Fast and Low GPU Memory Abdominal organ segmentation

YingteHe [1], GongningLuo [2]

Harbin Institution of technology

Abstract. *Abdominal organ segmentation plays an significant rolein clinical practice. Most solutions only focus on the accuracy of the segmentation, ignoring inference time andGPU memory, making it difficult to be taken into practice inclinic. In our method, we use a light-weight 3Dunet to detect the abdominal region, and then apply a 3D DenseVnet to segment each organ based on the ROI. Eventaully,we achieved average dice coefficient of 0.75,and took up nearly 2G GPU memory, and it takes nearly 30 seconds to inference one sample on average. To some extent, we achieved a tradeoff in accuracy, memory and inference time.*

Keywords: Abdominal organ, Segmentation,DenseVnet

1. Introduction

Abdominal organ segmentation is a popular task which has significant clinical value. Many solutions have already achieved good performence in most organs such as liver, except tubular organs like pancreas. However, most solutions either takes plenty of time when inference, or takes up lots of GPU memory. It is still a big challenge to segment abdominal organs accurately and efficiently. Firstly, we adopted 3D convlutional network for better accuracy compared to 2D network, which didn't take global information into consideration. Secondly, 3D convlutional network takes up lots of memory limiting the input size of data. We adopted 3D Densenet, which is memory efficient on the basis of high accuracy. Lastly, thanks to densenet, we don't need to crop data when training, instead, we reisize input data into a particular size, this could take less time when inference compared to the crop way.

2. Method

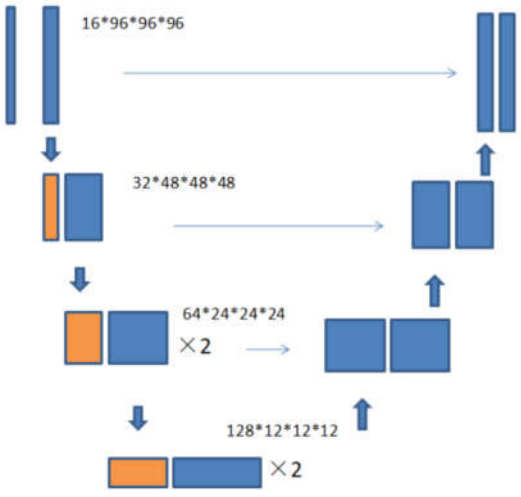


Figure 1. ROI detection network

2.1. ROI Detection

In some cases, abdominal cavity only occupies partial region in CT volume, it is worth taking seconds to crop abdominal region for further organ segmentation, which is helpful for less computation resources and more accuracy segmentation. We consider it more suitable to take it as a segmentation task than a detection task when cropping abdominal cavity, thus we use the label of all organs as foreground and process the the task as semantic segmentation. As Figure 1 shows, we use a light-weight 3DUnet with fewer parameters to predict the abdominal region, finding the boundary and then expand some pixels to get the ROI. First, we resize the input data into 96*96*96 as the input of network,and finally resize back to its original size.

2.2. Organ Segmentation

All the preprocesses end up with abdominal volume datas with different size. Considering training cropped data didn't taking global information into consideration, and takes more time when inference, we resize the input data into a particular size. By analysing the training dataset, we resize the input data into (96,128,192) to keep the scale of each dimension. Traditional 3DUnet is composed of deep convlutional layers and takes up a lot of GPU memory, it may not be suitable for the flare21 challenge even if it may achieve the best accuracy.Instead, we applied the DenseVet[1] to this task,as Figure 2 shown. The densevnet structure comprises three sequences of dense block connected by downsampling dilate convlution, The output of each dense block is followed by upsample layer which uses bilinear upsampling, the results of upsample layers are added together and then resized to the orignal size. Dense block is a good tool for feature fusion which is an alternative of convlutional layer with lots of channels. Beacuse the number of pixels differs in each orgin, we applied Focal Loss to train the DensenVnet in order to solve the label imbalance problem.The number of training parameters in the network is 871922 and the flops is 20339341715

