Abdominal multi-organ segmentation in CT using nnU-Net

Cheng Li

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
lchengx27@gmail.com

Abstract. Abdominal multi-organ segmentation in computed tomography (CT) is crucial for many clinical applications including disease detection and treatment planning. Deep learning methods have shown unprecedented performance in this perspective. However, it is still quite challenging to accurately segment different organs utilizing a single network due to the vague boundaries of organs, the complex background, and the substantially different organ size scales. In this work, we investigated the feasibility of applying the famous nnU-Net to performing abdominal multi-organ segmentation in CT [4]. By slightly modifying the configurations of nnU-Net, we obtained promising segmentation results. Specifically, quantitative evaluations on the FLARE2022 validation cases (20 cases) show that the method achieves an average Dice similarity coefficient (DSC) of 0.71 and average normalized surface distance (NSD) of 0.76. With further optimization, it is possible to obtain satisfactory segmentation results.

1 Introduction

Multi-organ segmentation in abdominal computed tomography (CT) is very important for many clinical applications. Two example images are shown in Fig. 1. Three main difficulties can be observed for the segmentation task: 1) The vague boundaries of different organs. 2) The complex and heterogeneous background. 3) The large size variations exist for different organs. Consequently, it becomes very challenging to segment all the organs simultaneously with a high accuracy.

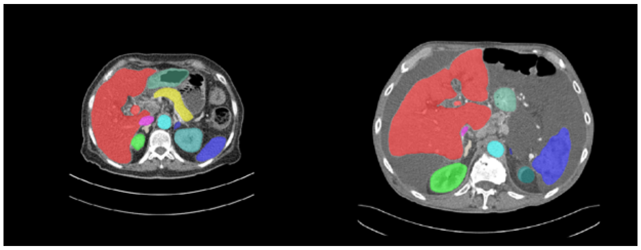

Fig. 1. Example abdominal CT images with different organs delineated manually.

nnU-Net is a highly effective self-configuring deep learning method for biomedical image segmentation [4]. It can automatically configure itself to achieve the best performance on a given dataset by searching for the optimal configurations of preprocessing, network architecture, training, and post-processing. The authors have experimented with 23 public datasets and state-of-the-art segmentation results have been achieved by utilizing nnU-Net [4].

In this work, we test the feasibility of applying nnU-Net to the task of multi-organ segmentation in abdominal CT. This work is only a proof of concept. Further improvements are needed to achieve satisfactory segmentation performance.

2 Method

Our model is developed based on the famous nnU-Net [4]. Compared to the naive nnU-Net, we set the target spacing to (5, 1.6, 1.6) instead of (2.5, 0.8, 0.8) for preprocessing. During inference, we set the step size to 1. For the model training, we utilized only the labeled training data. The unlabeled data were not exploited, which could be another direction for improvement. All the other settings can refer to the original paper [4].

Fig. 2 illustrates the applied 3D nnU-Net [4]. Basically, a U-Net with an encoder-decoder architecture is adopted. The input images are firstly downsampled to extract low-resolution high-level semantic features. Then, these features are upsampled to generate the segmentation outputs with the same resolution as that of the input images.

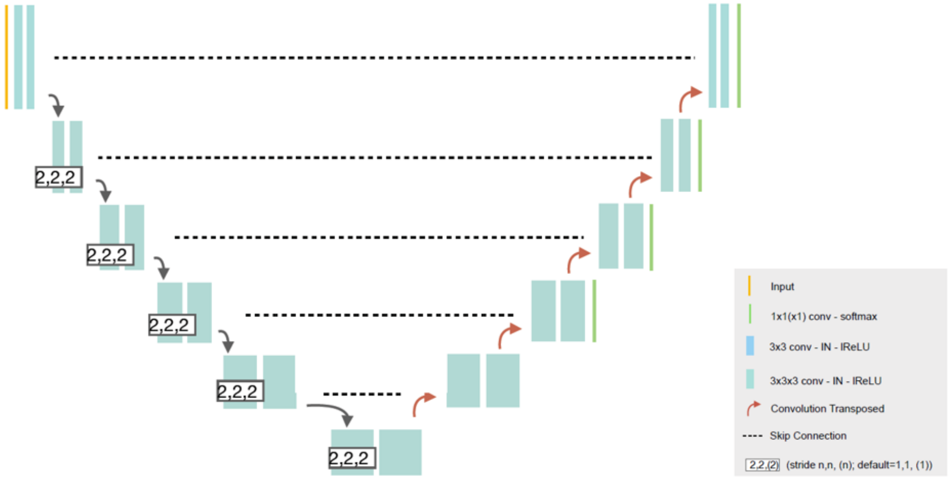

Fig. 2. Network architecture

To train the network, we use the summation between Dice loss and cross entropy loss because compound loss functions have been proved to be robust in various medical image segmentation tasks [8].

