3. We also highlight what we have changed compared with FLARE21 baseline method.

5. Results

Results generated through the trained nnUNet are listed in Figure 3. It can be seen that the abdominal organs are successfully segmented. The high-quality organ segmentation results construct clear 3D organ representation based on CT data, and provides potential for further medical image analysis with advanced technologies.

However, as shown in the segmentation result of some cases, the performance in some cases is still

Table 3. Training protocols

Data augmentation methods	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring.
Initialization of the network	“he” normal initialization
Patch sampling strategy	More than a third of the samples in a batch contain at least one randomly chosen foreground class which is the same as nn-Unet [3].
Batch size	2
Patch size	80×192×160
Total epochs	100
Optimizer	Adam
Initial learning rate	0.0003
Learning rate decay schedule	ReduceOnPlateau
Stopping criteria, and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch (100).
Training time	17 h

considerably limited with some obvious errors of misclassification, due to the following reasons. Firstly, because of the limited time and low GPU performance, we only train the model with 100 epochs, which is obviously not enough that can be clearly seen from the training loss (see in Figure 4). Our results show a tendency that the kidney, liver and spleen are segmented with a relatively higher quality compared with pancreas, so it can also be a hint that the proposed method needs further adjustment and improvement when dealing with pancreas. One can also attribute the error to a reason that the feature extracted by convolution layers are too similar in these regions for the network to make the right decision, so it is necessary to improve the feature exploitation with advanced convolutional operation.

6. Discussion and Conclusion

In this study, the abdominal organs are segmented with the nnUNet, which provides fundamental support for further medical image analysis with this 3D organ representation. The performance needs further improvement, and the resource consumption still requires further research to improve the segmentation efficiency.

Acknowledgement

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers.

References

[1] Mehta S, Mercan E, Bartlett J, et al. Y-Net: joint segmentation and classification for diagnosis of breast biopsy images[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018: 893-901.

[2] Shvets A A, Rakhlin A, Kalinin A A, et al. Automatic instrument segmentation in robot-assisted surgery using deep learning[C]//2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018: 624-628.

[3] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[4] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted*

intervention. Springer, 2015, pp. 234–241.

[5] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze *et al.*, “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” *arXiv preprint arXiv:1902.09063*, 2019.

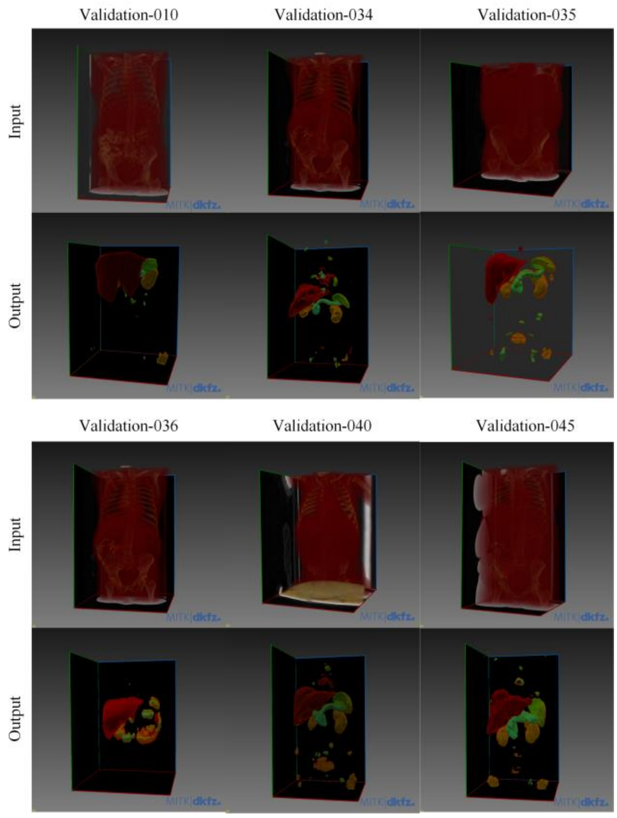

Figure 3. Examples of prediction results in FLARE21 challenge database

[6] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, “The liver tumor segmentation benchmark (lits),” a*rXiv preprint arXiv:1901.04056*, 2019.

[7] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, “Data from pancreas-ct. the cancer imaging archive (2016).”

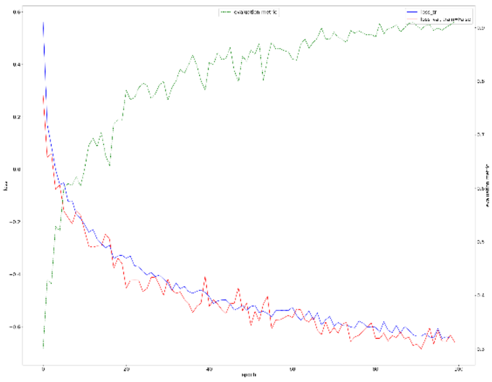

Figure 4. Training loss and evaluation metric

[8] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, “Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564.

[9] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, “The cancer imaging archive (tcia): maintaining and operating a public information repository,” *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[10] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge,” *Medical Image Analysis*, vol. 67, p. 101821, 2021.

[11] N. Heller, S. McSweeney, M. T. Peterson, S. Peterson, J. Rickman, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, J. Rosenberg *et al.*, “An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging.” *American Society of Clinical Oncology*, vol. 38, no. 6, pp. 626–626, 2020.

[12] J. Ma, Y. Zhang, S. Gu, Y. Zhang, C. Zhu, Q. Wang, X. Liu, X. An, C. Ge, S. Cao *et al.*, “Abdomenct-1k:

Is abdominal organ segmentation a solved problem?" *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2021.