Dear editor,

Thank you very much for your decision letter and helpful advices on our manuscript. We have paid our close attention to all comments and suggestions raised from the reviewers. Revision has been carefully made to the manuscript and point-by-point responses to the comments are listed below this letter.

We greatly appreciate the comments and suggestions from reviewers and your thorough consideration of our manuscript. We hope our replies and revision to the manuscript would satisfy the requirements from you and reviewers. Please feel free to let me know if you have further questions on our revised manuscript. Thank you again and best wishes to you and the reviewers.

Look forward to hearing from you soon.

With best wishes,

Yours sincerely,

Yi Lv

School of Mechanical Engineering and Automation, Beihang University, Beijing,

100191, China

E-mail: [dnkii@buaa.edu.cn](mailto:dnkii@buaa.edu.cn)

**To Reviewer CLBQ**

**Comment 1:** No explanation on the method used for semi-supervised (pseudo-label or consistency regularization)

***Re:*** *Thanks for your comment. We have extended the description of semi-supervised learning method used in our research.*

*“In order to further improve the robustness of the network on different data, we adopt the training strategy of semi-supervised learning. Since no research has proved that more unlabeled data in semi supervised learning is better, we set the unlabeled data as much as the number of labeled data. In the training process, we use 40 labeled data and 50 randomly selected unlabeled data as the training set used in the stage of global locating and organ locating. We use the labeled data to train the model in the first 50 epochs, and then introduce the unlabeled data. We use the trained model to segmentation the unlabeled dataset after each five epochs, and we use the results as the label for training. As the first two stages are the segmentation of complete CT, which is different from the third stage, we only use the semi-supervised learning strategy for the first two stages.”*

**Comment 2:** Made in MS Word without aligning with the provided style guide. The paper seems completely different than others.

***Re:*** *Thanks for your comment. We have changed the file into Latex style and the source file can be founded in the ‘Supplementary Material’.*

**Comment 3 & Comment 4:** It would be better to mention the name of the stages in the abstract instead of just mentioning their numbers.

Why is only the first-fold validation set result mentioned in the abstract?

***Re:*** *Thanks for your comment. We rewrote the abstract section to eliminate the problems mentioned.*

*“Abdominal multi-organ segmentation is fast becoming a key instrument in preoperative diagnosis. Using the results of abdominal CT image segmentation for three-dimensional reconstruction is an intuitive and accurate method for surgical planning. In this paper, we propose a stable three-stage fast automatic segmentation method for abdominal 13 organs: liver, spleen, pancreas, right kidney, left kidney, stomach, gallbladder, esophagus, aorta, inferior vena cava, right adrenal gland, left adrenal gland, and duodenum. Our method includes preprocessing the CT data, segmenting the multi-organ and post-processing the segmentation outputs. The results on the test set show that the average DSC performance is about 0.766. The average time and GPU memory consumption for each case is 81.42s and 1953MB.”*

**Comment 5:** What does making maximum in-plain resolution 128x176 mean for a 3D CT scan?

***Re:*** *Thanks for your comment. The purpose of this step is to prevent the data size from being too large. As the spacing in the dataset varies greatly, the resolution of data with large spacing will become very large after spacing normalization. So it is necessary to set a maximum in-plain resolution.*

**Comment 6:** The description of the model in section 2.2 is not clear, especially Each layer of the encoding path contains two 3 × 3 × 3 convolution layers, each followed by a ReLu layer, followed by a 2 × 2 × Maximum pool layer with step size of 2 in each direction of 2. In the decoding path, each layer contains a 2 with a step size of 2 × 2 × 2, followed by two 3 × 3 × 3, each followed by a RuLu layer.

***Re:*** *We are very grateful for you to point out this. We have corrected it to “The network includes an encoding path and a decoding path, each of which has four resolution levels. Each level of the encoding path contains two 3 × 3 × 3 convolution layers, and the convolution layers followed by a ReLu layer and a 2 × 2 ×2 Maximum pool layer with step size of 2. In the decoding path, each level also contains two 3 × 3 × 3 convolution layers, and the convolution layers followed by a ReLu layer and an upsampling layer.”*

**Comment 7:** The first stage can be formulated as a binary segmentation task; whether a pixel is an abdomen or not. Making multi-class output for the stage makes little to no sense.

