Self-supervised approach for organs at risk segmentation of abdominal CT images

Seenia Francis[1,2], *Coutinho Austin* Minino[1], *Pournami* P N[1], *Niyas P*uzhakkal[3], *and Jayaraj* P B[1]

[1] National Institute of Technology Calicut, Kerala, India.
[2]Jyothi Engineering College, Cheruthuruthy, Kerala, India.
[3]MVR Cancer Centre & Research Institute, Calicut, India.

Abstract. Accurate segmentation of organs at risk is essential for radiation therapy planning. However, manual segmentation is time-consuming and prone to inter and intra-observer variability. This study proposes a self-supervision based attention UNet model for OAR segmentation of abdominal CT images. The model utilizes a self-supervision mechanism to train itself without the need for manual annotations. The attention mechanism is used to highlight important features and suppress irrelevant ones, thus improving the model's accuracy. The model is evaluated on a dataset of 100 abdominal CT scans and compared its perfor mance with state-of-the-art methods. Our results show that the proposed model got comparable performance in terms of the dice similarity coefficient. More over, the inference time is much faster than traditional manual segmentation methods, making it a promising tool for clinical use.

1 Introduction

Medical image segmentation is a crucial task in medical imaging, as it plays a significant role in the precise diagnosis and treatment planning of various medical conditions. Organ At Risk (OAR) segmentation is a critical aspect of medical image segmentation, which involves distinguishing and contouring organs in medical images of the human body anatomy. OARs are healthy tissues near the tumour volume that may be affected by radiation treatment. Accurate segmentation of abdominal organs is essential for radiotherapy treatment planning of abdominal cancer patients, as it enables precise targeting of radiation to the tumour while minimizing exposure to healthy surrounding tissues. The abdominal cavity contains various organs, such as the liver, spleen, pancreas, stomach, and intestines, which differ in shape, size, and position. Computed Tomography (CT) images are commonly used for radiotherapy and for organ segmentation.

Manual segmentation methods, traditionally used for organ segmentation, are timeconsuming and prone to inter-observer variability. The process requires experts to accurately identify organs and tissues, making it highly dependent on their subjective interpretation. Automatic segmentation methods can provide a solution to the limitations of manual approaches by producing accurate and consistent segmentation results. This approach relies on advanced image processing techniques and anatomical knowledge. Deep learning-

based approaches have shown promising results in automating the segmentation process due to the increasing availability of large-scale medical imaging datasets [1, 2] However, these approaches typically require large amounts of labelled data for training, which may not always be available in the context of radiotherapy treatment planning. Despite this challenge, deep learning-based approaches have demonstrated significant potential in improving the efficiency and accuracy of organ segmentation for radiotherapy treatment planning.

Self-supervised learning [3] has emerged as a promising approach for addressing the issue of limited labeled data. As a type of unsupervised learning, it can learn meaningful representations by leveraging the inherent structure of the data without requiring explicit annotations. In this paper, we propose a UNet-based self-supervised learning approach for abdomen CT image segmentation that can be applied to radiotherapy treatment planning. Our approach can effectively capture the underlying structure of the data by learning to predict relationships between different views of the same image, without relying on explicit annotations. We demonstrate the effectiveness of our approach on a publicly available dataset of abdomen CT scans. Our proposed approach achieves state-of-the-art performance on multiple segmentation tasks, including liver, spleen, and kidney segmentation. The results show that self-supervised learning can be a promising alternative to supervised approaches for organ segmentation in medical imaging, particularly when labeled data is limited. Our approach can significantly improve the efficiency and accuracy of organ segmentation, which is crucial for radiotherapy treatment planning.

2 Related Works

Abdomen CT image segmentation is an important task in medical imaging that can aid in the treatment planning of cancer. Automatic segmentation has been approached in various ways, ranging from traditional image processing methods to advanced deep-learning architectures. These methods have shown promising results in improving the accuracy and efficiency of medical diagnosis and treatment planning. Some of the most relevant and recent works of literature on CT image segmentation using different types of approaches are discussed here.

Statistical models involve the process of co-registering images in a training dataset to establish anatomical correspondences. A statistical model of the distribution of shapes and/or appearances of corresponding anatomy in the training dataset is constructed and then fitted to new images to generate segmentation [4, 5]. On the other hand, multi-atlas label fusion methods register images in a training dataset for each new image and combine propagated reference segmentation to generate new segmentation [6]. However, both statistical models and multi-atlas methods face limitations in image registration accuracy. Although increasing the dataset may improve accuracy, it also results in higher computation costs. Despite extensive research, image registration remains a challenging task.