3 Experiments

3.1 Dataset

The FLARE2022 dataset is curated from more than 20 medical groups under the license permission, including MSD [11], KiTS [2,3], AbdomenCT-1K [9], and TCIA [1]. The training set includes 50 labelled CT scans with pancreas disease and 2000 unlabelled CT scans with liver, kidney, spleen, or pancreas diseases. We did not exploit the provided unlabeled data. The validation set includes 50 CT scans with liver, kidney, spleen, or pancreas diseases.

3.2 Evaluation Metrics

The evaluation measures consist of two accuracy measures: Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD), and three running efficiency measures: running time, area under GPU memory-time curve, and area under CPU utilization-time curve. All measures will be used to compute the ranking. Moreover, the GPU memory consumption has a 2 GB tolerance.

3.3 Implementation details

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Gold 6240 CPU@2.60GHz
RAM	503 GB
GPU (number and type)	Five NVIDIA V100 32G
CUDA version	11.2
Programming language	Python 3.9
Deep learning framework	Pytorch (Torch 1.11.0, torchvision 0.12.0)

Training protocols The training protocols of the adopted method is shown in Table 2. The applied augmentation methods are the same as those utilized in nnU-Net, including rotations, scaling, Gaussian noise, Gaussion blur, brightness, contrast, simulation of low resolution, gamma correction, and mirroring.

Table 2. Training protocols.

Network initialization	“he" normal initialization
Batch size	2
Patch size	40×224×192
Total epochs	1000
Optimizer	SGD with nesterov momentum ($\mu = 0.99$)
Initial learning rate (lr)	0.01
Lr decay schedule	Poly learning rate decay $0.1 \times (1 - epoch/1000)^{0.9}$
Training time	8 hours
Number of model parameters	8.14M
Number of flops	35.6G
CO_2eq	1 Kg

4 Results and discussion

4.1 Quantitative results on validation set

The average running time is 56.73s per case in inference phase. The maximum used GPU memory is 1701MB. The average area under GPU memory-time curve is 84059.24, and average the area under CPU utilization-time curve is 910.172. Table 3 lists the results on the validation set. Overall, better results are achieved for larger and regular organs like the liver and the kidney. Worse results are achieved for smaller and complex organs like the gallbladder and the duodenum. These results indicate that it is difficult to handle the size variations utilizing the baseline nnU-Net, and specific modules should be designed to particularly address the issue.

Table 3. Quantitative results of validation set in terms of DSC and NSD.

Organ	Liver	RK	Spleen	Pancreas	Aorta	IVC	RAG
DSC	0.93	0.79	0.79	0.65	0.89	0.79	0.62
Organ	LAG	Gallbladder	Esophagus	Stomach	Duodenum	LK	
DSC	0.52	0.52	0.71	0.75	0.50	0.75	
Organ	Liver	RK	Spleen	Pancreas	Aorta	IVC	RAG
NSD	0.92	0.80	0.80	0.74	0.92	0.77	0.78
Organ	LAG	Gallbladder	Esophagus	Stomach	Duodenum	LK	
NSD	0.65	0.48	0.82	0.79	0.67	0.76	

4.2 Qualitative results on validation set

Fig. 3 presents the segmentation results of two cases for which satisfactory segmentation accuracy is achieved. Fig. 4 shows the segmentation results of two cases for which low segmentation accuracy is achieved. It can be observed that for those easy samples, the background is quite simple, whereas for hard samples, the background is quite complex. Meanwhile, the most obvious problem for the low accuracy is under-segmentation. Possible reasons could be that the model training is not sufficient or that the model complexity is not enough due to the small training data we utilized. I believe more accurate results can be obtained if we exploit the unlabeled data in an effective way.

