U-Net with Optimal Depth and Width for Abdominal Organ Segmentation

Ziyan Huang
Shanghai Jiao Tong University
Shanghai China
ziyanhuang@sjtu.edu.cn

Zhikai Yang
Shanghai Jiao Tong University
Shanghai China
xiaoerlaigeid@sjtu.edu.cn

Abstract

No new U-Net (nnU-Net) is currently the state-of-the-art method for medical image segmentation which design task-specific pipelines for different tasks. Inspired by nnU-Net, we argue that the depth and width in U-Net are two key task-specific hyperparameters, which have a great impact on the performance and efficiency of algorithms in different segmentation tasks. In this paper, we design the task-specific depth and width of U-Net for obdominal organ segmentation task and achieve efficient performance gain. Code and model are available in https://github.com/Ziyan-Huang/FLARE21-Challenge-TeamLetsGo.

1. Introduction

Abdominal organ segmentation plays an important role in clinical practice. With the development of deep learning, various deep models have been proposed to improve the segmentation accuracy of abdominal organs. The vast majority of successful algorithms are base on the U-Net architecture. [1] proposed a framework for the task-specific pipeline design of medical image segementation, which configured itself automatically and achieved state-of-the-art results on a variety of medical image segmentation benchmarks. Inspired by nnU-Net, we argue that the depth and width in U-Net are two key task-specific hyperparameters, which have a great impact on the performance and efficiency of algorithms in different segmentation tasks. In this paper, we design the task-specific depth and width of U-Net for obdominal organ segmentation task and achieve efficient performance gain.

2. Method

Without spells and whistles, we simply adjust the depth and width of 3d nnuNet. Figure 1 illustrates the applied modified 3D nnU-Net, where a U-Net [2] architecture is adopted.

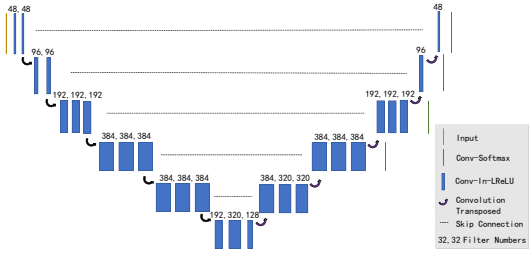

Figure 1. Network architecture

2.1. Preprocessing

We adopt the same preprocessing procudure as nnU-Net baseline. The following preprocessing steps are conducted:

- Cropping strategy: None.
- 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.

2.2. Proposed Method

- Network architecture details: The detailed architecture is shown in figure1, we only modifify the depth and width of the architecture generated by nnU-Net, other hyperparameters are the same as nnU-Net.
- Loss function: 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 [3].
- Number of model parameters: 63.07M (We computer the model parameters by torchinfo library for Pytorch)

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

- Number of flops: 1.25T (We computer the FLOPs of our model by torchinfo library for Pytorch)

2.3. Post-processing

The whole procedure of post-processing is the same as nnU-net[1]. A connected component analysis of all ground truth labels is applied on training data.

3. Dataset and Evaluation Metrics

3.1. Dataset

- A short description of the dataset used:
  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].

- Details of training / validation / testing splits:
  The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split 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

- Dice Similarity Coefficient (DSC)
- Normalized Surface Distance (NSD)
- Running time
- Maximum used GPU memory (when the inference is stable)

4. Implementation Details

4.1. Environments and requirements

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

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.3 LTS
CPU	Intel(R) Xeon(R) Platinum 8168 CPU @ 2.70GHz
RAM	1.5T
GPU	Nvidia V100
CUDA version	10.1
Programming language	Python3.7
Deep learning framework	Pytorch 1.6
Specification of dependencies	nnUNet
code is publicly available at	Code of LetsGo

4.2. Training protocols

We trained the model from scratch. The training protocols of our method is shown in Table 3.

4.3. Testing protocols

- The same pre-processing strategy is applied as traing-ing steps.

- 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 for 5-fold cross validation.

The provided results analysis is based on the 5-fold cross validation results and validation cases.

Table 4 illustrates the results of 5-fold cross validation. While high DSC and NSD scores are obtained for liver, kidney and spleen, DSC and NSD scores for pancreas indicating unsatisfactory performance.

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	1000
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 (1000).
Training time	94.5 hours
CO_2eq[1]	

5.2. Quantitative results on validation set.

Table 5 illustrates the results on validation 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 NSD, the obtained values and the dispersed distributions indicate unsatisfying segmentation performance for all four organs.

Comparison between Table 4 illustrates better performance is obtained for the 5-fold cross validation than the validation set. This phenomenon may caused by the trained model over-fitted on training set.

5.3. Qualitative results

Figure 2 presents some challenging examples.

6. Discussion and Conclusion

Our proposed method can outperform the baseline and increase acceptable computational budget.

Acknowledgment

Thanks to authors of nnU-Net and organizers of FLARE challenge.

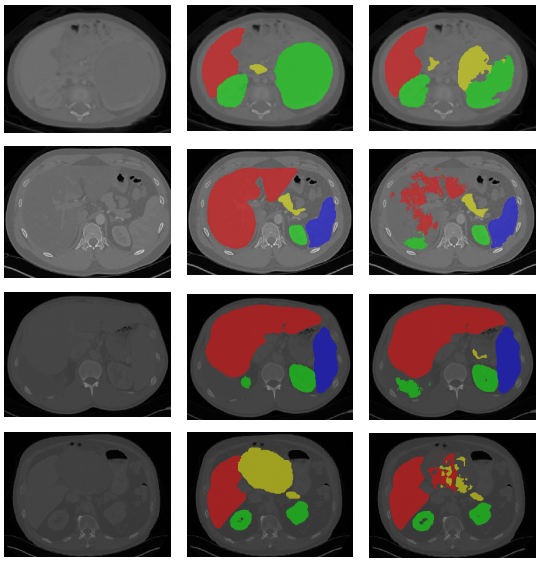

Figure 2. Challenging examples.First column is the image, second column is the ground truth, and third column is the predicted results by our proposed method [11].

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*, 2015, pp. 234–241. 1

[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. 1

[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

Table 4. Quantitative results of 5-fold cross validation in terms of DSC and NSD.

Training	Liver		Kidney		Spleen		Pancreas	
	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)
Fold-0	98.6±0.6	94.1±3.5	96.8±2.4	93.0±7.1	98.3±1.4	97.3±3.7	85.1±6.9	66.3±14.2
Fold-1	98.5±1.2	93.9±4.2	97.2±1.6	94.0±5.9	98.5±0.7	97.5±2.9	86.6±4.4	68.4±12.4
Fold-2	98.6±0.8	93.9±4.2	96.6±4.7	93.4±6.1	98.4±0.8	97.6±2.6	84.5±6.6	65.9±12.6
Fold-3	98.6±0.8	93.8±4.0	96.6±4.1	93.2±6.6	98.5±0.9	97.7±2.6	84.5±7.5	67.1±13.7
Fold-4	98.4±1.3	93.5±4.4	97.1±1.6	93.8±5.5	98.4±0.9	97.4±2.8	85.0±7.1	68.6±11.5
Average	98.5±1.0	93.8±4.1	96.9±3.2	93.5±6.3	98.4±1.0	97.5±2.9	85.1±6.7	67.2±13.0

Table 5. Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	95.0±6.38	80.3±14.8
Kidney	80.0±18.3	71.3±18.7
Spleen	90.6±16.7	83.9±19.6
Pancreas	61.7±23.0	51.5±18.8

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, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2, 3