Uncertainty-aware Mean Teacher Framework with Inception and Squeeze-and-Excitation Block for MICCAI FLARE22 Challenge

Hui Meng[1], Haochen Zhao[2], Ziniu Yu[2], Qingfeng Li[2], and Jianwei Niu[2,3]

[1] School of Intelligent Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, 1 Sub-lane Xiangshan, Hangzhou 310024, China.
[2] Research center of Big Data and Computational Intelligence, Hangzhou Innovation Institute of Beihang University, Hangzhou 310051, China.
[3] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, also with the Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC), Beihang University, Beijing 100191, China.

Abstract. Semi-supervised learning has attracted extensive attention in the field of medical image analysis. However, as a fundamental task, semi-supervised segmentation has not been investigated sufficiently in the field of multi-organ segmentation from abdominal CT. Therefore, we propose a novel uncertainty-aware mean teacher framework with inception and squeeze-and-excitation block (UMT-ISE). Specifically, the UMT-ISE consists of a teacher model and a student model, in which the student model learns from the teacher model by minimizing segmentation loss and consistency loss. Additionaly, we adopt an uncertainty-aware algorithm to make the student model learn accurate and reliable targets by making full use of uncertainty information. To capture multi-scale features, the inception and squeeze-and-excitation block are incoporated into the UMT-ISE. It is worth noting that abdominal CT of test cases are first extracted before multi-organ segmentation in the inference phase, which significantly improves segmentation accuracy.

Keywords: semi-supervised learning, multi-organ segmentation, uncertainty estimation, multi-scale features

1 Introduction

Accurate segmentation of medical images is essential for many clinical applications, such as disease diagnosis and tumor localization [3]. Nowadays, manual segmentaion results given by radiologists are widely regarded as gold standards. However, manual segmentation is tedious and time consuming. Additionally, manual segmentation heavily depends on radiologists' experience and suffers from intra- and inter-observer variabilities. Therefore, many researchers have developed different automatic segmentation methods [12], which are supposed to assist radiologists to make accurate diagnosis.

For abdominal organ segmentation, most research work focus on single organ segmentation, such as kidney [6] or blood vessels [9]. Compared with single-organ segmentation, multi-organ segmentation faces two major challenges. The first one is that large morphological differences between multiple organs limit accurate segmentation of all organs. The second one is that it's difficult to obtain large dataset with accurate annotations for multi-organ segmentation. Therefore, it is necessary to make full use of unlabeled medical images to improve the multi-organ segmentation accuracy [4].

To utilize unlabeled medical images effectively, we propose a novel uncertainty-aware mean teacher framework with inception and squeeze-and-excitation block (UMT-ISE) for segmenting multiple organs from 3D abdominal CT. The UMT-ISE is constructed based on conventional teacher-student model [2], which consists of a teacher model and a student model. For the same unlabeled data under different perturbations, the segmentation predictions of the teacher model and the student model are constrained to be consistent [13]. Different from the conventional teacher-student model, the UMT-ISE adopts framework of uncertainty-aware mean teacher. The teacher model in the UMT-ISE generates multiple predictions for each target under Monte Carlo sampling and gives uncertainty evaluation. The predictions with high uncertainty are filtered out and the predictions with low uncertainty are retained to compute consistency loss. Based on the design of the uncertainty evaluation, the teacher model tends to generate high-quality predictions and the student model can be constantly optimized. Considering multiple organs have different sizes, the inception and squeeze-and-excitation (ISE) block are incoporated into the UMT-ISE to capture multi-scale features.

2 Method

2.1 Preprocessing

The following preprocessing techniques are utilized in this work:

- Cropping strategy in z-axis direction:
  The range of CT scans varies depending on the situation. For example, some patients may have CT scans not only of abdominal area, but of entire chest, lower abdomen and even legs. In some cases, only the abdominal region containing target organs is presented. Therefore, it is necessary to filter out some irrelevant and redundant slices in CT scans. In this study, we train an uncertainty-aware mean teacher (UA-MT) network to perform a rough abdominal segmentation.
  To extract CT scans only containing abdominal region, we implement different cropping strategies in z direction. For training data with labels, we tailor them according to the range of target organs in annotations. For training data without labels, validation data and test data, we first implement rough segmentation of target organs based on the trained UA-MT network and then crop CT scans according to the scope of segmentated organs.