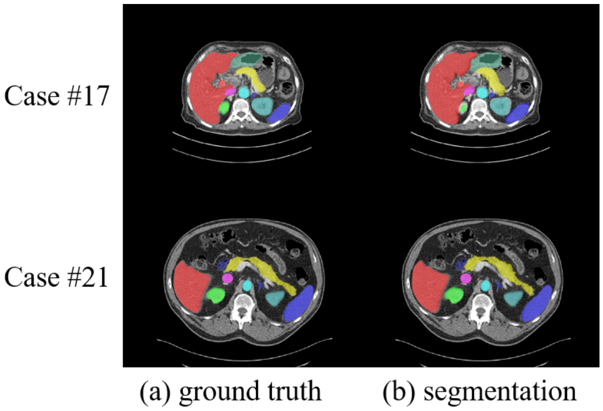

Fig. 3. Example segmentation results with high segmentation accuracy

5 Limitation and future work

In this work, we test the feasibility of utilizing nnU-Net for multi-organ segmentation in abdominal CT. nnU-Net is a powerful tool. Its effectiveness has been validated on different tasks [4]. However, only moderate segmentation performance is achieved in this work. The three difficulties of multi-organ segmentation in abdominal CT we discussed were not properly addressed by the baseline nnU-Net we implemented.

There are many limitations for this work that should be noticed for future improvements. First, the large amount of unlabeled data were not exploited. Currently, there are many semi-supervised methods, such as the co-training and pseudo-labeling methods [6,7,10], that can be utilized to extract information from the unlabeled data and improve the model's accuracy and robustness. Second, the data imbalance issue was not solved. Different weights might be needed

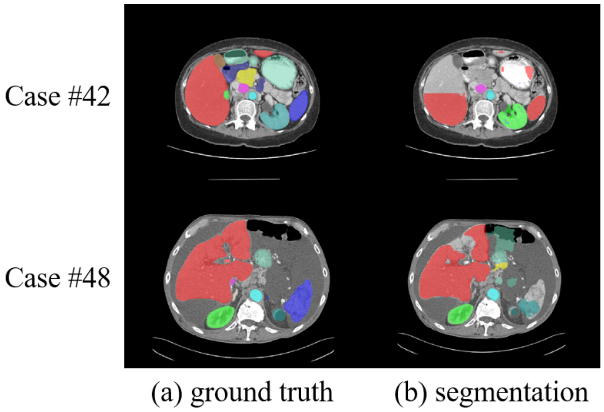

Fig. 4. Example segmentation results with low segmentation accuracy

for the loss calculation of different organs considering the different sizes. Calculating boundary losses might be more suitable than calculating volume losses [5]. Lastly, no special preprocessing or post-processing was conducted.

6 Conclusion

We implemented the famous nnU-Net to address the issues encountered during segmenting multi-organs in abdominal CT. We found that although promising results are obtained for large and regular organs, the segmentation performance on small and irregular organs are not satisfactory. Further optimizations in method configurations are needed. Besides, the current work did not exploit the unlabeled data. Further attempts should be made to utilize those data to improve the segmentation performance.

Acknowledgements The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2022 challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention.

References

1. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013) 3

2. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., 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 **67**, 101821 (2021) 3
3. Heller, N., McSweeney, S., Peterson, M.T., Peterson, S., Rickman, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Rosenberg, J., 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 **38**(6), 626–626 (2020) 3
4. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**, 203–211 (2021) 1, 2, 5
5. Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ben Ayed, I.: Boundary loss for highly unbalanced segmentation. In: International conference on medical imaging with deep learning. pp. 285–296 (2019) 6
6. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: ICML 2013 Workshop: Challenges in Representation Learning (2013) 5
7. Li, Y., Chen, J., Xie, X., Ma, K., Zheng, Y.: Self-loop uncertainty: A novel pseudo-label for semi-supervised medical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 614–623. Springer (2020) 5
8. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021) 3
9. Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., Cao, S., Zhang, Q., Liu, S., Wang, Y., Li, Y., He, J., Yang, X.: Abdomenct-1k: Is abdominal organ segmentation a solved problem? IEEE Transactions on Pattern Analysis and Machine Intelligence (2021). https://doi.org/10.1109/TPAMI.2021.3100536 3
10. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: European Conference on Computer Vision. pp. 135–152 (2018) 5
11. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019) 3