Data source invariant segmentation of abdominal organs.

Talha Anwar
chtalhaanwar@gmail.com

Abstract

The segmentation of the abdominal organs plays a vital role in the radiological evaluation of abdominal diseases. Most studies focus on the segmentation of single organ segmentation or data taken from single sources in a single phase, making the deep learning procedure less predictive at unknown data. I participated in the Flare21 competition and tried to resolve this issue. SegResNet is used as a segmentation model, and average dice similarity coefficient of 65% is achieved.

1. Introduction

Abdominal organ segmentation provides detailed analysis such as structural analysis of different abdominal organs helping in diagnosis, surgical planning, and making a clinical decision. Several researchers have used segmentation techniques, but their approach is limited to a single organ or data collected from a single source. Zhou used Faster RCNN to locate live region; then Gaussian mixture model-based signed distance function is used to increase the flexibility of shape. They used three different datasets, but they trained on one organ only [1]. Medical Image Computing and Computer-Assisted Interventions society (MICCAI) organized a competition named Fast and Low GPU Memory Abdominal Organ Segmentation (FLARE) with a goal to segment abdominal organs *i.e* liver, kidney, spleen and pancreas from 3D CT scan. The CT scans are taken from different medical centers from different CT scan machines in different phases, making it difficult to segment. The challenge is to test the system on a low GPU system. Jun Ma *et.al* combined different existing dataset and proposed a generalized model which is reliable for diverse data testing [2]. Liu *et.al* used a weakly supervised learning technique to train an accurate segmentation model. First, they generate pseudo-masks using KMeans clustering and then train a 3D U-Net. They achieved an accuracy of 95.19%, 92.11% and 94.15% for liver, spleen and kidney segmentation [3]. I used SegResNet [4] to develop an independent data source approach that can segment multiple organs in different CT scan phases.

2. Method

2.1. Dataset

The CT scan 3D images comprised of four classes *i.e.* liver, kidney, spleen and pancreas. The dataset is multi-vendor, multi-center, multi-phase, and multi-disease based which make it more generalized toward the practical usage. Dataset has 511 CT scan volumetric images collected from 12 medical centers. The dataset used of FLARE2021 is adapted from MSD [5] (Liver [6], Spleen, Pancreas), NIH Pancreas [7, 8, 9], KiTS [10, 11], and Nanjing University under the license permission. More details about dataset is available on the competition website and baseline paper[2]. There are 361, 50 and 100 3D scans in training, validation and testing. Table 1 shows the number of images collected in different phases from different medical centers.

2.2. Proposed Method

For segmentation of the data, MONAI framework is used [12]. Before feeding, data to deep learning architecture (UNET) is pre-processed, and output from architecture is post-processed.

2.3. UNET

SegResNet architecture is used as the backbone UNET for the segmentation task because it has few parameters. This architecture has approximately 4.7 M parameters.

2.4. Pre-processing

Because of the large number of parameters, units are pruned to over-fitting, so I used more transformation in train data than validation data; this makes train data less prone to over-fitting.

- **Spacing** As 3D scans are from different medical centers, so re-sample all images to the same spacing. Spacing of (1.0, 1.0, 1.0) is used for x, y, and z dimensions, respectively.
- **Orientation** As images are collected in different phases, I changed their orientation to RAS orientation which is (Left, Right), (Posterior, Anterior), (Inferior, Superior) orientation.

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

- **Scale Intensity** 3D scans are scaled between 0 and 1
- **Crop Foreground** Foreground is cropped based on value 0; if a portion of the image has 0 value, it means there are no organs in that area, so that area is removed.
- **Random patch cropping** Images come with x and y dimensions of 512 and 512, and their z dimensions range from 50 to 1000 approximately. So I cropped 4 patches of 128,128,64. During validation, images are not cropped; instead, whole images are fed and model learning the whole images in patches.
- **Axis Flipping** Images are flipped on all three axes randomly.

All these above transformations are applied to training data; for validation data, only spacing, orientation, scale intensity, crop foreground is applied.

2.5. Post processing

In post-processing, images are returned backed to their original orientation, spacing, and size.

2.6. Training protocol

I have used Tesla P100 with 16 GB ram. Pytorch lightning framework V.1.4.0 is used to run MONAI framework. I used gradient accumulation and Stochastic Weight Averaging (SWA) techniques. Gradient accumulation is a technique to train the model with larger batch sizes. It updates the network weights after some batches instead of every batch. It accumulates the gradient of some batches and then back-propagates them. SWA improves the generalization of the model with no additional cost. I have also trained the model in mixed precision (float 16) mode, which increases the performance by 2x to 8x [13]. Table 3 shows the system used to train the model.