Deep learning has shown significant promise in CT image segmentation by enabling the automatic identification and accurate segmentation of anatomical structures with high precision and efficiency.[7]. Convolutional Neural Network (CNN) based models have become popular in segmentation problems [8] and performed better than ATLAS-based methods with better accuracy [9]. Auto-segmentation of organs has been implemented using various deep neural networks[10] and CNN models like UNet [11], ResNet [10], Dense V-Networks [12], etc. UNet has proved to be an improvement in segmentation with less memory requirement as only the important features are extracted instead of all and it performs faster than basic CNNs. Due to the scarcity of annotated medical images, a lesser number of medical images are available and training of a model needs to be done with this limited data.

The transfer learning approach can solve this limited data issue to an extent. This approach entails making use of a pre-existing model that has been trained on a different task,

and using it as a foundation for a new task, rather than starting from scratch and training a new model [13]. Segmentation of multiple organs is also possible at the same time [14]. Implementation of neural networks and training of models has become faster with transfer learning. It basically extends the knowledge gained from a different trained network to another network. In simple words, the knowledge gained by training on ”X” can be used to train a model for ”Y”. Self-supervised learning is an approach in machine learning where a model is trained to predict certain properties or transformations of the data without the need for explicit labeling. This approach can be particularly useful for medical images, where labeled data can be scarce or expensive to obtain [15].

3 Methodology

The proposed model workflow is illustrated in Figure 1, which comprises data collection and pre-processing, self-supervision model training, segmentation model training, and testing steps. The data collection and pre-processing steps involve gathering CT image data from patients along with corresponding masks for model training. The pre-processing step analyzes the data and converts it into a 3-dimensional vector format suitable for model training. It also performs normalization of vector values, scaling them into a range between 0 to 1 for consistency. The proposed workflow combines self-supervised learning with a UNet-based segmentation model for accurate organ segmentation in CT images of the abdomen region.

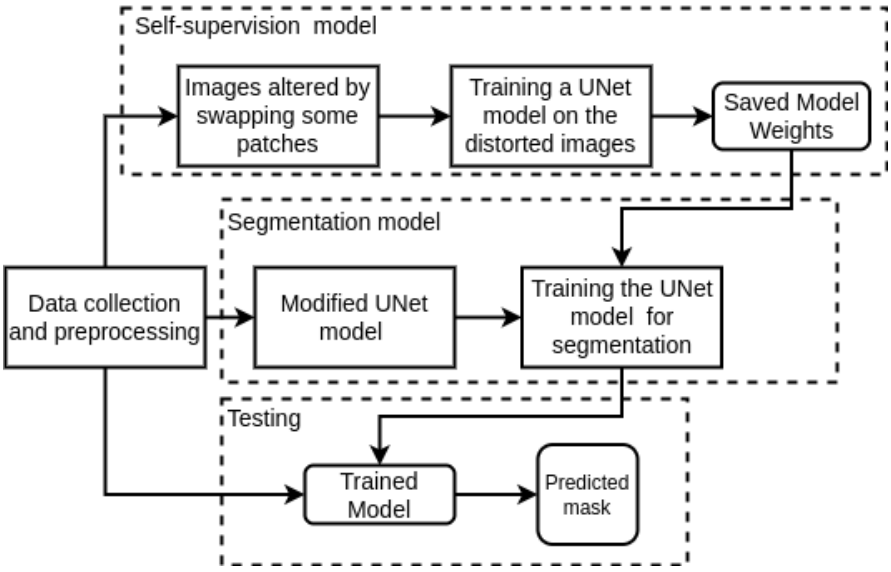

Figure 1. Workflow diagram

After preprocessing, the data is used to train the self-supervision model by applying some self-supervision tasks. The weights of this trained model are then loaded onto the segmentation model, which is trained for segmentation using the same data along with manual annotations. Since UNet models have shown remarkable results in medical image segmentation, we applied a modified UNet-based architecture for segmentation. We enhanced the standard network by incorporating new techniques to improve segmentation performance. The trained and modified UNet model is then saved and utilized for making predictions.

