

Boundary-Aware Network for Abdominal Multi-Organ Segmentation

Shishuai Hu, Zehui Liao, and Yong Xia✉

National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
yxia@nwpu.edu.cn

Abstract. Automated abdominal multi-organ segmentation is a crucial yet challenging task in the computer-aided diagnosis of abdominal organ-related diseases. Although numerous deep learning models have achieved remarkable success in many medical image segmentation tasks, accurate segmentation of abdominal organs remains challenging, due to the varying sizes of abdominal organs and the ambiguous boundaries among them. In this paper, we propose a boundary-aware network (BA-Net) to segment abdominal organs on CT scans and MRI scans. This model contains a shared encoder, a boundary decoder, and a segmentation decoder. The multi-scale deep supervision strategy is adopted on both decoders, which can alleviate the issues caused by variable organ sizes. The boundary probability maps produced by the boundary decoder at each scale are used as attention to enhance the segmentation feature maps. We evaluated the BA-Net on the Abdominal Multi-Organ Segmentation (AMOS) Challenge dataset and achieved an average Dice score of 89.29% for multi-organ segmentation on CT scans and an average Dice score of 71.92% on MRI scans. The results demonstrate that BA-Net is superior to nnUNet on both segmentation tasks.

Keywords: Boundary-aware network · Medical image segmentation · Multi-organ segmentation

1 Introduction

Accurate abdominal anatomy segmentation using computerized tomography (CT) or magnetic resonance imaging (MRI) provides crucial information such as the interrelations among the organs as well as individual positions and shapes in the standardized space, playing an essential role in the computer-aided diagnosis applications such as disease diagnosis and radiotherapy planning [4]. Since manual segmentation of abdominal organs is time-consuming and requires high concentration and expertise, automated segmentation methods are highly demanded to accelerate this process. However, this task remains challenging due to two reasons: (1) abdominal organ regions, *e.g.*, liver and adrenal glands, vary significantly in the volumes, as shown in Fig. 1 (b); and (2) the contrast between

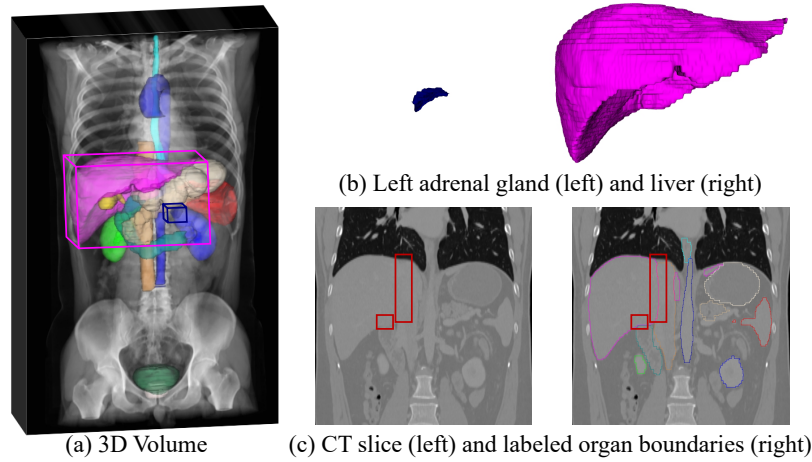


Fig. 1. Illustrations of (a) 3D abdominal organ volumes, (b) imbalanced organ volumes, *e.g.*, left adrenal gland and liver, and (c) ambiguous boundaries between different organs and their surroundings.

different organs and their anatomical surroundings is particularly low, resulting in the blurry and ambiguous boundary, as shown in Fig. 1 (c).

