Abdominal organs segmentation based on nnUNet of the FLARE21

YiFei Wang
Bioengineering College of CQU
Shapingba, ChongQing
wngyifei@foxmail.com

Abstract

Segmenting Abdominal organs is challenging, as classic methods fail to produce satisfactory results and modern supervised learning (deep learning) is often time-consuming and can some time be infeasible. Thus, We provide a baseline implementation of the FLARE21 challenge based on 3D nnU-Net, which is the SOTA method of multi-organ segmentation. The proposed network is trained in an end-to-end fashion, segmenting Abdominal CT images that seperate the background of the irrelevant parts and preserve the organ structures. We train and evaluate nnUNet model on a dataset of 60 digital Abdominal images, and experiments demonstrate the supreme accuracy of the baseline method on Abdominal organs segmentation.

1. Introduction

The aim of semantic segmentation in biomedical field is to highlight specific regions in raw medical images (such as CT), which makes it play an essential role for disease detection and organ location.

However, there are mainly two key points to build a medical segmentation method. The first challenge comes from the diversity of the dataset, which including multi-center, multi-phase, multi-vendor, and multi-disease cases. Well generalization of the applied method is required.The second challenge comes from the efficiency requirement for the proposed solutions.

To addressing above difficulties, we employed the 3D nnU-Net [1] as our solution as a result of the following reasons. First, nnU-Net can automatically configures itself for the entire segmentation pipeline, including preprocessing, network architecture, training and post-processing. Second, it is fast. Third, it shows powerful performance on several segmentation tasks.

2. Method

nnU-Net is a deep learning framework that condenses the current domain knowledge and autonomously takes the key decisions required to transfer a basic architecture to different datasets and segmentation tasks.

Figure 1 illustrates the applied 3D nnU-Net, where a U-Net [2] architecture is adopted.

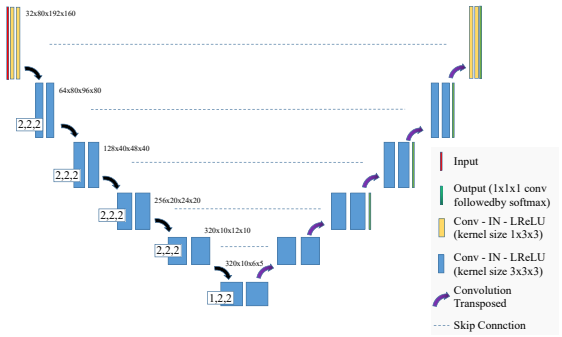

Figure 1. Network architecture

2.1. Preprocessing

In order to adapt into any kind of data from differet scene, nnUNet proposes some preprocessing strategies to clean the data.

nnUNet includes the following preprocessing steps:

- Cropping each data into several fixed-size patches to reduce the parameters.
- Resampling method for anisotropic data:
  In-plane with third-order spline interpolation, out-of-plane with nearest neighbor interpolation.
- Intensity normalization method:
  First, the dataset is clipped to the [0.5, 99.5] percentiles of the intensity values of the training dataset. Then a z-score normalization is applied based on the mean and standard deviation of the intensity values.

Table 1. Data splits of FLARE2021.

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

2.2. Proposed Method

U-Net like architectures enable state of the art segmentation when the pipeline is well-configured.

- Network architecture details: detail of each layer, hyper-parameters, such as strides, weights size, etc. If a standard network is used, indicate the modification:

  3D nnU-Net [1, 2] is used as shown in Figure 1, all the hyper-parameters are set as the defaulted ones.

- Loss function: we use the summation between Dice loss and cross entropy loss because it has been proved to be robust [3] in medical image segmentation tasks.

- Number of model parameters: 41268192 (can be computed via such as torchsummary library for Pytorch)

- Number of flops: 590861472000 (can be computed via such as fvcore library for Pytorch)

3. Dataset and Evaluation Metrics

3.1. Dataset

- The dataset used of FLARE2021 is adapted from MSD [4] (Liver [5], Spleen, Pancreas), NIH Pancreas [6, 7, 8], KiTS [9, 10], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [11].

- The total number of cases is 511 and we selected 60 samples as the dataset for the experiment. An approximate 70%/10%/20% train/validation/testing split is employed resulting in 42 training cases, 6 validation cases, and 12 testing cases. The information of the total data of FLARE2021 is presented in Table 1.

3.2. Evaluation Metrics

- Dice Similarity Coefficient (DSC)

- Jaccard

- Running time

- Maximum used GPU memory (when the inference is stable)

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.5 LTS
CPU	Intel(R) Core(TM) i9-7900X CPU@3.30GHz
RAM	16×4GB
GPU	GeForce RTX 2080 Ti
CUDA version	10.1
Programming language	Python3.6
Deep learning framework	Pytorch (Torch 1.8.1, torchvision 0.2.2)
Specification of dependencies	nnUNet
(Optional) code is publicly available at	FLARE21nnUNetBaseline

4. Implementation Details

4.1. Environments and requirements

The experiment was conducted on the GeForce RTX 2080Ti GPU with CUDA10.1 and nnUNet was trained in the enviroment of Pytorch1.8.1.