Dice cross-entropy is used as a loss function. AdamW optimizer with 0.001 learning rate is used. Cosine Annealing Warm Restart is used for learning rate schedules. 300 epochs are trained with a batch size of 4. During validation sliding window interface is used with a batch size of 1. Table 3 shows the training protocol.

Table 2. Training protocols.

Data augmentation methods	Random Axis flipping, scaling, orientation, spacing, random crop by positive and negative label, random rotate 90
Patch sampling strategy	Random cropping of patches on the base of positive and negative label.
Batch size	4
Patch size	128×128×64
Total epochs	300
Optimizer	AdamW
Initial learning rate	0.001
Learning rate decay schedule	Cosine annealing rate

Table 3. Environments and requirements.

Ubuntu version	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) CPU @ 2.00GHz
RAM	12 GB
GPU	Nvidia P100
CUDA version	10.1
Programming language	Python3.7.11
Deep learning framework	Pytorch 1.9.0 and Pytorch-lightning 1.4.0
Specification of dependencies	moani-weekly 0.7.dev2130
Code is publicly available at	Github Repo

2.7. Evaluation Metrics

The official Flare21 competition metrics comprised of following

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

Table 4. Results showing average and individual organ DSC and NSD

	DSC		NSD
Mean	65.30	Mean	50.70
Liver	87.92±14.31	Liver	64.32±15.93
Kidney	61.94± 31.30	Kidney	52.79±28.30
Spleen	59.01± 38.09	Spleen	51.37±30.93
Pancreas	52.31± 27.65	Pancreas	34.31±19.92

This paper showed DSC and NSD on validation data.

3. Result

Table 4 shows the mean DSC and NSD and the score for individual subjects. Average DSC value of 65% is achieved.

Figure 1 shows the dice similarity coefficient and loss for the validation data which is created by splitting train data into 80:20.

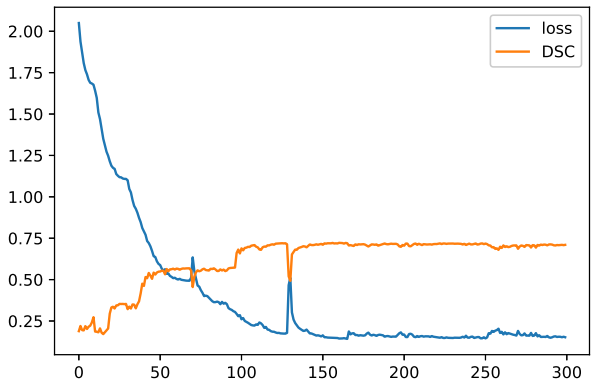

Figure 1. Validation loss and DSC

Acknowledgment

I would like to thanks the maintainers of MONAI and organizers of Flare21 challenge. I also declared that the segmentation method implemented for participation in the FLARE challenge had not used any pre-trained models nor additional datasets other than those provided by the organizers.

References

[1] L. Zhou, L. Wang, W. Li, B. Lei, J. Mi, and W. Yang, "Multi-stage liver segmentation in ct scans using gaussian pseudo variance level set," *IEEE Access*, 2021. 1

[2] 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. 1

[3] Y. Liu, Q. Hui, Z. Peng, S. Gong, and D. Kong, "Automatic ct segmentation from bounding box annotations using convolutional neural networks," *arXiv preprint arXiv:2105.14314*, 2021. 1

[4] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 311–320. 1

[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. 1

[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)," *arXiv preprint arXiv:1901.04056*, 2019. 1

[7] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from pancreas-ct. the cancer imaging archive (2016)." 1

[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. 1

[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. 1

[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. 1

[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. 1

[12] N. Ma, W. Li, R. Brown, Y. Wang, Behrooz, B. Gorman, H. Johnson, I. Yang, E. Kerfoot, charliebudd, M. Adil, Y. Li, Y.-T. Hsieh, A. Aggarwal, masadcv, C. Trentz, adam aji, myron, M. Graham, B. Murray, G. Daroach, P.-D. Tudosiu, M. McCormick, A. Hatamizadeh, Ambros, Balamurali, C. Baker, H. Roth, and J. Sellner, "Project-monai/monai: 0.6.0," Jul. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5083813 1

[13] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, "Mixed precision training," *arXiv preprint arXiv:1710.03740*, 2017. 2