With the recent success of deep convolutional neural networks (DCNN) in many vision tasks, a number of DCNN-based methods have been proposed for abdominal organ segmentation. Milletari *et al.* [8] proposed a simple but efficient V-shape network and a Dice loss for volumetric medical image segmentation. Based on that, Isensee *et al.* [3] integrated the data attributes and network design using empirical knowledge and achieved promising performance on the abdominal organ segmentation task. Following the same philosophy, Peng *et al.* [9] and Huang *et al.* [2] searched networks from data directly using NAS and further enhanced the accuracy of abdominal organ segmentation. Although achieved promising performance, these methods [8,3,9,2] ignore the boundary information, which is essential for abdominal organ segmentation. Compared to the largely imbalanced volume proportions of different organs' regions, the boundary (*a.k.a.* surface in 3D) proportions of different organs are less sensitive to the varying sizes. Therefore, a lot of boundary-involved segmentation methods have been recently proposed for multi-organ segmentation. Kervadec *et al.* [7] introduced a boundary loss that uses a distance metric on the space of boundaries for medical image segmentation to enhance the segmentation accuracy near the boundary. Karimi *et al.* [6] proposed a hausdorff distance loss to reduce the hausdorff distance directly to improve the similarity between the predicted mask and ground truth. Shit *et al.* [11] incorporated morphological operation into loss function calculation of thin objects, and proposed a cDice loss for accurate boundary segmentation. Jia *et al.* [5] introduced a 2D auxiliary boundary detection branch that shares the encoder with the 3D segmentation branch and

achieved superior performance on MRI prostate segmentation. Despite the improved performance, the boundary-involved loss function-based methods [7,6,11] are time-consuming during training since these boundary-involved loss functions are computing-unfriendly, *i.e.*, the distance map of each training image mask should be computed during each iteration. Whereas the previous auxiliary boundary task-based method [5] does not fully utilize the extracted boundary information since the auxiliary boundary branch will be abandoned at inference time.

In this paper, we propose a boundary-aware network (BA-Net) based on our previous work [1] for abdominal multi-organ segmentation. The BA-Net is an encoder-decoder structure [10]. To force the model to pay more attention to the error-prone boundaries, we introduce an auxiliary boundary decoder to detect target organs’ ambiguous boundaries. Compared to our previous work in [1], both the boundary and segmentation decoders are supervised at each scale of the decoder in this paper to further improve the model’s robustness to varying sizes of target organs. Also, we modified the feature fusion mechanism and used the detected boundary probability maps as attention maps on the corresponding segmentation features instead of directly concatenating the boundary feature maps with segmentation feature maps. We have evaluated the proposed BA-Net model on the AMOS Challenge training dataset using 4-fold cross-validation and achieved a Dice score of 89.29% on average using CT scans and a Dice score of 71.92% on average using MRI scans.

2 Dataset

The large scale abdominal multi-organ segmentation dataset published by AMOS challenge¹ [4] was used for this study. The AMOS dataset contains 500 CT and 100 MRI scans with voxel-level annotations of 15 abdominal organs, including the spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, pancreas, right adrenal gland, left adrenal gland, duodenum, bladder, and prostate/uterus. Among all the scans, 200 CT and 40 MRI scans are provided as training cases, 100 CT and 20 MRI scans are used for online validation, and the left 200 CT and 40 MRI scans are withheld for final evaluation. Only the voxel-level annotations of training cases are publicly available, while the annotations of validation cases and testing cases cannot be accessed.

3 Method

The proposed BA-Net contains a shared encoder, a boundary decoder, and a segmentation decoder, as shown in Fig. 2. The encoder extracts features from the input images. Then the boundary decoder performs target organ boundary detection using the extracted features. The segmentation decoder takes the extracted features as input and outputs segmentation maps of target organs,