3.1 Dataset

The dataset used in this study was collected from the AbdomenCT-1K official repository [16].It comprises 1000 3D CT images with manual annotations of four organs: liver, kidneys, spleen, and pancreas. The CT images are stored in NIfTI format and have resolutions of 512 × 512 pixels, with a varying number of slices. The data used for model training consists of 800 CT images with corresponding masks, while 100 CT images are reserved for validation, and another 100 for testing. Figure 2 shows a sample CT slice from the dataset with its corresponding ground truth mask.

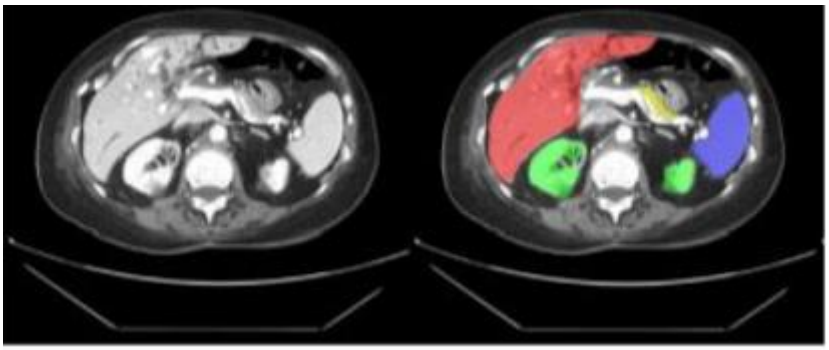

Figure 2. A sample CT image (left) and ground truth mask (right) from the dataset

3.2 Self-supervision

Self-supervised learning can be a powerful approach for addressing the issue of limited labeled data in medical image segmentation. It allows leveraging the abundance of unlabeled data to learn generic features that can be transferred to downstream tasks. One self-supervised method involves removing patches of images from the unlabelled dataset, and training a Unet model to fill in the missing patches [17]. The UNet model learns to exploit the relationship between various input image features to make accurate predictions. The learned weights from the self-supervised model are then used to initialize the weights of the segmentation model, which is then fine-tuned using the complete labeled dataset.

3.3 UNet architecture

The UNet architecture is so called because of its U-shaped symmetric structure. It has a contracting path and an expanding path made of convolutional layers. It works by classifying the image voxel by voxel. Hence the input and output images are the same size. To improve the segmentation performance of UNet a modified structure of UNet is proposed which has been shown in Figure 3. The modified structure consists of residual layers in the encoder part and attention gates in the skip connections of the Unet.

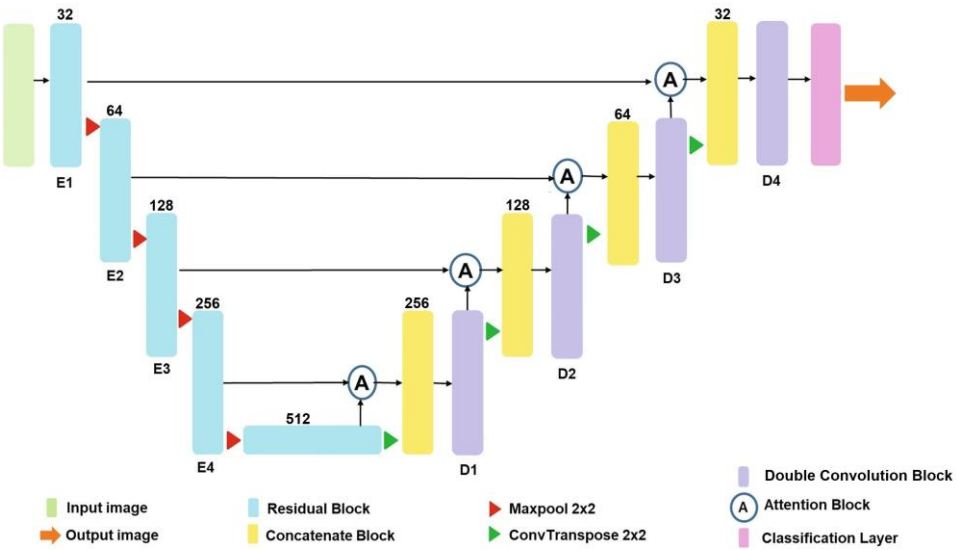

Figure 3. The complete model of Attention ResUNet. The number of channels is denoted at the top of the layers.

