A 3D ResNet with multi-scale block and shape constrains for MICCAI FLARE21 Challenge

Yinan Xu[1], Siying Cao[1], Wenhao Dong[1], Nuo Tong[1]
[1] Xidian University, Xi'an, Shaanxi 710071, China
ynxu_1@stu.xidian.edu.cn

Abstract. Abdominal multi-organ segmentation is very important for preoperative planning, At present, there are many open abdominal data sets, and many more accurate methods have been proposed. However, the problem of polycentric, spatio-temporal efficiency remains unsolved. In order to solve the problem of inconsistency of multi-center data, we have done a lot of work in the preprocessing, including adjusting window level and window width, cutting and resize image to the same size, and we used data augmentation to expand the data tenfold. In order to improve the spatio-temporal efficiency, we adopted a lightweight network based on 3D-ResNet (ISE-Net), and added Inception and squeeze-and-excitation (SE) block to obtain multi-scale features, which reduce the differences in resolution of the polycentric data. And Shape constraints are added to the proposed method in order to mitigate the effects of liver and pancreatic disease on its segmentation. In post-processing, small connected regions are removed to solve the problem of missegmentation. It is important to note that we choose to only use CPU in the inference phase, which reduces the hardware requirements.

Keywords: multi-organ segmentation, ResNet, multi-scale features, shape constrains

1. Introduction

Accurate contouring of target organs is essential in applications including disease diagnosis, radiotherapy treatment planning, and delivery [1]. However, abdominal organ segmentation on CT images is a demanding task. Manual organ delineation is not only tedious but also suffers from substantial intra- and inter-observer variabilities [2]. An accurate and robust automated segmentation technique would be highly desirable to replace or augment the manual process. However, the accuracy of automated segmentation is limited by the morphological complexities of abdominal organs, the large inter-subject variations, and the low CT soft-tissue contrast. Moreover, artifacts from respiratory andperistaltic motion further blur the boundaries of many abdominal organs[3].

Many different CNNs have been proposed and have achieved great success. However, they usually contain large number of parameters which leads to the difficulties of not being clinically useful. And the training data and test data are usually from the same medical center.

In order to meet the needs of fast speed and low memory occupancy, We design a lightweight network based on the 3D ResNet [4] structure and use Inception and SE block [5, 6] to adapt to different resolution of different images. This network has only one tenth of the parameters of baseline 3D nnUnet. In order to adapt the model to multi-center data, we have done a lot of data preprocessing. For example, adjust the window position and width of the CT image, cut off the slices without target organ, etc. At the same time, our network adopts a multi-scale mechanism, which can also help overcome the problem of large data resolution gap. And in order to mitigate the effects of liver and pancreatic disease on its segmentation, shape constraints are added to the proposed method.

2. Method

Figure 1 illustrates the applied 3D ResNet with Inception and SE block (ISE Block), where a U-Net architecture is adopted.