During reference, we adopt specific preprocessing for large samples containing more than 800 slices and with z-axis spacing of 1. For these samples, we first equally divide the whole CT scans into three parts. Then, the rough segmentation of target organs is implemented for each part. Finally, the CT scans containing target organs are extracted based on the segmentation results.

- Cropping strategy in x-axis and y-axis directions:
  According to observations, different samples have different proportions of target region to CT images in x-axis and y-axis directions. In some samples, the target organs only occupy a small region in CT images. It is necessary to cropping redundant background in x-axis and y-axis direction to enlarge the target organs. Conversely, the target organs in some samples occupy a large region in CT images. The target organs in these samples are close to the edge of the CT images, which resulting in mis-segmentation of target organs. For these samples, we pad the CT images with zero in x-axis and y-axis directions to ensure appropriate proportions of the target region to the corresponding CT images.
- Adjusting window level and window width:
  In order to achieve high contrast between the target organs and the background area, we adjust window width and window level of the original CT images. According to doctor's experience, the window width and window level of the CT images are adjusted to 40 and 255, respectively.
- Image Resampling:
  In this study, the network input of UMT-ISE is randomly cropped patches from whole CT images. The input size of the UMT-ISE is $112\times112\times80$, while the size of the whole CT images is much larger than $112\times112\times80$. To ensure the cropped patches contain efficient information, we resample all CT images to $192\times192\times96$ after the above preprocessing.
- Image normalization:
  After the above preprocessing, we implement z-score normalization on CT images based on the mean and standard deviation of the intensity values.
- Data augmentation:
  In this study, we implement random cropping on the whole CT images to obtain network input. Additionally, horizontal flipping are performed to achieve data augmentation.

2.2 Proposed Method

Strategies to use the unlabelled cases:

The input of teacher model and student model are the same CT images with different noises, and the output of the two models are constrained by unsupervised loss function.

Network architecture details:

The network architecture of UMT-ISE is shown in Fig.1. The UMT-ISE is composed of two modified V-Net models, i.e. the teacher model and the student model, and the two models share the network structure. In multi-organ

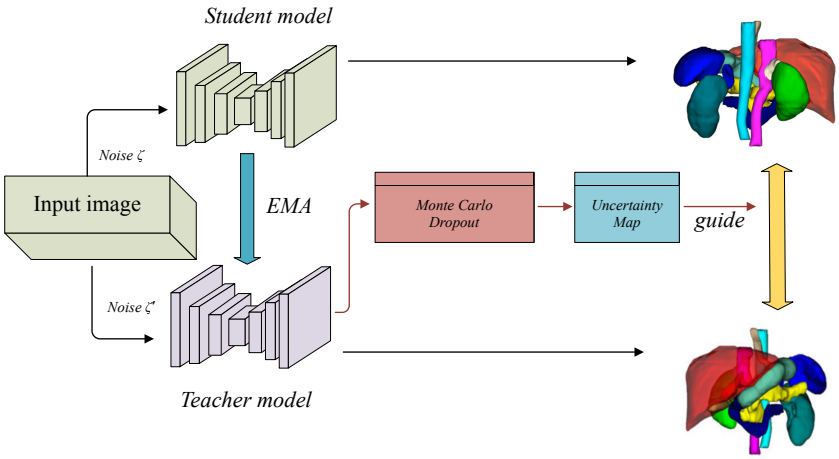

Fig. 1. Network architecture of the UMT-ISE. The UMT-ISE is composed of a student model and a teacher model. The estimated uncertainty from the teacher model guides the student model to learn from the more reliable targets.