The standard UNet architecture is modified by introducing the residual blocks instead of basic convolution blocks in the encoding layers. The residual blocks help to overcome the vanishing gradient problem of deep networks. Figure 4 describes a residual block with its layers. Connections in the CNN are skipped i.e each layer feeds to the layer which is 2-3 layers away [18]. The model can be trained easily and the loss can be reduced with optimal weights. It also addresses the vanishing gradient problem where the loss function calculated shrinks to zero hence no learning is performed. The general output of a network layer is as follows,

$$yl = h(xl) + F(xl, wl) \quad (1)$$

$$xl+1 = f(yl) \quad (2)$$

where l is the lth layer of the network, F(.) is the residual function f(y) is the activation function and h(x) is an identity mapping function.

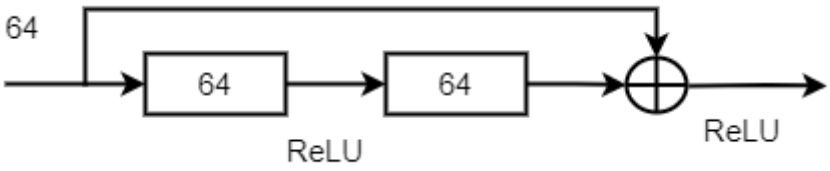

Figure 4. A residual block in a ResNet model. The map dimension can be 64, 128, 256, or 512

To focus on the relevant information in an image we use an attention gate. Attention gates suppress irrelevant features and highlight important features [19]. There is an attention coefficient α which ranges from 0-1. The attention gate described in Figure 5 takes 2 inputs, 1st is the gating signal from the lower layer and the 2nd is from the skip connection [20]. The formula for the output layer is as follows:

$$xl+1 = xl.\alpha \quad (3)$$

where l is the input layer, l+1 is the output layer.

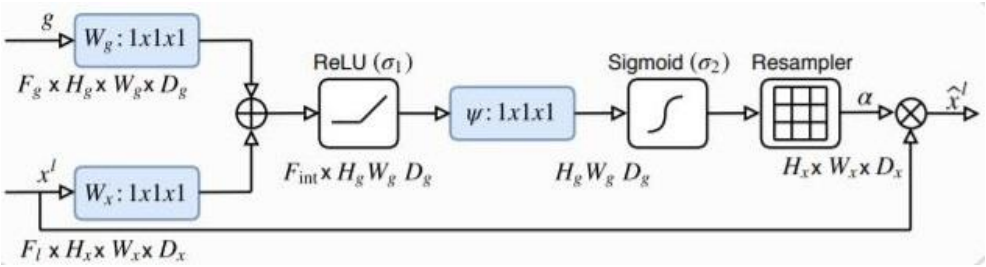

Figure 5. Basic Attention gate mechanism [20]. g is the gating signal, xl is the input layer features.

4 Implementation and Training Details odolog

The model is implemented using PyTorch which is an open-source machine-learning framework for computer vision and natural language processing based on the Torch library. SimpleITK package is used for processing 3D 'nii' images. Various pre-processing methods can be used which are provided by this library. Matplotlib library is used for visualizing data and results. It is also used for plotting various graphs. The model was trained on an Nvidia DGX station with 4 A100 GPU cards having 40 GB of graphics memory.

The self-supervision model has been trained on 200 images where small patches have been removed from these images and the same images have been given as labels for the model. After training the model the weights of this model have been used for training our modified UNet model. The model is trained for 50 epochs using Adam optimiser having a learning rate of 0.0001. The model achieves a mean squared error of 0.004.

The segmentation UNet model has been trained on 800 images with 100 images for validation. After loading the self-supervision network weights, the model was trained for 100 epochs with a batch size of one. The loss function used is dice loss, and the optimiser is Adam, and the achieved a training loss of 0.029 and a validation loss of 0.061. The training curves of the self-supervision and segmentation models are illustrated in Figure 6. Figure (a) presents the training loss details of the self-supervision model over 50 epochs, utilizing the mean square error as the applied loss function. On the other hand, Figure (b) displays the training curve of the segmentation model, exhibiting the train loss and validation loss for a training duration of 100 epochs. The weight updation process of the segmentation model employed the Dice loss. Notably, both curves indicate a parallel progression, ensuringing the absence of overfitting issues.

