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

Yinan Xu[1], Siying Cao[1], Wenhao Dong[1], Nuo Tong[2], Shuiping Gou[1,2]

[1]Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi, 710071, China

[2]AI-based Big Medical Imaging Data Frontier Research Center, Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an, Shaanxi 710071, China

ynxu_1@stu.xidian.edu.cn

Abstract. Abdominal multi-organ segmentation is of great significant for preoperative treatment planning. At present, there are many public abdominal data sets and deep learning based segmentation methods have been proposed. However, the problem of polycentric and spatio-temporal inefficiency still remain unsolved. To solve the domain gaps among the datasets from different source, a lot of preprocessing techniques are commonly utilized, including adjusting the window level and window width, cropping and resizing the images to a uniform size. Data augmentation is also employed to enlarge the dataset. In this work, a lightweight network based on 3D-ResNet (ISE-Net) are adopted to improve the spatio-temporal efficiency. The inception and squeeze-and-excitation (SE) block are incorporated into the ISE-Net to obtain multi-scale features, which is able to reduce the differences in the resolution of the polycentric data. To mitigate the effects of liver and pancreatic diseases on segmentation, shape constraint is also employed. In post-processing, small connected regions are removed to reduce the false positive islands. It is worth noting that only CPU was utilized in the inference phase, which reduces the hardware requirements.

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

1. Introduction

Accurate contouring of target organs is essential in clinical applications, including disease diagnosis, radiotherapy treatment planning, and delivery [1]. Thus, 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]. Accurate and robust automated segmentation techniques 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 deep learning based segmentation methods have been proposed and achieved great success in abdominal organ segmentation. However, they usually contain large number of parameters which leads to the difficulties of not being clinically useful. Additionally, the training data and test data are

usually from the same medical center.

To meet the needs of fast speed and low memory occupancy, a lightweight network based on the 3D ResNet [4] structure is designed, which has only one tenth of the parameters of the baseline 3D nnUnet. Moreover, inception block is used to extract multi scale features and SE block [5, 6] is used to get channel attention, which allows the proposed method can easily adapt to different CT image resolutions. To deal with the domain gaps from multi-center datasets, a lot of data preprocessing techniques are utilized. For example, adjusting the window level and window width of the CT images, dropping the slices without target organ, data augmentation with CLAHE, random flip, and random rotate. Moreover, to mitigate the effects of liver and pancreatic diseases on segmentation, shape constraint is also commonly employed.

2. Method

The structure of the proposed segmentation network is illustrated in Fig1, which is a 3D ResNet with Inception and SE blocks (ISE block).