¹ <https://amos22.grand-challenge.org/>

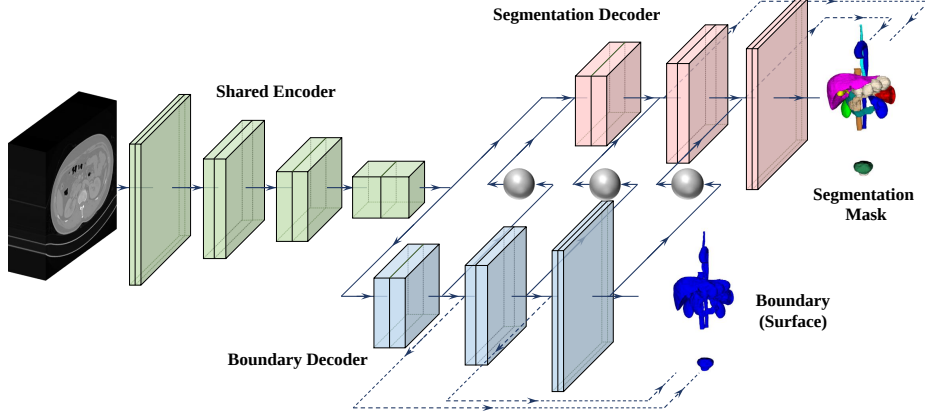


Fig. 2. Illustration of BA-Net for abdominal multi-organ segmentation. The green, blue, and pink layers represent convolutional layers of the encoder, boundary decoder, and segmentation decoder respectively. The gray nodes represent the soft-max operations. The dashed lines in both decoders represent deep supervision. The skip connections from the shared encoder to both decoders are omitted for simplicity.

wherein the boundary probability maps produced by the boundary decoder are adopted as attention maps to enhance the segmentation features.

3.1 Shared Encoder

Since nnUNet [3] performs well on abdominal multi-organ segmentation, we adopt it as the backbone of our BA-Net. Therefore, the shared encoder design is the same as nnUNet, which is generated from the statistics of the training data. It contains N encoder blocks (N depends on the specific imaging modalities), each of which is composed of 2 convolutional layers. Each convolutional layer is followed by instance normalization and the LeakyReLU activation. The down-sample operations are performed using strided convolutions, followed after each encoder block. In the encoder, the number of convolutional filters is set to 32 in the first layer, then doubled in each next block, and finally fixed with 320 when it becomes larger than 256.

3.2 Boundary Decoder

Symmetrically, the boundary decoder contains N decoder blocks to upsample the feature map extracted by the encoder and gradually refine it to generate the boundary map. The transposed convolution with a stride of 2 is used to improve the resolution of input feature maps. The upsampled feature map is concatenated with the low-level feature map extracted from the corresponding encoder block and then fed to the decoder block. The output feature map f_i^b of each decoder

block is processed by a convolutional layer and a soft-max layer to generate the boundary probability map p_i^b at each scale i , where $i = 1, 2, \dots, N - 1$.

3.3 Segmentation Decoder

The structure of the segmentation decoder is similar to the boundary decoder, except for an additional step to enhance the segmentation feature map at each scale. In each block of the segmentation decoder, the upsampled feature map $U(f_{i-1}^s)$ is enhanced with the corresponding boundary probability map p_i^b , shown as follows

$$U(\widehat{f_{i-1}^s}) = (1 + p_i^b) * U(f_{i-1}^s) \quad (1)$$

where $U(\widehat{f_{i-1}^s})$ is the enhanced feature map, and $U(\cdot)$ represents upsample operation.

The enhanced feature map is concatenated with the feature map extracted from the corresponding encoder block before being further processed by the segmentation decoder block. Similar to the boundary decoder, the output feature map f_i^s of each segmentation decoder block is also processed by a convolutional layer and a soft-max layer to generate the segmentation probability map p_i^s at each scale i .

3.4 Training and Inference

The combination of Dice loss and cross-entropy loss is adopted as objective for both the boundary detection and organ segmentation tasks. The joint loss at each scale can be calculated as

$$\begin{aligned} \mathcal{L}_i = & 1 - \frac{2 \sum_{v=1}^V p_{iv} y_{iv}}{\sum_{v=1}^V (p_{iv} + y_{iv} + \epsilon)} \\ & - \sum_{v=1}^V (y_{iv} \log p_{iv} + (1 - y_{iv}) \log (1 - p_{iv})) \end{aligned} \quad (2)$$

where p_{iv} and y_{iv} denote the prediction and ground truth of the v -th voxel in the output of the i -th decoder block, V represents the number of voxels, and ϵ is a smooth factor to avoid dividing by 0. For the segmentation task, p_i is the segmentation probability map p_i^s , and y_i is the downsampled segmentation ground truth y_i^s . For the boundary detection task, p_i is the boundary probability map p_i^b , and y_i is the boundary y_i^b of the segmentation ground truth y_i^s .