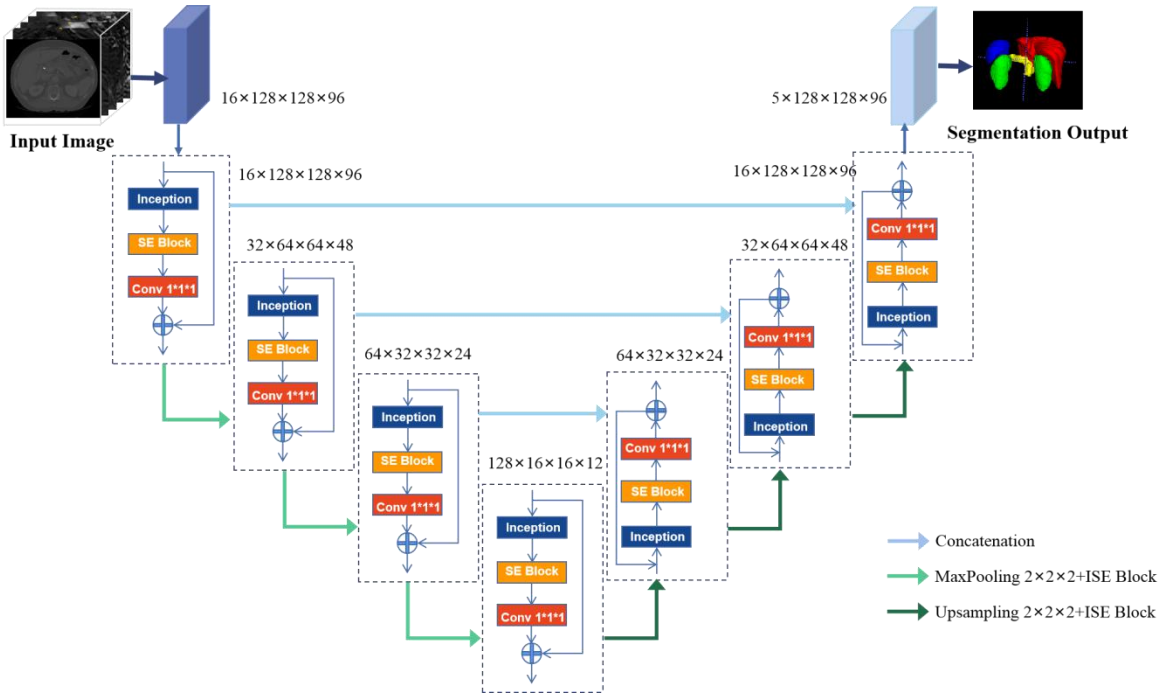

Figure 1: The overall architecture of the proposed ISE-Net.

2.1 Inception and SE block (ISE Block)

The ISE block for this work is shown in Figure 2. It adopts the structure of residual block, and adds Inception block and SE block.

The Inception block adopts convolution kernels of different sizes and fuses features, so it can have receptive fields of different sizes to alleviate the resolution difference of polycentric data. The Inception block for this work is shown in Figure 3.

Although abundant features are obtained, a deal of redundant features are collected, which weaken the important and target-related features and can reduce the discriminability of the network. Thus, a squeeze-and-excitation block is employed as the SE block here to recalibrate the importance of the features obtained by the Inception block. The specific components and structure of the SE block is illustrated in Figure 4. In the SE block, a global average pooling layer is used to aggregate the global information, which is followed by two full connection layers to capture the channel-wise relationships. Then, the features obtained by the Inception block is recalibrated by the channel-wise relationships through point-wise multiply operation.

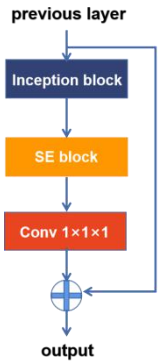

Figure 2: The detailed architecture of the ISE block.

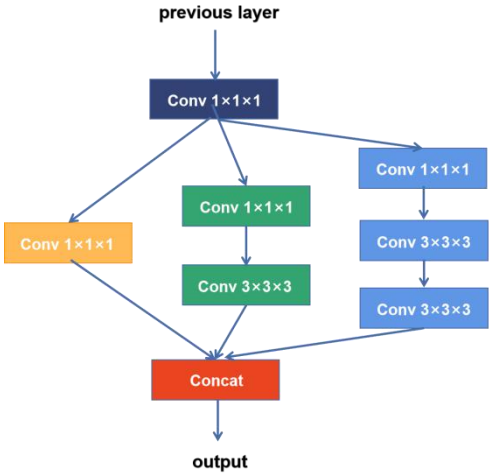

Figure 3: The detailed architecture of the Inception block.

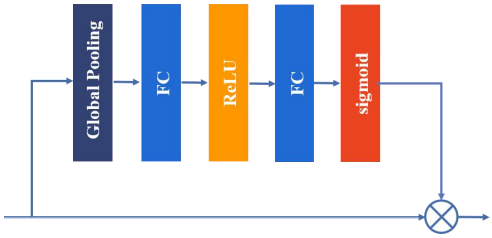

Figure 4: The detailed architecture of the SE block.