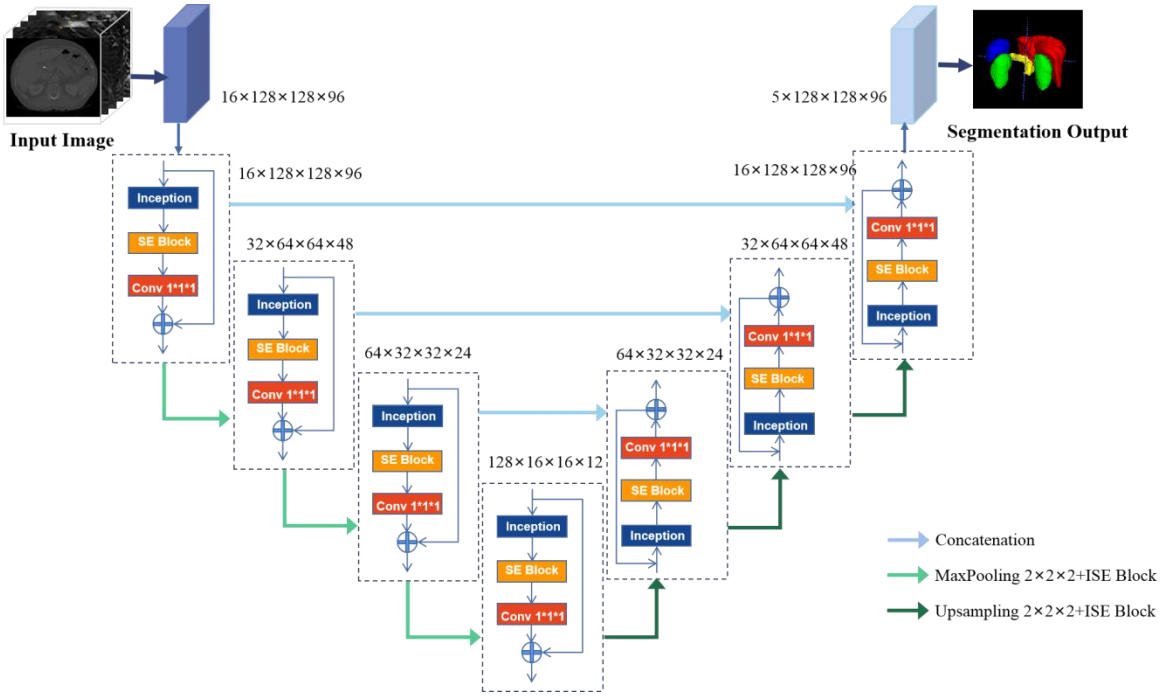

Fig1: The overall architecture of the proposed ISE-Net.

2.1 Inception and SE block (ISE Block)

The structure of the proposed ISE block in the proposed ISE-Net is shown in Fig2, which integrates the residual block, Inception block, and SE block.

The convolution kernels of different sizes are used in the Inception block to obtaint the receptive fields of different sizes. Then, features from different kernels are fused to get multi scale features to reduce the impact of different resolutions of anisotropic CT images on organ segmentation. The Inception block for this work is shown in Fig3.

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 (SE block) is employed here to recalibrate the importance of the multi scale features obtained by the Inception block. The specific components and structure of the SE block is illustrated in Fig4. 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.

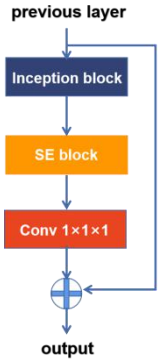

Fig2: The detailed architecture of the ISE block.

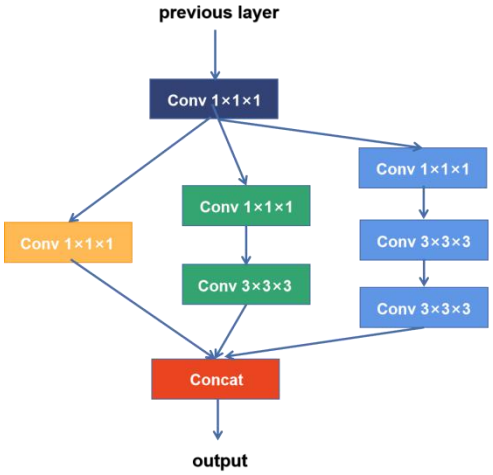

Fig3: The detailed architecture of the Inception block.

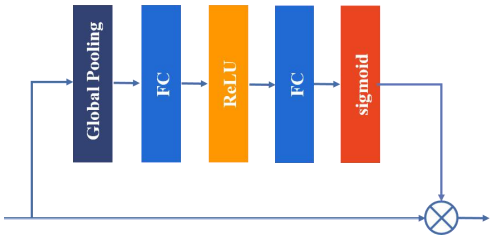

Fig4: The detailed architecture of the SE block.

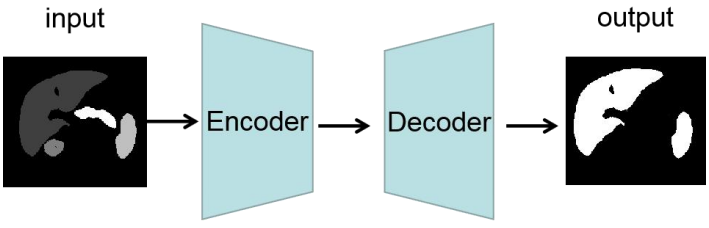

Fig5: The detailed architecture of the Shape constraint model.

2.2 Shape constraint model

A pre-trained binary segmentation network is employed as the shape constraint model for the training of the segmentation network. Specifically, the input of the shape constraint model is the ground truths of the four abdominal organs, and the output is the binary segmented results of liver and pancreas. Since masks are utilized as the model input, the shape constraint model can be regarded as a

network for extracting shape features. The architecture of the model is shown in Fig5.