We use deep supervision for both decoders to ensure the boundary can be detected at each scale to enhance the segmentation feature map, and improve the model's robustness to targets' sizes. Totally, the loss is defined as

$$\mathcal{L} = \sum_{i=1}^{N-1} \omega_i (\mathcal{L}_i^s + \mathcal{L}_i^b) \quad (3)$$

where ω_i is a weighting vector that enables higher resolution output to contribute more to the total loss [3], $\mathcal{L}_i^s$ is the segmentation loss, and $\mathcal{L}_i^b$ is the boundary detection loss.

During inference, given a test image, the shared encoder extracts features at each scale, and then each boundary decoder block processes these features and produces a boundary probability map. Based on the extracted features and the boundary probability maps, the segmentation decoder is only required to output the segmentation result of the last decoder block.

4 Experiments and Results

4.1 Implementation Details and Evaluation Metric

The AMOS challenge contains two tasks: (1) segmentation of abdominal organs using only CT scans; and (2) segmentation of abdominal organs using both CT and MRI scans.

In both tasks, the Hounsfield Unit (HU) values of CT scans were clipped to the range of $[-991, 373]$ according to the data statistics, and then subtracted the mean and divided by the standard deviation. For task 1, following the nnUNet [3] framework, we trained the 3D low resolution model (*i.e.*, BA-Net-L), which takes the input image with a voxel size of $3.3mm \times 1.7mm \times 1.7mm$ as input, and the 3D high resolution model (*i.e.*, BA-Net-H), which takes the input image with a voxel size of $2.0mm \times 0.7mm \times 0.7mm$ as input. The segmentation results produced by BA-Net-L and BA-Net-H were averaged to produce the final result (*i.e.*, BA-Net-E) for task 1. For task 2, the CT scans and the MRI scans were preprocessed by different policies due to their different imaging mechanisms. The preprocess workflow and parameters for CT scans were the same as that of task 1. The MRI scans were resampled to a uniform voxel size of $2.0mm \times 1.2mm \times 1.2mm$, and then normalized by subtracting the mean and dividing by the standard deviation. The preprocessed CT volumes and MRI volumes were fed to a uniform model (*i.e.*, BA-Net-U) randomly so that the model can perform abdominal multi-organ segmentation for both imaging modalities. The segmentation results produced by BA-Net-U were adopted as the final results for MRI scans. Whereas the segmentation results produced by BA-Net-L in task 1 and BA-Net-U in task 2 were averaged as the final results (*i.e.*, BA-Net-R) for CT scans in task 2. Limited to the GPU memory, the size of the input image volume was set to $64 \times 160 \times 160$ and the batch size was set to 2. The number of down-sampling layers N was set to 5 for both tasks. The SGD algorithm was used as the optimizer. The initial learning rate lr_0 was set to 0.01 and decayed according to $lr = lr_0 \times (1 - t/T)^{0.9}$, where t is the current epoch and $T = 1000$ is the maximum epoch. The whole framework was implemented in PyTorch and trained using an NVIDIA 2080Ti.

The segmentation performance was measured by the Dice Similarity Coefficient. All experiments were performed using 4-fold cross-validation with no postprocessing.

Table 1. Performance (%) of nnUNet and our BA-Net in abdominal multi-organ segmentation on **CT** scans of **task 1**. The best results are highlighted with **bold**.

Methods	Spleen	Right Kidney	Left Kidney	Gall Bladder	Esophagus	Liver	Stomach	Aorta
nnUNet-L	96.42	95.66	94.62	81.43	84.04	97.44	90.07	94.88
nnUNet-F	95.85	95.46	94.40	83.29	84.61	97.28	90.42	95.40
nnUNet-E	96.33	96.03	94.78	83.84	85.29	97.68	90.77	95.63
BA-Net-L	95.91	95.88	94.85	83.25	84.22	97.57	90.32	94.87
BA-Net-F	96.34	95.70	94.56	83.22	84.93	97.46	90.77	95.69
BA-Net-E	96.46	96.25	95.03	84.16	85.46	97.79	91.11	95.75
Methods	Postcava	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Duodenum	Bladder	Prostate /Uterus	Average
nnUNet-L	90.41	84.41	76.39	77.26	80.73	88.99	84.80	87.84
nnUNet-F	91.04	85.97	78.77	79.87	82.06	88.69	84.14	88.48
nnUNet-E	91.38	85.94	78.77	79.83	82.30	90.09	85.77	88.96
BA-Net-L	90.54	84.89	76.41	77.16	80.78	90.27	85.35	88.15
BA-Net-F	91.42	86.56	79.11	80.06	82.54	89.67	86.65	88.98
BA-Net-E	91.58	86.43	78.95	79.97	82.55	91.13	86.80	89.29