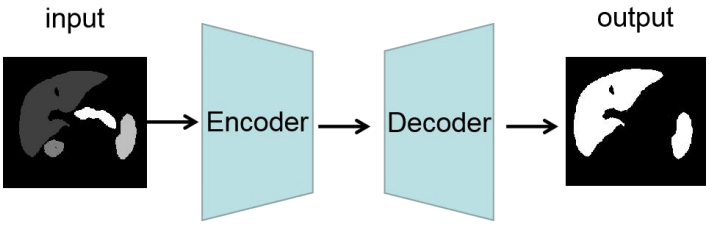

Figure 5: The detailed architecture of the Shape constraints model.

2.2 Shape constraints model

The proposed method first trains a shape constrained model, which is a binary segmentation network. The input of the network is the segmented label of multiple abdominal organs, and the output is the binary segmented result of liver and pancreas. You can think of the network as learning the shape of the liver and the pancreas. The architecture of the network is shown in Figure 5.

2.3 The combination of shape constraints model and ISE-Net

The shape constraints are added as regular terms in the loss function, which is composed of two types of loss functions, that is, segmentation loss and shape constraints loss. The loss function is shown in Eq.(1).

$$Loss = DiceLoss(P,T) + 0.01 \times MSELoss(E(p),E(T)) \quad (1)$$

Where, P is the prediction result of ISE-Net, T is the true label of the input image, and E represents the encoder of the shape constraint model. Diceloss is a commonly used segmentation loss, which can solve the problem of imbalance between positive and negative samples, and MSELoss is the mean square error loss function.

2.4 Preprocessing

The proposed method includes the following preprocessing steps:

• Cropping strategy:

The range of CT scans varies greatly from patient to patient, some people's CT scans include the entire chest and lower abdomen, while others only include the target organs. Therefore, it is necessary to drop out some useless slices. For the training data, we crop images according to the labels. For unknown validation data, we trained a very lightweight binary segmentation network, which can roughly locate the four target organs for cropping.

• Adjust window level and window width:

The difference between the maximum and minimum values of CT HU values of different patients is relatively large, but the HU values of the organs we pay attention to are in the same range. Therefore, we need to adjust the window level and window width to make the contrast of the organs of concern higher.We set the window level to 40 and the window width to 400.

• Resize method for anisotropic data:

Drawbacks of such patch-based segmentation methods are redundant computation and inability to learn global Features, so we input the whole picture. To make the network lightweight, we resize each patient's image to $128 \times 128 \times 96$. And In-plane with third-order spline interpolation, out-of plane with nearest neighbor interpolation.

• Intensity normalization method:

A z-score normalization is applied based on the mean ans standard deviation of the intensity values.

• Data augment method:

First we selected a few difficult kidney tumor samples and individually expanded them. And we use CLAHE, a method of data contrast enhancement, as a way of data expansion , and we use random rotations, flips, and scales to further expand the data, finally, the number of data increased tenfold.

2.5 Proposed Method

• Network architecture details: The network architecture is shown in Figure 1.

• Loss function: the loss function is described in Section 2.3.

• Number of the parameters of ISE-Net: 4326525

• Number of flops:57632408

2.6 Post-Processing

Post-processing operations include removing small connected areas and filling the holes to solve the missegmentation problem.

3. Dataset and Evaluation Metrics

3.1 Dataset

• A short description of the dataset used:

The dataset used of FLARE2021 is adapted from MSD [7] (Liver [8], Spleen, Pancreas), NIH Pancreas [9, 10, 11], KiTS [12, 13], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [14].

• 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.

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

Table 2:Environments and requirements

Windows version	Win10
CPU	Intel(R) Core(TM) i7-6900K CPU @3.20GHz
RAM	64.0GB
GPU	NVIDIA GeForce GTX 1080 Ti (11GB)
CUDA version	10.1
Programming language	Python 3.7
Specification of dependencies	None
Deep learning framework	Tensorflow-gpu 2.2.0 Keras 2.3.1

Table 3:Training protocols

Data augmentation methods	Rotations, scaling and filp
Initialization of the network	“he” normal initialization
Patch sampling strategy	Patch sampling strategy is not be used in our method, the input is the entire image
Batch size	1
Input size	128×128×96
maximum epochs	100
Optimizer	Adam
Initial learning rate	0.001
Learning rate decay schedule	The learning rate is reduced by 0.5 times per five epochs
Training time	48h
EarlyStop	If val-loss not fall after five epochs, the training is stopped

3.2 Evaluation metrics

- Dice Similarity Coeffificient (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 the proposed method is shown in Table2, and the CPU can be selected in inference process.