2.3 The combination of shape constraint model and ISE-Net

The shape constraint model is combined with the ISE-Net through an additional loss term in the loss function. Thus, the loss function used in the proposed method is composed of two parts, the segmentation loss and shape constraint loss. The loss function is as follows:

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

where P is the prediction result of ISE-Net, T is the corresponding ground truth, and $E(.)$ represents the encoder of the shape constraint model. Dice loss is used as the segmentation loss, which can solve the imbalance between positive and negative samples, and *MSELoss* is the mean square error loss function, which is used to calculate the differences between the shape features of the prediction and ground truth.

2.4 Preprocessing

The following preprocessing techniques were utilized in this work:

• Cropping strategy:

The CT scan ranges are varies greatly from patient to patient. Specifically, some patient'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, they are cropped according to the labels. For unknown validation or test data, a lightweight binary segmentation network are trained to roughly locate the four target organs for cropping.

• Adjusting window level and window width:

To make the contrast of the organs of concern higher, adjusting the window level and window width are necessary. According to the knowledge of CT images, the window level and width were adjusted to 40 and 400, respectively.

• Resizing method for anisotropic data:

The main shortcoming of the patches-based segmentation methods is the redundant computation in the inference process and unable to learn global features. To overcome the problem, the whole CT images were utilized as the network input in this work. To match the input size of ISE-Net, the CT images were resized to a uniform size of 128×128×96.

• Intensity normalization method:

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

• Data augmentation method:

Firstly, a few difficult kidney tumor samples were selected and augmented individually. Then, CLAHE, random rotation, flips, and scales were used to augment the whole training dataset.

2.5 Proposed Method

• Network architecture details: The network architecture is shown in Fig1.

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

• Number of the parameters of ISE-Net: 4,326,525

• Number of flops: 57,632,408

2.6 Post-Processing

The post-processing operations used in the work include removing small connected areas and filling the holes to reduce false positive islands.

3. Dataset and Evaluation Metrics

3.1 Dataset

• Dataset details:

The dataset used in FLARE2021 was collected 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 was employed and resulted 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 was applied as training steps.
- Post-processing steps of the network outputs: removing small connected areas to reduce the false positive islands and resizing the predictions back to the original size of the predicted image.

5. Result

5.1 Quantitative results on validation set.

Table 4 illustrates the results on the validation cases. Liver has a large volume, so that its 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 have large kidney tumors, which makes segmentation difficult. Thus, such cases were augmented individually by changing the HU value distributions of the few kidney tumors in the training set to improve the performance of the segmentation network on kidney. The pancreas is a very small organ, which is difficult to segment, so the DSC value of pancreas 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

Fig6 presents some cases that are relatively easy to be segmented. It can be observed that the proposed network shows promising segmentation performance.