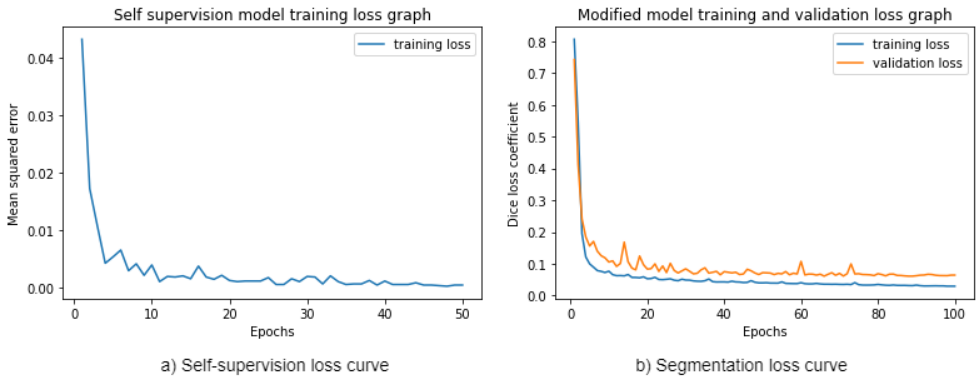

Figure 6. Loss curves a) Self-supervision model b) Segmentation model

5 Results

The output of the self-supervision model has been tested on 50 CT images. A sample output from the tested images has been shown in Figure 7 a). Dice Similarity Coefficient metric is used to evaluate the performance of the model, which is a commonly used metric for evaluating the performance of image segmentation algorithms. It measures the similarity between the predicted segmentation mask and the ground truth mask. The Dice coefficient is particularly useful for evaluating segmentation tasks where the foreground is relatively small compared to the background. The Dice coefficient is defined as twice the intersection of the predicted and ground truth masks divided by the sum of their sizes:

$$Dice = (2 * |P*G|) / (|P| + |G|) \quad (4)$$

Where P and G represent predicted and ground truth masks, respectievely.

The segmentation model was tested on 100 CT images and the average dice score of these images has been recorded. The average dice score for liver, spleen, kidneys, and pancreas has been given in Table 1. The table also gives a comparison between state-of-the-art models including, nnUNet [16], Attention UNet [21], and MargExcIL [22]. The proposed model achieved comparable Dice values to other models. Figure 7 b) shows the results of testing the model on a single CT image. The output is very similar to the ground truth. The inference time is around 10 seconds, which is significantly faster than the manual segmentation approach.

Table 1. Comparison of Dice scores of our model with other models

Organ	nnUNet	UNet	ResUNet	Attention UNet	MargExcIL	Proposed
liver	**0.979**	0.943	0.956	0.945	0.978	0.966
kidneys	**0.974**	0.898	0.930	0.900	0.939	0.955
spleen	**0.973**	0.936	0.944	0.933	0.969	0.965
pancreas	0.825	0.685	0.719	0.712	**0.839**	0.827