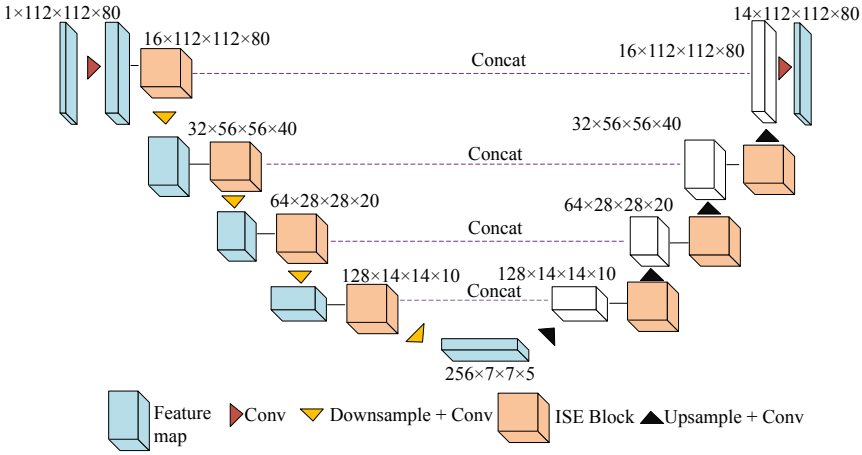

Fig. 2. Network architecture of the modified V-Net. The ISE blocks are inserted at each layer of encoder and decoder.

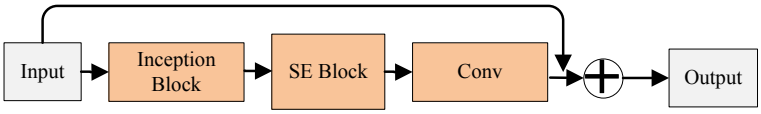

Fig. 3. The detailed architecture of the ISE block.

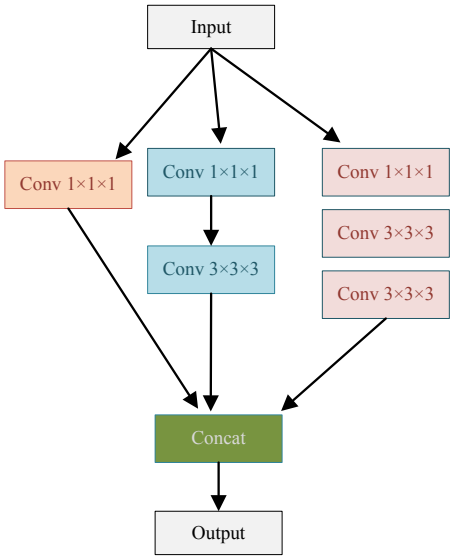

Fig. 4. The detailed architecture of the Inception block.

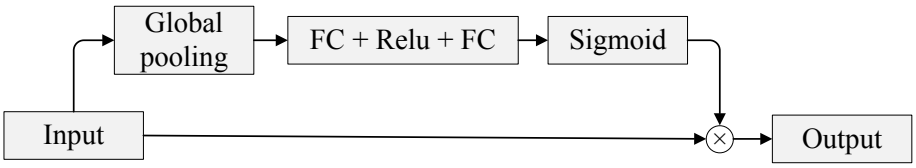

Fig. 5. The detailed architecture of the SE block.

segmentation, different organs have different size and the segmentation accuracy of small organs is low. In order to improve the accuracy of multi-organ segmentation, we add ISE blocks in V-Net to obtain the modified V-Net. The ISE block mainly contains inception block and SE block, which can obtain multi-scale feature maps with channel attention. The network structure of the modified V-Net is shown in Fig.2.

The structure of the ISE block is shown in Fig.3. The ISE block integrates the residual block, Inception block, and a squeeze-and-excitation (SE) block. Multiple convolution layers with different convolution kernels are used in the Inception block to obtain multiple feature maps with different receptive fields. Then, the feature maps are fused to generate multi-scale features to alleviate the impact of size diversity in multi-organ segmentation. The structure of the Inception block is shown in Fig.4.