4.2 Training protocols

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

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: removing small connected areas to solve the missegmentation problem. And resize the predicted results back to the original size of the predicted image.

5. Result

5.1 Quantitative results on validation set.

Table 4 illustrates the results on validation cases. Liver has a large volume, so its segmentation DCS value is relatively high, and the variance is relatively small. However, the poor performance of its NSD value indicates that the organ boundary segmentation is not good. Some of the data in the validation set had large renal tumors, which made segmentation difficult. By changing the HU distribution of the few renal tumors in the training set and separately expanding the data, the segmented DSC of the kidney was improved. The pancreas is a very small organ, which is difficult to divide, and the segmentation DSC value is unsatisfactory.

Table 4. Quantitative results on validation set.

Organ	DSC (%)	NSD (%)
Liver	94.5±2.77	66.2±12.8
Kidney	91.0±6.99	72.4±10.9
Spleen	89.1±17.6	70.1±16.7
Pancreas	59.1±23.3	42.2±19.0

5.2 Qualitative results

Figure 6 presents some easy examples. It can be seen that the proposed network shows great performance.

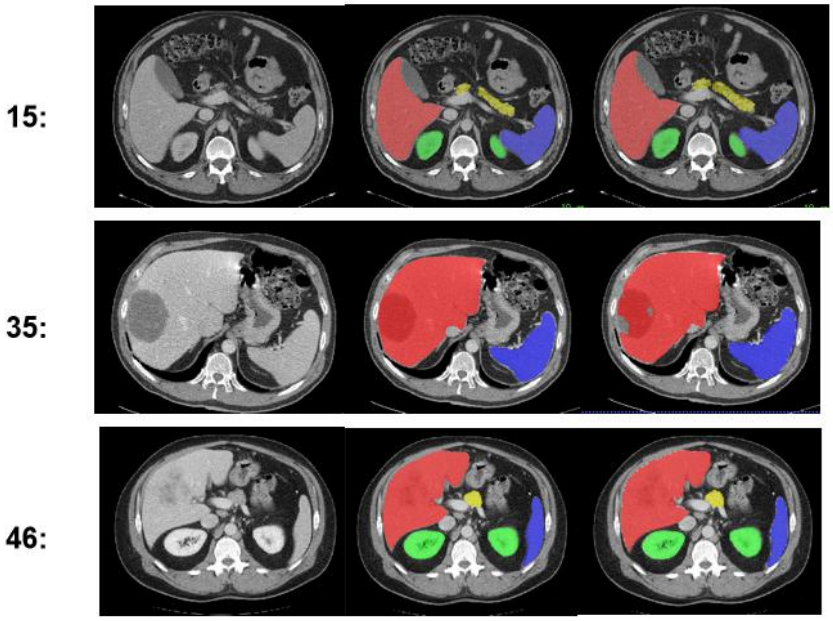

Figure6. Easy examples. First column is the image, second column is the ground truth, and third column is the predicted results by the proposed method.

Figure 7 presents some challenging examples. The first row of Figure 7 illustrates a fatty liver case where the liver is darker than healthy ones. The baseline method fails to segment the spleen (blue) and the liver (red) in this case in the template, but our proposed method can segment them. Second row of Figure 7 shows an example with abdominal effusion which causes incorrect segmentation (green). Third row of Figure 7 shows an example of a large pancreatic tumor, our proposed method misclassifies it into liver tissue. The example in the fourth row contains a kidney tumor that our method can identify.