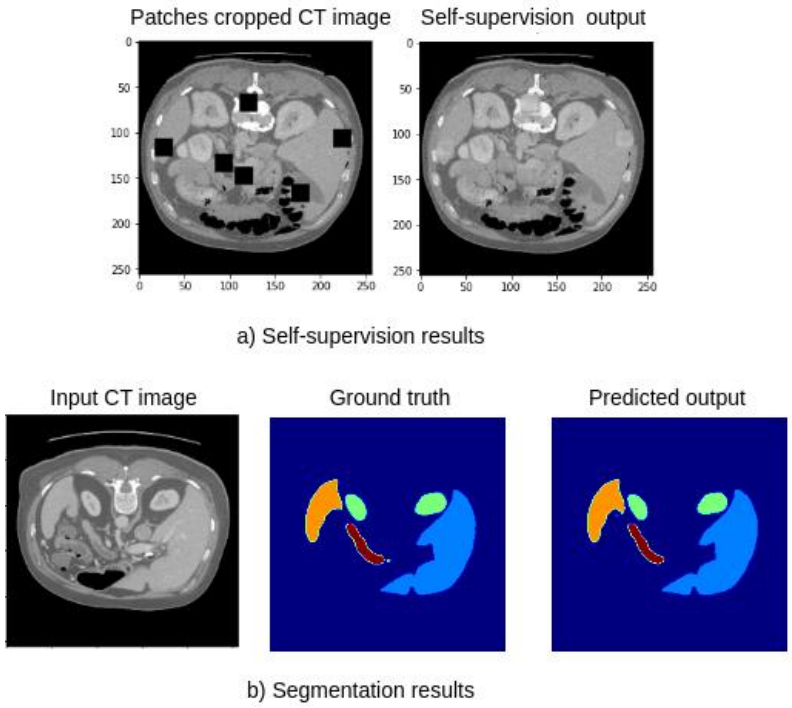

Figure 7. Visualisation of self-supervision and segmentation results

6 Conclusion

Deep learning methods can be utilized to develop various tools and algorithms for segmenting CT images. Our model outperforms most of the existing models and is capable of working on dataset with limited labelled images. The average dice score of our model is 0.9283, which indicates accurate results for most CT image segmentation tasks. The automatic segmentation of abdominal CT images can significantly reduce the workload on doctors and assist them in developing effective treatment plans for patients. The segmentation time has reduced significantly to around 5 seconds, which can speed up the treatment process considerably. The accuracy of the model can be further improved with the help of transformer networks.

References

1. X. Liu, K.W. Li, R. Yang, L.S. Geng, Frontiers in Oncology 11, 717039 (2021)
2. B. Qiu, H. van der Wel, J. Kraeima, H.H. Glas, J. Guo, R.J. Borra, M.J.H. Witjes, P.M. van Ooijen, Journal of personalized medicine 11, 629 (2021)
3. A. Jaiswal, A.R. Babu, M.Z. Zadeh, D. Banerjee, F. Makedon, Technologies 9, 2 (2020)
4. T. Zhang, Y. Yang, J. Wang, K. Men, X. Wang, L. Deng, N. Bi, Medicine 99 (2020)

5. Y. Zhou, J. Bai, IEEE Transactions on Information Technology in Biomedicine 11, 348 (2007)
6. X. Han, M.S. Hoogeman, P.C. Levendag, L.S. Hibbard, D.N. Teguh, P. Voet, A.C. Cowen, T.K. Wolf, Atlas-based auto-segmentation of head and neck CT images, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part II 11 (Springer, 2008), pp. 434–441
7. Y. Fu, Y. Lei, T. Wang, W.J. Curran, T. Liu, X. Yang, Physica Medica 85, 107 (2021)
8. X. Liang, N. Li, Z. Zhang, J. Xiong, S. Zhou, Y. Xie, Medical Image Analysis 73, 102156 (2021)
9. J. Zhu, J. Zhang, B. Qiu, Y. Liu, X. Liu, L. Chen, Acta Oncologica 58, 257 (2019)
10. Y. Chen, D. Ruan, J. Xiao, L. Wang, B. Sun, R. Saouaf, W. Yang, D. Li, Z. Fan, Medical physics (2020)
11. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015" (Springer International Publishing, 2015), pp. 234–241
12. E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S.P. Pereira, M.J. Clarkson, D.C. Barratt, IEEE transactions on medical imaging 37, 1822 (2018)
13. Q. Chen, M. Bernard, J. Duan, X. Feng, International Journal of Radiation Oncology*Biology*Physics 111, e125 (2021)
14. X. Chen, S. Sun, N. Bai, K. Han, Q. Liu, S. Yao, H. Tang, C. Zhang, Z. Lu, Q. Huang et al., Radiotherapy and Oncology 160, 175 (2021)
15. L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, D. Rueckert, Medical image analysis 58, 101539 (2019)
16. J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu et al., IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 6695 (2021)
17. H. Zheng, J. Han, H. Wang, L. Yang, Z. Zhao, C. Wang, D.Z. Chen, Hierarchical self-supervised learning for medical image segmentation based on multi-domain data aggregation (2021)
18. M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, arXiv preprint arXiv:1802.06955 (2018)
19. J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, IEEE Access PP, 1 (2020)
20. O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz et al., Attention u-net: Learning where to look for the pancreas (2018), 1804.03999
21. C.E. Lee, M. Chung, Y.G. Shin, Voxel-level siamese representation learning for abdominal multi-organ segmentation (2022)
22. P. Liu, X. Wang, M. Fan, H. Pan, M. Yin, X. Zhu, D. Du, X. Zhao, L. Xiao, L. Ding et al., Learning incrementally to segment multiple organs in a ct image (2022)