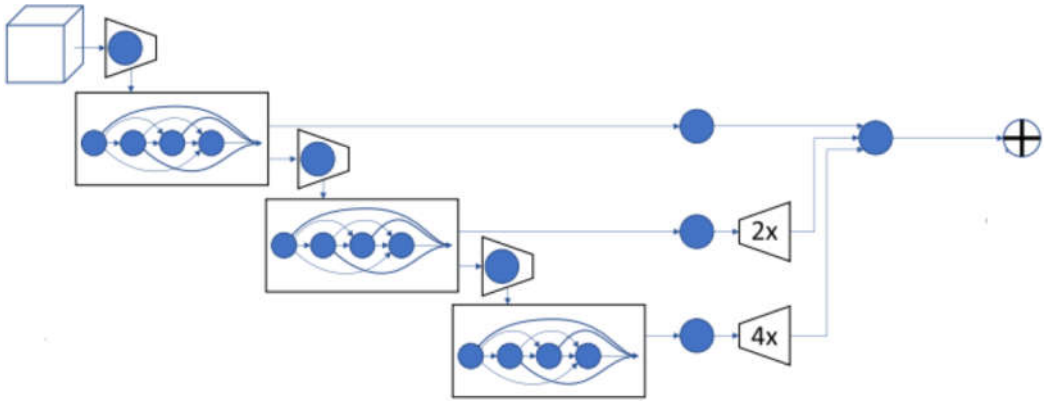


Figure 2. organ segmentation network

2.3. Post-processing.

We apply the connected domain analysis on the results of segmentation network.

3. Dataset and EvaluationMetrics

3.1. Dataset

- A short description of the datasetused:
  The dataset used of FLARE2021 is adapted fromMSD [2] (Liver [3], Spleen, Pancreas), NIH Pancreas [4,5,6], KiTS [7, 8], and Nanjing University
  underthelicensepermission.Formoredetailinformationofthedataset,pleaserefertothechallengewebsite and[9].

- Details of training / validation / testingsplits:
  The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split is employedresultingin361trainingcases,50validation cases, and 100 testing cases. The detail information is presented in Table 1.

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

3.2. EvaluationMetrics

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

4. ImplementationDetails

4.1. Environments andrequirements

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

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 16.04.7 LTS
CPU	Intel(R) Core(TM) i7-6800k CPU@3.40GHz
RAM	62GB
GPU	Nvidia TITAN X
CUDA version	10
Programming language	Python3.6
Deep learning framework	Tensorflow = 1.12.0

4.2. Trainingprotocols

Full description of the training protocols, including but not limited to the items illustrated in Table 3.
The training protocols of the baseline method is shown in Table 3.

Table 3. Training protocols.

Data augmentation methods	Rotations, scaling, Gaussian noise.
Initialization of the network	“he” normal initialization
Patch sampling strategy	No patch sampling
Batch size	1

Patch size	96×128×192
Total epochs	600
Optimizer	Stochastic gradient descent
Initial learning rate	0.01
Learning rate decay schedule	poly learning rate policy:(1 - epoch/600)0.9
Stopping criteria,and optimal model selection criteria	Stopping criterion is reaching the maximum number of epoch (600).
Training time	72 hours

4.3. Testing protocols

- Pre-processing steps of the network inputs: The same strategy is applied as training steps.
- Post-processing steps of the network outputs: connected domain analysis.
- Patch-based strategy: None.

5. Results

5.1. Quantitative results on validationset.

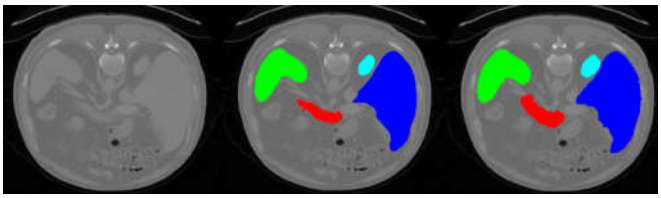
Table 5 illustrates the results on validation cases. The results of livers shows good performance, and pancreas has the worst performance. Obviously, organ with large volume and regular shape tends to have better segmentation result. The NSD shows bad performance as we can see in the figure. The reason maybe our segmentation model is trained by focal loss ,which only consider the pixel-wise classification, which ignored the boundary information. And our method used downsampling and upsampling , which caused the jagged boundary leading to worse NSD..Overall, compared to the result of training data, the worse performance on validation data may be caused by overfitting or different data distribution.