Table 2. Performance (%) of nnUNet and our BA-Net in abdominal multi-organ segmentation on **CT** scans of **task 2**. The best results are highlighted with **bold**.

Methods	Spleen	Right Kidney	Left Kidney	Gall Bladder	Esophagus	Liver	Stomach	Aorta
nnUNet-L	96.42	95.66	94.62	81.43	84.04	97.44	90.07	94.88
nnUNet-U	95.96	95.47	94.35	83.05	84.56	97.20	90.29	95.35
nnUNet-R	96.35	96.05	94.84	83.09	85.28	97.65	90.75	95.56
BA-Net-L	95.91	95.88	94.85	83.25	84.22	97.57	90.32	94.87
BA-Net-U	96.20	95.70	94.84	83.46	84.73	97.44	90.45	95.56
BA-Net-R	96.43	96.26	95.13	84.27	85.36	97.74	90.98	95.71
Methods	Postcava	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Duodenum	Bladder	Prostate /Uterus	Average
nnUNet-L	90.41	84.41	76.39	77.26	80.73	88.99	84.80	87.84
nnUNet-U	90.86	85.81	78.40	79.31	81.99	88.71	84.52	88.39
nnUNet-R	91.28	85.86	78.59	79.50	82.28	90.01	85.74	88.85
BA-Net-L	90.54	84.89	76.41	77.16	80.78	90.27	85.35	88.15
BA-Net-U	91.21	86.33	78.59	79.45	82.06	89.70	85.53	88.75
BA-Net-R	91.46	86.20	78.51	79.52	82.29	91.00	86.60	89.16

Table 3. Performance (%) of nnUNet, BA-Net, and their variants in abdominal multi-organ segmentation on **MRI** scans of **task 2**. The best results are highlighted with **bold**.

Methods	Spleen	Right Kidney	Left Kidney	Gall Bladder	Esophagus	Liver	Stomach	Aorta
nnUNet-F (CT→MRI)	55.96	39.23	62.03	26.30	29.82	36.96	49.68	65.09
nnUNet-M	97.14	94.93	93.87	75.27	67.07	97.96	84.95	92.10
nnUNet-U	93.59	95.36	94.80	77.75	66.23	97.91	87.35	92.15
nnUNet-D	94.07	95.45	95.04	77.01	66.19	97.89	87.92	92.23
BA-Net-U	93.62	95.59	95.02	74.58	67.31	98.10	88.12	92.52

Methods	Postcava	Pancreas	Right Adrenal Gland	Left Adrenal Gland	Duodenum	Bladder	Prostate/Uterus	Average
nnUNet-F (CT→MRI)	51.08	40.85	20.61	34.94	34.54	0.00	0.00	36.47
nnUNet-M	88.73	84.82	63.43	60.51	69.13	0.00	0.00	71.33
nnUNet-U	88.08	84.65	64.62	63.65	71.39	0.00	0.00	71.84
nnUNet-D	88.45	84.26	64.36	62.35	72.07	0.00	0.00	71.82
BA-Net-U	88.46	85.17	64.54	63.04	72.67	0.00	0.00	71.92

4.2 Results

We compared the proposed BA-Net with nnUNet [3] on both tasks. The performance of nnUNet and BA-Net on CT scan segmentation for task 1 is shown in Table 1. It shows that the overall performance of our BA-Net is superior to nnUNet in both the low-resolution model (nnUNet-L v.s. BA-Net-L) and high-resolution model (nnUNet-H v.s. BA-Net-H). It demonstrates that our BA-Net can effectively improve the segmentation performance on CT scans. Table 1 also reveals that the averaged performance of our BA-Net (BA-Net-E) achieves the best result, confirming the importance of taking boundary into consideration for model design.