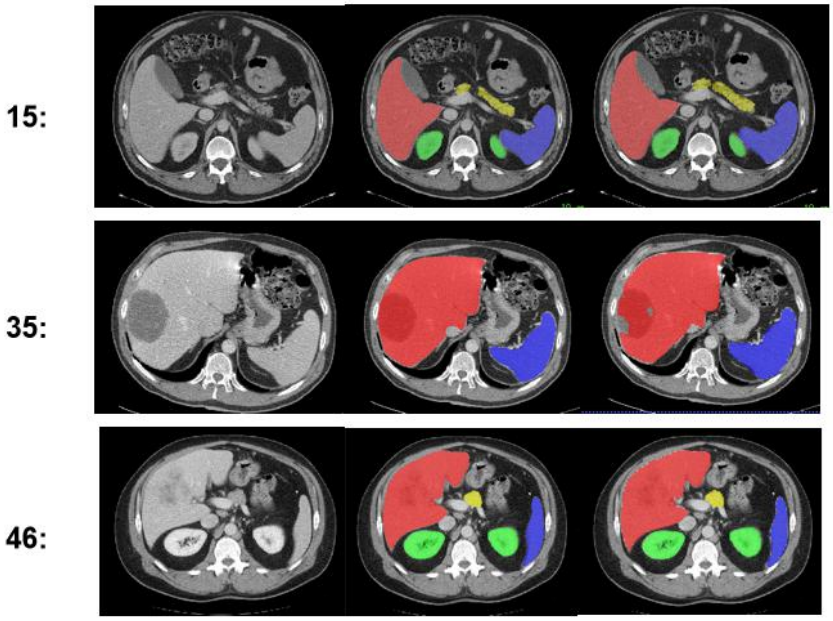

Fig6. Segmentation results by the proposed ISE-Net. The first column is the CT images, the second and third columns show the ground truth and the segmentation results, respectively.

Fig7 shows some cases that are challenging to segment. The first row of Figure 7 illustrates a fatty liver case where the liver is darker than healthy ones. The baseline method failed to segment the spleen (blue) and the liver (red) in this case, but the proposed ISE-Net segmented them. The second row shows an example with abdominal effusion which causes incorrect segmentation (green). The third row shows an example of a large pancreatic tumor and the proposed method misclassified it into liver tissue. The example in the fourth row contains a kidney tumor that the proposed method correctly identified, however, the segmentation performance of pancreas is not satisfactory.

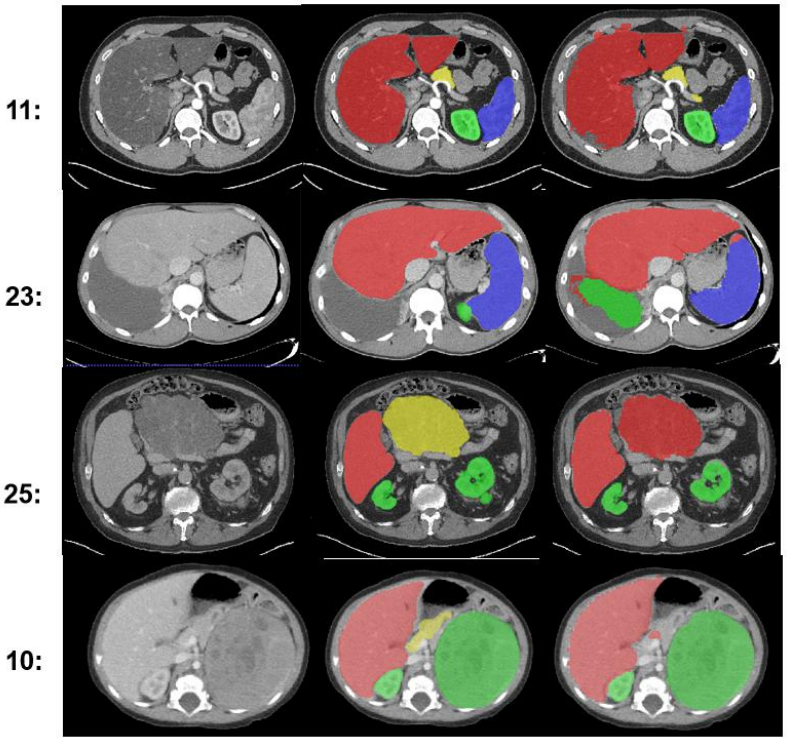

Fig7. Segmentation results by the proposed ISE-Net. The first column is the CT images, the second and third columns show the ground truth and the segmentation results, respectively.

6. Discussion and Conclusion

Generally, the proposed method can be better applied to simple cases where there are no diseases or not serious diseases. When patients have liver diseases that make the liver darker than normal, the proposed method is also performs better than the baseline method. At the same time, the proposed method can effectively segment large kidney tumors. Additionally, 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 also has some limitations. Because the pancreases are small and their shape and location are easily affected by tumors, the segmentation results of pancreases are unsatisfactory. Also, the method does not performs well in the face of some diseases, such as hydrops. Additionally, this method has a certain prior knowledge, we know what kind of cases there may be in the validation set before designing this method, there is no guarantee that this method can be applied to more patients with unknown diseases. How to make the segmentation method more robust is still a problem that worth further discussing.

Acknowledgment

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE challenge didn't use any pre-trained models or 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.