Table 5. Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	90.5±9.4	58.6.0±13.5
Kidney	73.6±23	58.4±19
Spleen	84.7±18.4	62.4±18.1
Pancreas	52.4±21.1	35.7±13.1

5.2. Qualitative results

Figure 4 presents some examples..the first two rows are the easy examples.Obviously,the result of segmentation is very similar to the ground truth. But the last two rows shows the hard examples.It failed to distinguish between liver and kidney, sometimes it wrongly recognize kidney as liver.And in the hard examples, it is very difficult to segment pancreas.



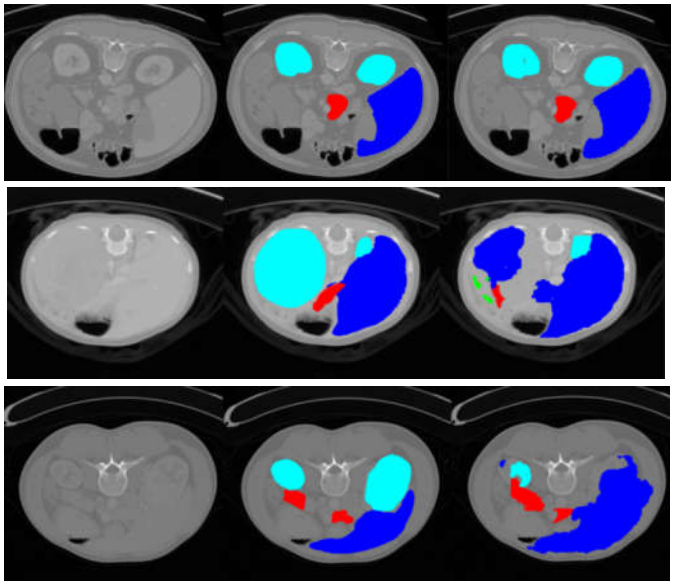


Figure4. First column is the image,second column is the ground truth, and third column is the predicted results by our baseline method[11].

6. Discussion and Conclusion

The method can work well on cases where the scan is high quality and has high degree of differentiation between organs. Besides, the DSC and NSD scores of liver segmentation is higher than the other organs, indicating liver 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.

We paied more attention on the less inference time and low gpu memory, so our method used less training parameter and downsampling the input data. As a result, our method achieved good performance on easy case, but bad performance on hard cases. Bad cases and organs with anatomical variability of volume and shape tends to need more training parameters to learn the feature, so our method didn't work well in these conditions.

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] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal ct with dense v-networks," *IEEE Transactions on Medical Imaging*, pp. 1822–1834, 2018. 2

[2] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani,B. Van Ginneken,A.Kopp-Schneider,B.A.Landman,G.Litjens,B.Menze*etal.*,"Alargeannotatedmedicalimage dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.2

[3] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H.Chen,Q.Dou,C.-W. Fu,X.Han,P.-A. Heng,J.Hesser*etal.*,"The liver tumor segmentation benchmark (lits)," *arXivpreprint arXiv:1901.04056*, 2019.2

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

[5] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey,andR.M.Summers,"Deeporgan:Multi-leveldeepconvolutional networks for automated pancreas segmentation," in*Internationalconferenceonmedicalimagecomputingand computer-assisted intervention*. Springer, 2015, pp.556–564. 2

[6] 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

[7] 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

theartinkidneyandkidneytumorsegmentationincontrastenhancedct imaging: Results of the kits19 challenge,"*Medical Image Analysis*, vol. 67, p. 101821, 2021.2

[8] 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

[9] 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. 2,3,4