Although the Inception block can enhance features of targets with different size, the redundant features in the multi-scale feature map reduce the discriminability of the network. Thus, a SE block is adopted 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 Fig.5. In the SE block, a global average pooling layer is used to aggregate the global information, which is followed by two fully connected (FC) layers to capture the channel-wise relationships. Then, the features given by the Inception block is recalibrated by the channel-wise relationships through point-wise multiplication.

Loss function:

In this study, we adopt Dice loss and cross entropy loss to calculate supervised loss on labelled data. Additionally, consistency loss on both unlabelled data and labelled data are calculated to optimize the network.

2.3 Post-processing

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

3 Experiments

3.1 Dataset and evaluation measures

The FLARE2022 dataset is curated from more than 20 medical groups under the license permission, including MSD [11], KiTS [7,8], AbdomenCT-1K [10], and TCIA [1]. The training set includes 50 labelled CT scans with pancreas disease and 2000 unlabelled CT scans with liver, kidney, spleen, or pancreas diseases. The validation set includes 50 CT scans with liver, kidney, spleen, or pancreas diseases. The testing set includes 200 CT scans where 100 cases has liver, kidney, spleen, or pancreas diseases and the other 100 cases has uterine corpus endometrial, urothelial bladder, stomach, sarcomas, or ovarian diseases. All the CT scans only have image information and the center information is not available.

The evaluation measures consist of two accuracy measures: Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD), and three running efficiency measures: running time, area under GPU memory-time curve, and area under CPU utilization-time curve. All measures will be used to compute the ranking. Moreover, the GPU memory consumption has a 2 GB tolerance.

3.2 Implementation details

Environment settings The environments and requirements are presented in Table 1.

Table 1. Environments and requirements.

Windows/Ubuntu version	Windows 10
CPU	Intel(R) Core(TM) i9-9900K CPU @ 3.60GHz
RAM	16×2GB;
GPU (number and type)	1 NVIDIA Tesla V100 GPU (48G)
CUDA version	11.1
Programming language	Python 3.6
Deep learning framework	Pytorch (Torch 1.7.0, torchvision 0.8.0)
Specification of dependencies	None
(Optional) Link to code	

Training protocols The training protocols are presented in Table 2.

Table 2. Training protocols.

Network initialization	“he" normal initialization
Batch size	16
Patch size	192×192×96
Total epochs	2000
Optimizer	SGD with nesterov momentum ($\mu = 0.99$)
Initial learning rate (lr)	0.05
Lr decay schedule	halved by 1000 epochs
Training time	48 hours
loss function	cross entropy loss + Dice loss+λ *consistency loss
Number of model parameters	9.44M[4]
Number of flops	41.40G[5]
CO_2eq	1 Kg[6]

4 Results and discussion

4.1 Quantitative results on validation set

Table 3 and Table 4 illustrates the results on the validation cases.

Table 3. Quantitative results on validation set(with unlabelled data).

Organ	DSC (%)
Mean DSC	0.7458
Liver	0.9547
RK	0.8500
Spleen	0.8885
Pancreas	0.7143
Aorta	0.8497
IVC	0.7650
RAG	0.6116
LAG	0.5108
Gallbladder	0.6429
Esophagus	0.6825
Stomach	0.8169
Duodenum	0.5471
LK	0.8608

Table 4. Quantitative results on validation set(only labelled data).

Organ	DSC (%)
Mean DSC	0.7193
Liver	0.9221
RK	0.8280
Spleen	0.8118
Pancreas	0.7148
Aorta	0.8020
IVC	0.7331
RAG	0.6158
LAG	0.5473.
Gallbladder	0.5677
Esophagus	0.6548
Stomach	0.7629
Duodenum	0.5748
LK	0.8164