The environments and requirements of the baseline method is shown in Table 2.

4.2. Training protocols

The dataset contains 361 samples which are from MSD (Liver, Spleen, Pancreas), NIH Pancreas, and KiTS under their license permission. The CT set also include a hidden testing set with 50 abdomen CT cases from Nanjing University. Annotations include four organs: liver (label=1), kidney (label=2), spleen (label=3), and pancreas (label=4). We adopted nnUNet as the baseline to train the dataset.

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 [1].
Batch size	2
Patch size	80×192×160
Total epochs	50
Optimizer	Stochastic gradient descent with nesterov momentum ($\mu = 0.99$)
Initial learning rate	0.01
Learning rate decay schedule	poly learning rate policy: $(1 - epoch/1000)^{0.9}$
Stopping criteria, and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch (50).
Training time	72.5 hours
CO_2eq[1]	

The training protocols of the baseline method is shown in Table 3.

4.3. Testing protocols

In testing stage, we randomly selected 20% samples as the testing set.

- Pre-processing steps of the network inputs:
  The same strategy is applied as trainging steps.

- If using patch-based strategy, describing the patch aggregation method:
  The same patch-based strategy is applied as nnU-Net [1]. Voxels close to the center are weighted higher than those close to the border, when aggregating predictions across patches.

5. Results

5.1. Quantitative results on testing set.

Table 4. Quantitative results on validation set.

Organ	DSC (%)	Jaccard (%)
Liver	92.0±6.21	85.4±8.90
Kidney	87.3±16.4	78.2±8.90
Spleen	92.7±10.6	86.7±17.0
Pancreas	30.8±21.7	18.7±11.5

Table 4 illustrates the results on testing cases. For DSC, though the high DSC values and low dispersed distributions of the liver segmentation indicate great performance, the results degrade for other organs. For Jaccard, the obtained values and the dispersed distributions indicate unsatisfying segmentation performance for all four organs. It is worth pointing out that for liver segmentation, the DSC scores are 92.0% , indicating great segmentation performance in terms of region overlap between the ground truth and the segmented region. Jaccard values are 85.4% demonstrating that the boundary regions contain more segmentation errors, which need further improvements [11]. What's more, it is obviously that the performence of the liver segmentation is the best while nnUNet performed not really well on the pancreas segmentation.

5.2. Qualitative results

Figure 2 presents some testing examples. It can be found that the baseline method can not segment pancreas well. The first row of Figure 2 illustrates a complete liver, kidney and spleen cases except the pancreas which mask can not cover the whole pancreas well. The baseline method performed well on segmenting the spleen (blue), kidney (green) the liver (red) in the other three cases. Raising the number of epochs may boost the performence of pancreas segmenting.

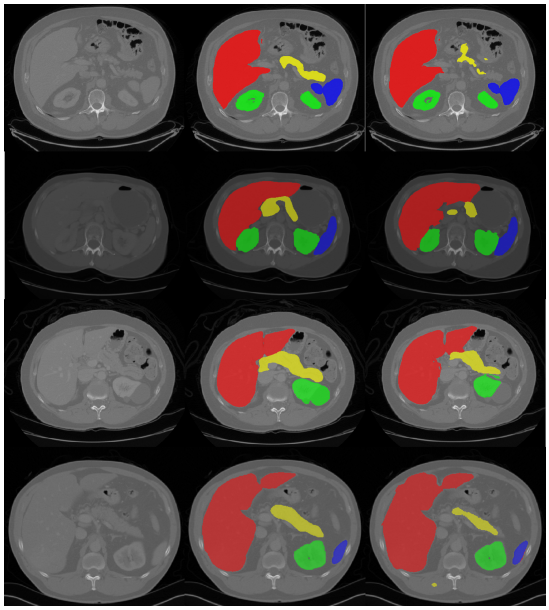

Figure 2. First column is the image, second column is the ground truth, and third column is the predicted results by our baseline method [11].

6. Discussion and Conclusion

The baseline method can work well on cases where no diseases exist. Besides, the DSC and Jaccard scores of liver and kidney segmentation is higher than the other organs, indicating the two organs maybe a comparable easier task as a result of its bigger size and consistent shape. Disappointing performance is obtained for pancreas segmentation as a result of the inter-patient anatomical variability of volume and shape.

The existence of leison is a critical factor for the segmentation performance. Although the testing samples are not contain lesion parts, it is important to figure out how to properly segment those cases. Besides, obtaining an accurate boundary segmentation need further investigate. Moreover, disappointing performance is obtained for pancreas segmentation as a result of the inter-patient anatomical variability of volume and shape.

Acknowledgment

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] 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. 1, 2, 3

[2] 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. 1, 2

[3] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, p. 102035, 2021. 2

[4] 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. 2

[5] 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)," *arXiv preprint arXiv:1901.04056*, 2019. 2

[6] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from pancreas-ct. the cancer imaging archive (2016)." 2

[7] 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. 2

[8] 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. 2

[9] 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. 2

[10] 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. 2

[11] 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?" *arXiv preprint arXiv:2010.14808*, 2020. 2, 3