***Re:*** *Thanks for your considerate comment. We did consider this option before submission. The reason for using multi-class output is that, on the one hand that can make it available to evaluate the accuracy of each organ in the first stage, and on the other hand that has little impact on reasoning time. What’s more, the multi-class output is more suitable for the maximum region growth algorithm.*

**Comment 8:** Since the training strategy of semi-supervised learning is not mentioned, I suppose it is using the pseudo-labels as a label. But you need to mention the strategy explicitly for the readers.

***Re:*** *Thanks for your considerate comment.* *We added a description of semi supervised learning strategies. “We use the labeled data to train the model in the first 50 epochs, and then introduce the unlabeled data. We use the trained model to segmentation the unlabeled dataset after each five epochs, and we use the results as the label for training.”*

**Comment 9:** No reasoning behind why only 50 unlabeled data were used and how there were selected.

***Re:*** *Thanks for your considerate comment.* *We added a description of the reason. “In order to further improve the robustness of the network on different data, we adopt the training strategy of semi-supervised learning. Since no research has proved that more unlabeled data in semi supervised learning is better, we set the unlabeled data as much as the number of labeled data. In the training process, we use 40 labeled data and 50 randomly selected unlabeled data as the training set used in the stage of global locating and organ locating.”*

**Comment 10:** No reasoning behind why semi-supervised learning was used in the first two stages.

***Re:*** *Thanks for your considerate comment.* *We added a description of the reason. “…As the first two stages are the segmentation of complete CT, which is different from the third stage, we only use the semi-supervised learning strategy for the first two stages.”*

**Comment 11:** The input data size seems to be variable from section 2.1. How did you convert the CT scan to 64×128×176?.

***Re:*** *Thanks for asking this.* *In this step, when the data size is larger than 64×128×176, 64×128×176 data will be intercepted from the image center. When the data size is smaller than 64×128×176, 0 will be added outside the image to make the size reach 64×128×176.*

**To Reviewer 3GXU**

**Comment 1:** Lack of DSC and NSC analysis at each stage.

***Re:*** *Thanks for your considerate comment. We added a table to analyze the accuracy between different stages.*

**Table 4.** Comparison on 13 Structures on official validation and testing results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Official Validation | | | Testing Result |
| Stage 1 | Stage 2 | Stage 3 |
| Liver | 0.902 | 0.868 | 0.903 | 0.866 |
| Right Kidney | 0.798 | 0.864 | 0.896 | 0.885 |
| Spleen | 0.802 | 0.898 | 0.926 | 0.910 |
| Pancreas | 0.497 | 0.708 | 0.683 | 0.676 |
| Aorta | 0.837 | 0.911 | 0.930 | 0.915 |
| Inferior Vena Cava | 0.710 | 0.802 | 0.846 | 0.836 |
| Right Adrenal Gland | 0.490 | 0.600 | 0.653 | 0.695 |
| Left Adrenal Gland | 0.354 | 0.528 | 0.610 | 0.653 |
| Gallbladder | 0.393 | 0.463 | 0.562 | 0.569 |
| Esophagus | 0.573 | 0.652 | 0.671 | 0.666 |
| Stomach | 0.672 | 0.780 | 0.799 | 0.807 |
| Duodenum | 0.379 | 0.577 | 0.593 | 0.589 |
| Left Kidney | 0.789 | 0.856 | 0.867 | 0.893 |
| Average | 0.630 | 0.731 | 0.765 | 0.766 |

**To Reviewer Ms9D**

**Comment 1:** Most training details are descripted, but the adopted semi-supervised learning method is missing.

***Re:*** *Thanks for your comment.* *We have extended the description and comparison of semi-supervised learning method used in our research.*

*“In order to further improve the robustness of the network on different data, we adopt the training strategy of semi-supervised learning. Since no research has proved that more unlabeled data in semi supervised learning is better, we set the unlabeled data as much as the number of labeled data. In the training process, we use 40 labeled data and 50 randomly selected unlabeled data as the training set used in the stage of global locating and organ locating. We use the labeled data to train the model in the first 50 epochs, and then introduce the unlabeled data. We use the trained model to segmentation the unlabeled dataset after each five epochs, and we use the results as the label for training. As the first two stages are the segmentation of complete CT, which is different from the third stage, we only use the semi-supervised learning strategy for the first two stages.”*

|  |  |  |
| --- | --- | --- |
|  | Average DSC | Standard Deviation of DSC |
| With Unlabeled Data | *0.731* | *0.142* |
| Without Unlabeled Data | *0.678* | *0.186* |