The performance of nnUNet and BA-Net on CT scans segmentation for task 2 is shown in Table 2. It shows that the joint training of CT scans and MRI scans brings no benefit to CT scan segmentation (nnUNet-F v.s. nnUNet-U). Still, our BA-Net achieves better performance than nnUNet on every organ.

The performance of nnUNet, BA-Net, and their variants on MRI scan segmentation for task 2 is shown in Table 3. We also trained a nnUNet model only on MRI scans, denoted as nnUNet-M. Based on the uniform nnUNet model for both CT scans and MRI scans, we introduced domain-specific convolution to process the CT volumes and MRI volumes respectively in the first encoder block, denoted as nnUNet-D. Also, we directly generate the segmentation results for MRI scans using nnUNet-F that trained using CT scans, denoted as nnUNet-F (CT→MRI), to quantize the domain gap between MRI and CT scans.

Table 3 shows that the joint training can improve the segmentation performance on MRI scans (nnUNet-M v.s. nnUNet-U). We believe that plenty of CT data for training improves the model’s feature extraction ability, which enhances the segmentation performance on limited MRI scans. Surprisingly, Table 3 reveals that the domain-specific convolution makes no contribution to the MRI segmentation performance. We assume that preprocessing CT scans and MRI scans in different ways can effectively reduce the domain gap between CT and MRI (nnUNet-F v.s. nnUNet-U). It reveals in Table 3 that our BA-Net still can improve the segmentation performance on MRI scans (nnUNet-U v.s. BA-Net-U).

5 Conclusion

In this paper, we propose a BA-Net for abdominal multi-organ segmentation. It is composed of a shared encoder for feature extraction, a boundary decoder for boundary detection, and a segmentation decoder for target organ segmentation. At each scale of the boundary decoder and the segmentation decoder, the boundary and segmentation mask at the corresponding scale are used to supervise the predicted probability map. The predicted boundary map at each decoder block is used as an attention map to enhance the segmentation features. Experimental results on the AMOS challenge dataset demonstrate that our BA-Net is superior to nnUNet on CT scan segmentation and MRI scan segmentation.

References

1. Hu, S., Zhang, J., Xia, Y.: Boundary-aware network for kidney tumor segmentation. In: Liu, M., Yan, P., Lian, C., Cao, X. (eds.) MICCAL 2020 - MLMI. pp. 189–198. Lecture Notes in Computer Science, Springer International Publishing. https://doi.org/10.1007/978-3-030-59861-7_20
2. Huang, Z., Wang, Z., zhikai yang, Gu, L.: Adwu-net: Adaptive depth and width u-net for medical image segmentation by differentiable neural architecture search. In: Medical Imaging with Deep Learning (2022), <https://openreview.net/forum?id=kF-d1SKWJpS>
3. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-net: a self-configuring method for deep learning-based biomedical image segmentation **18**(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z>, <https://www.nature.com/articles/s41592-020-01008-z>
4. Ji, Y., Bai, H., Yang, J., Ge, C., Zhu, Y., Zhang, R., Li, Z., Zhang, L., Ma, W., Wan, X., Luo, P.: AMOS: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation . <https://doi.org/10.48550/arXiv.2206.08023>, <http://arxiv.org/abs/2206.08023>
5. Jia, H., Song, Y., Huang, H., Cai, W., Xia, Y.: Hd-net: hybrid discriminative network for prostate segmentation in mr images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 110–118. Springer (2019)
6. Karimi, D., Salcudean, S.E.: Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on medical imaging* **39**(2), 499–513 (2019)

7. Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, I.B.: Boundary loss for highly unbalanced segmentation. In: International conference on medical imaging with deep learning. pp. 285–296. PMLR (2019)
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
9. Peng, C., Myronenko, A., Hatamizadeh, A., Nath, V., Siddiquee, M.M.R., He, Y., Xu, D., Chellappa, R., Yang, D.: Hypersegmas: Bridging one-shot neural architecture search with 3d medical image segmentation using hypernet. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20741–20751 (June 2022)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. pp. 234–241. Lecture Notes in Computer Science, Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
11. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16560–16569 (2021)