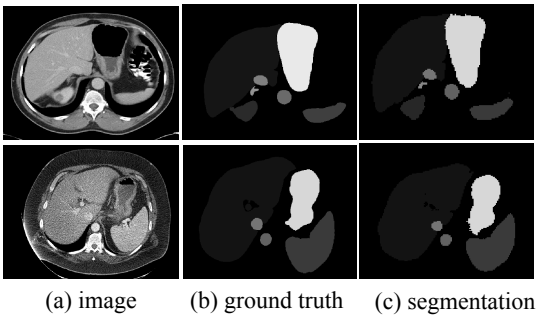

Fig. 6. Well-segmented examples from validation sets

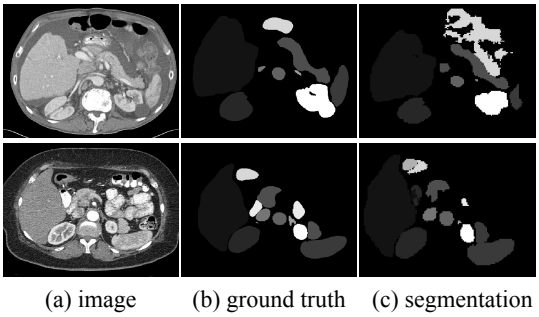

Fig. 7. Challenging examples from validation sets

4.2 Segmentation efficiency results

Fig.6 and Fig.7 show examples with good segmentation results and bad segmentation results, respectively. As for the bad segmentation cases, we think there are two reasons. One is that the low imaging quality of the CT images cause the bad segmentation. Specifically, there are some dark holes in some organs which results in broken segmentation results of the organs. The other one is that our model is not robust enough for accurate segmentation of small organs.

5 Conclusion

In this study, we propose a novel UMT-ISE for multi-organ segmentation in abdominal CT. The UMT-ISE achieves fast and accurate multi-organ segmentation. Additionally, our method can be tested on CPU, which is convenient to complete some clinical tasks. However, our method still has some limitations.

For some small organs, their shapes and positions are easily affected by tumors and edema. Our method is not robust enough for segmentation of small organs. Additionally, it is difficult to extract abdominal regions for the cases with many CT slices, and the segmentation results of these cases are not satisfied. Our future work will focus on the accurate segmentation of small organs in multi-organ segmentation and develop more robust segmentation methods.

6 Acknowledgment

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE22 challenge didn't use any pre-trained models or additional datasets other than those provided by the organizers.

References

1. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013) 6
2. Cui, W., Liu, Y., Li, Y., Guo, M., Li, Y., Li, X., Wang, T., Zeng, X., Ye, C.: Semi-supervised brain lesion segmentation with an adapted mean teacher model. In: International Conference on Information Processing in Medical Imaging. pp. 554–565. Springer (2019) 2
3. Fu, Y., Mazur, T.R., Wu, X., Liu, S., Chang, X., Lu, Y., Li, H.H., Kim, H., Roach, M.C., Henke, L., et al.: A novel mri segmentation method using cnn-based correction network for mri-guided adaptive radiotherapy. Medical physics **45**(11), 5129–5137 (2018) 1
4. Ganaye, P.A., Sdika, M., Benoit-Cattin, H.: Semi-supervised learning for segmentation under semantic constraint. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 595–602. Springer (2018) 2
5. Hall, E.J.: Intensity-modulated radiation therapy, protons, and the risk of second cancers. International Journal of Radiation Oncology* Biology* Physics **65**(1), 1–7 (2006)
6. Heimann, T., Van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., et al.: Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE transactions on medical imaging **28**(8), 1251–1265 (2009) 2
7. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., 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 **67**, 101821 (2021) 6
8. Heller, N., McSweeney, S., Peterson, M.T., Peterson, S., Rickman, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Rosenberg, J., 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 **38**(6), 626–626 (2020) 6
9. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Computing Surveys (CSUR) **36**(2), 81–121 (2004) 2

10. Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., Cao, S., Zhang, Q., Liu, S., Wang, Y., Li, Y., He, J., Yang, X.: Abdomenct-1k: Is abdominal organ segmentation a solved problem? IEEE Transactions on Pattern Analysis and Machine Intelligence (2021). https://doi.org/10.1109/TPAMI.2021.3100536 6
11. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019) 6
12. Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E.K., Yuille, A.L.: Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical image analysis **55**, 88–102 (2019) 1
13. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 605–613. Springer (2019) 2