Paper Reading No.3

nnU-Net: Breaking the Spell on Successful Medical Image Segmentation

Sheng Lian

June 2019

1 Brief Paper Intro

- ***Paper ref:*** Arxiv preprint , https://arxiv.org/abs/1904.08128 code is available on https://github.com/MIC-DKFZ/nnunet
- ***Authors:*** See Figure 1

Fabian Isensee[1,2], Jens Petersen[1,3], Simon A. A. Kohl[1⋆], Paul F. Jäger[1], and Klaus H. Maier-Hein[1]

[1] Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Faculty of Biosciences, University of Heidelberg, Heidelberg, Germany
[3] Dept. of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany

Figure 1: authors' brief intro.

- ***Paper summary:*** A medical image segmentation framework that use only naive U-Net model and can automatically adjust hyper-parameters to new dataset is proposed.
- ***Reading motivation:*** This paper is in the reading list given by Dr.Li. In the field of medical image segmentation, a large number of variants

of FCN based model are proposed for different subtasks. Many of them fail to deal with new datasets. For me, I've done some research based on U-Net. I tried to introduce residual connections, dense connections, attention mechanism, etc., to U-Net for solving different medical image segmentation tasks. For different datasets, we need to add different components to the baseline model, and use different hyperparameters. This is normal for the computer vision community, but it is also unreasonable. It's disappointed that our so-called 'AI' is not intelligent with such weak generalization performance. This paper aims to solve this problem.

2 Backgrounds

Semantic segmentation can be regarded as the most important task in the domain of medical image analysis. This is very challenging due to the diversity and individual peculiar of imaging datasets. All these issues make it more diffcult to generalize findings from one task to others. In the era of deep learning, researchers proposed so many FCN-based, especially UNet-based architectures. So many modifications, included residual connections, attention mechanism, feature recalibration, etc., have been introduced. But, many new design concepts did not improve, or sometimes even worsened the performance of a well-designed baseline.

3 Methods

This part is more like a bag of training schemes and tricks for different medical image segmentation datasets. The segment algorithm can be formalized as a function $f_\theta(x) = \hat{y}$, with x being the image, $\hat{y}$ the predicted segmentation, and θ the set of hyperparameters for training. Former publications usually focus one the best choice of θ for one task, and this scheme usually have poor generalization performance, as is introduced above. Here, this paper aims to seek for a function $g(X, Y) = \theta$, that is, making our model, namely 'nnU-Net', adapts itself without user interaction to previously unseen datasets.

3.1 preprocessing

3.1.1 Image Normalization

Given the modality of input data, nnU-Net decides the way of image normalization, which goes as follows:

- **CT:** all foreground voxels in the training set are collected and an automated level-window-like clipping of intensity values is performed based on the 0.5 and 99.5th percentile of these values. The data is then normalized with the global foreground mean and standard deviation.

- **not CT:** normalizes intensity values by subtracting the mean and dividing by the standard deviation.

3.1.2 Voxel spacing

nnU-Net collects all spacings within the training data and for each axis chooses the median as the target spacing. All training cases are then resampled with third order spline interpolation.

3.2 Training procedure

- A 2D U-Net, a 3D U-Net and a cascade of two 3D U-Net models are configured. There are limited changes towards the original U-Net, including the usage of padded convocations and Leaky ReLU.

- nnU-Net automatically sets the batch size, patch size and number of pooling operations for each axis while keeping the memory consumption within a certain budget.

- All U-Net architectures are trained in a five-fold cross-validation.

- The sum of the cross-entropy loss and the dice loss are used as loss function.

- Choose Adam as optimizer. Whenever the exponential moving average of the training loss does not improve within the last 30 epochs the learning rate is dropped by a factor of 0:2. Training is stopped when the learning rate drops below 10^{-6} or 1000 epochs are exceeded.

- A series of data augmentation operations are adopted.

3.3 Experiments

Authors implemented nnU-Net on the Medical Segmentation Decathlon challenge [1]. In this phase I of this challenge, researchers can develop model on seven training datasets, including medical images of various parts of the body. And in phase II, 3 unknown datasets are given for measuring the models' performance.

In the first and second phases (see Figure 2 and 3), nnU-Net has obvious advantages in the decathlon challenge.