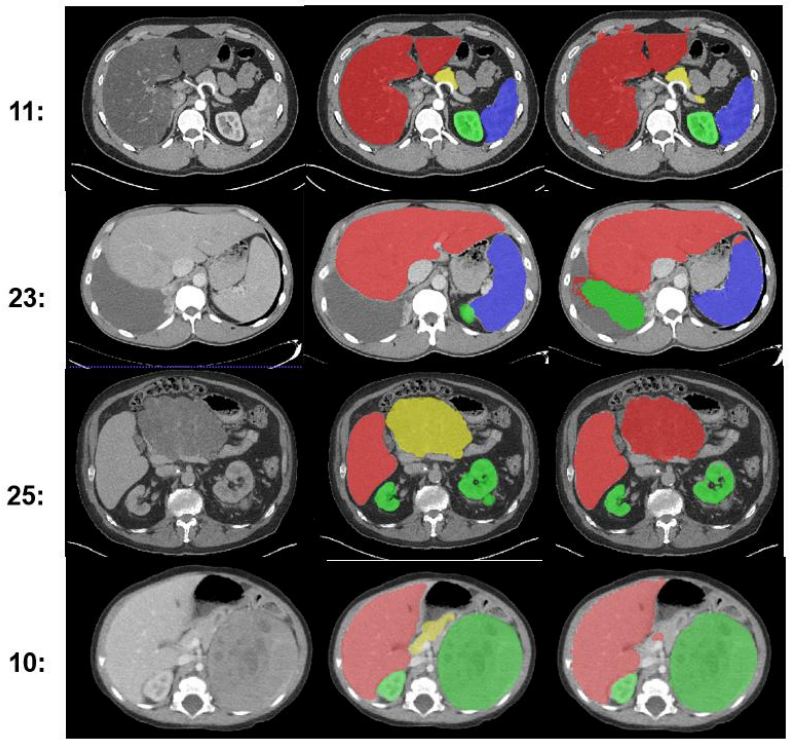

Figure7.Challenge examples. First column is the image, second column is the ground truth, and

third column is the predicted results by the proposed method.

6. Discussion and Conclusion

Generally, the proposed method can be better applied to simple cases where there is no disease or disease is not serious. When patients have liver disease that makes the liver darker than normal, our method is also an improvement over the baseline method. At the same time, our method can effectively segment large kidney tumors. And it is important to note that our method can use the CPU to test, which can better meet the needs of clinical applications. However, the proposed method has some limitations. Because the pancreas is small, its shape and location are easily affected by tumors, the segmentation results of pancreas are unsatisfactory, and the method does not segment the four organs well in the face of some diseases, such as hydrops in the abdomen. In addition, this method has a certain prior knowledge, we know what kind of cases there may be in the validation set before designing this method, can not guarantee that this method can be applied to more unknown patients. How to make the segmentation method more robust is still a problem worth discussing.

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] Hall E J, Phil D and Sc D , ″ Intensity-modulated radiation therapy, protons, and the risk of second cancers, ″ *Int. J. Radiat. Oncol*, vol.65, pp.1–7, 2006.

[2] Nelms B, Etom W A, Robinson G and Heeler J W, ″ Variations in the contouring of organs at risks: test case from a patient with oropharyngeal cancer, ″ *Int. J. Radiat. Oncol. Biol. Phys*, vol.82, pp. 368–378, 2012.

[3] Fu Y , ″ A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy, ″ *Med. Phys,* vol.45, pp.29–37, 2018.

[4] He K and Sun J, ″ Deep Residual Learning for Image Recognition, ″ in *IEEE Conf. Comput. Vis. Pattern Recognit.,*2016, pp. 770–778.

[5] Szegedy C, Liu W, Jia Y Q, ″ Going deeper with convolutions, ″ in *IEEE Conference on Computer Vision and Pattern Recognition,* 2015, pp. 1-9.

[6] Hu J, Shen L and Sun G, ″ Squeeze-and-excitation networks, ″ in Proc. *IEEE Conf. Comput. Vis. Pattern Recog,* 2018, pp. 7132–7141.

[7] A. L. Simpson, M. Antonelli and S. Bakas, ″ A large annotated medical image dataset for the development and evaluation of segmentation algorithms, ″ *arXiv preprint arXiv:1902.09063*, 2019.

[8] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng and J. Hesser, ″ The liver tumor segmentation benchmark (lits), ″ *arXiv preprint arXiv:1901.04056*, 2019.

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

[10] 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

[11] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffifitt, 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.

[12] 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 contrastenhanced ct imaging: Results of the kits19 challenge," *Medical Image Analysis*, vol. 67, p. 101821, 2021.

[13] 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 artifificial intelligence to defifine 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.

[14] 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.