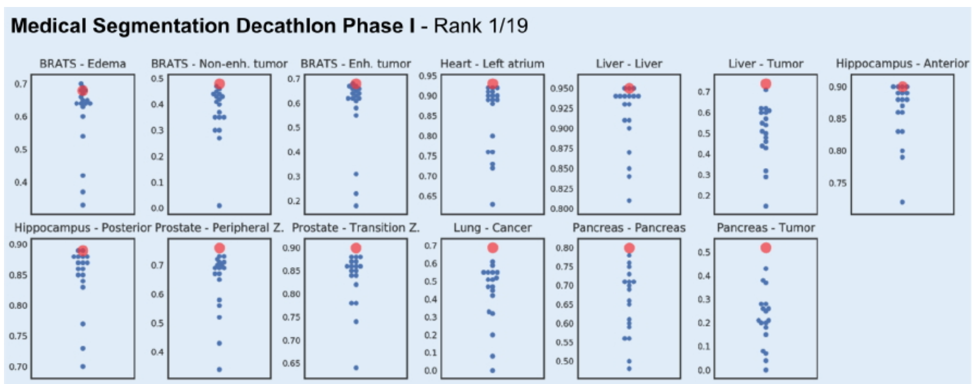

Figure 2: Summary of nnU-Net performance on phase I.

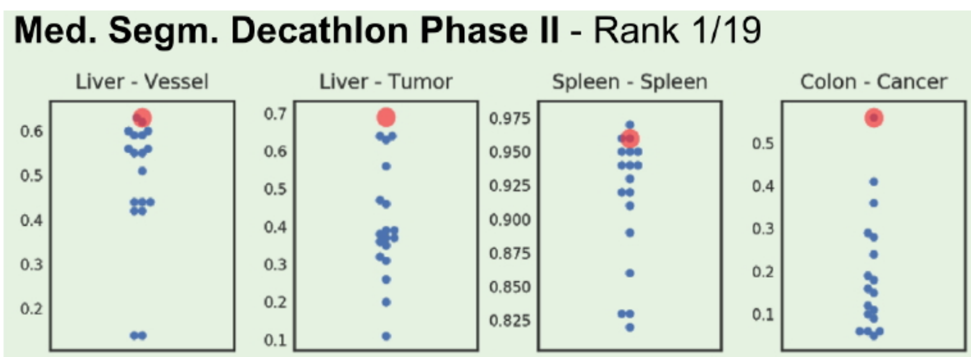

Figure 3: Summary of nnU-Net performance on phase II.

Ablation studies on the design choices of nnU-Net have been discussed in Figure 4.

	BraTS	Liver lowres	Liver fullres	Hippocampus	Prostate	Lung nodule	Pancreas
Vanilla nnU-Net	0.72	0.79	0.78	0.89	0.77	0.65	0.65
Batch norm instead of Inst. norm	1.0%	-0.1%	2.9%	-0.1%	-1.3%	-14.2%	-3.7%
No feature map normalization	1.1%	-4.6%	-22.8%	-0.2%	-4.2%	3.0%	-100.0%
ReLU instead of LeakyReLU	0.6%	0.0%	1.0%	-0.1%	-0.2%	-0.4%	0.5%
No data augmentation	-0.8%	-4.9%	1.5%	-1.5%	-0.4%	4.2%	-11.3%
Only cross-entropy loss	-0.6%	-12.0%	-6.3%	0.0%	-1.4%	-25.4%	-8.8%
Only dice loss	0.9%	-2.5%	-10.1%	-0.3%	-3.0%	-11.5%	1.6%

Figure 4: Ablation studies on the design choices of nnU-Net. Experiments were done on representative datasets from the Medical Segmentation Decathlon using one split of the training data and a 3D U-Net. Numerical values for nnU-Net represent the average foreground Dice scores (i.e. mean between liver and tumor dice for the liver dataset), values for the ablation studies represent the percentage-wise change in Dice score.

4 My thoughts

1) The starting point and motivation of this paper is good and rather innovative. However, in my opinion, the goal of this article is too idealistic. It's true that our so called 'AI' models are still so weak, especially with poor data generalization performance. But, many subtasks of medical image segmentation are vitally important, and the diversity and individual peculiarities are far beyond our imagination. This research may be able to try in other tasks, but in the field of medical imaging, any small error, any miss of small lesion, may cause severe results. So, we still need to fine-tune the models for specific tasks for best performance.

2) But, still, this research is a nice try, and, nnU-Net could be considered as the strongest U-Net baseline to date (perform well in various tasks). We can learn from the experiments. However, according to Figure 4, the proposed hyperparameters setting of nnU-Net is still far from global optimum. Many improvements are still only applicable to certain datasets, which indicates the need to redesign the network for each subtask, especially the life-critical task of medical image segmentation.

3) post-processing is not considered in this paper, which is more relative to the kind of image dataset.

4) Maybe for the medical segmentation decathlon task (or a similar challenge), we need to see something deeper from the data distribution.

References

[1] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.