Semi-supervised organ segmentation with Mask Propagation Refinement and Uncertainty Estimation for Data Generation

Minh-Khoi Pham[1,2][0000-0003-3211-9076], Thang-Long Nguyen-Ho[1,2][0000-0003-1953-7679] *, and Minh-Triet Tran[1,2,3][0000-0003-3046-3041]

[1] University of Science, Ho Chi Minh City, Vietnam
[2] Viet Nam National University, Ho Chi Minh City, Vietnam
[3] John von Neumann Institute, Ho Chi Minh City, Vietnam

Abstract. We present a novel two-staged method that employs various 2D-based techniques to deal with the 3D segmentation task. In most of the previous challenges, it is unlikely for 2D CNNs to be comparable with other 3D CNNs since 2D models can hardly capture the temporal information. In light of that, we propose using the recent state-of-the-art technique in video object segmentation, combining with other semi-supervised training techniques to leverage the extensive unlabeled data. Moreover, we also generate pseudo-labeled data that is both plausible and consistent for further retraining by using uncertainty estimation. Overall, our method achieves ... on the validation set of FLARE22.

Keywords: 2D semi-supervised segmentation · Mask Propagation · Uncertainty Estimation

1 Introduction

Organs segmentation hold an important step in clinical stage because it affects factors such as abnormal organ detection, disease diagnosis, etc. However, quantifying agencies accurately is quite expensive and time consuming, while medical labels need to be evaluated and labeled by experts to ensure accuracy before releasing a valuable data set to the community. Because of the lack of data or human resources with high expertise in labeling, or the trade off between spending too much time on labeling, the work of medical staff must always be prioritized, so it is very difficult to generate data when human resources are scarce.

In the supervised scenario, that is time consuming to manually annotate the pixel locations of interest. As for the unsupervised lack in the low-quality segmentation mask, it is difficult to apply in real scenario. With the semisupervise approach, we can take advantage of the labeled data combined with the unlabelled data, the approach still provide accurate evaluation based on the labeled data.

* The first two authors share the equal contribution.

However, in any scenario, data labeling must be done manually and sequentially to ensure absolute accuracy in medical data. But now, state of the art method are gradually proposed, but there is not aim to the reusability or adaptability, so when labeling a completely new data, the work does not take advantage of the knowledge from previous models. Our propose is not just about how we utilize unlabelled data, but also the potential of how we can reuse it to help with efficient data generation.

The objective of our approach is to propose a novel method for interactive scenario, which is a semi-automated process include interact and refine, this labeling process speeds up the annotate process. By recommending propagation module using binary manner learning strategy, we do not need to care about the number of classes when doing the inference stage. The model can understand the context or class we want to reference, and then infer the entire volume. Proposed architecture design referencer that takes advantage of unlabelled data and suggests the propagation model valuable slices. The propagation result is satisfying all designed goals, including quickly, propagated less often to satisfy results, the results are always exactly comparable to the reference level.

2 Method

2.1 Preprocessing

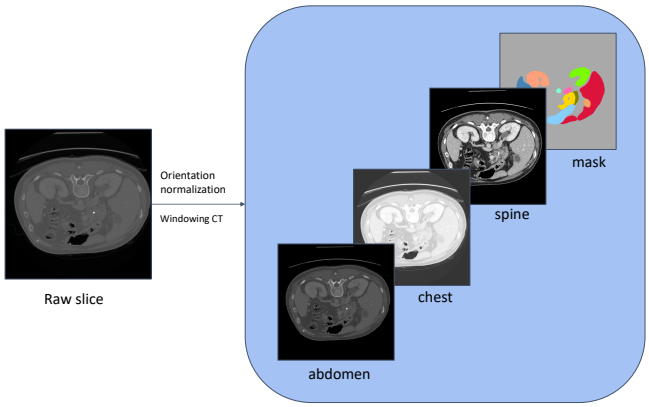

Fig. 1. Windowing CT

With some organs that are relatively small, it might be better to keep the original size of the slices without any cropping, resampling or resizing methods.

Because the Hounsfield Unit scale has a very wide range, and might not perform well with the approach on conventional visual models, we apply a window

based on prior knowledge to display map to 256 range. The motivation is with disease tissue it is only visible when applying windows, if we use raw data, the diseased tissue is mixed with normal tissue, it is almost impossible to extract information from the visual image.

Windowing [2], also known as grey-level mapping, contrast stretching, histogram modification or contrast enhancement is the process in which the CT image grayscale component of an image is manipulated via the CT numbers; doing this will change the appearance of the picture to highlight particular structures. The brightness of the image is adjusted via the window level. The contrast is adjusted via the window width. In or experiments, we create 3 different versions of a single slice by highlight the abdomen, chest, and spine groups and stack it to one as a three-channel image (Fig. 1).

In addition, we choose the axial plane to cut the slices from the CT volumes since this plane has various dimension sizes. The image is rotated to a predefined angle, then is divided by 255 for normalization before going through the next step.

2.2 Proposed Method

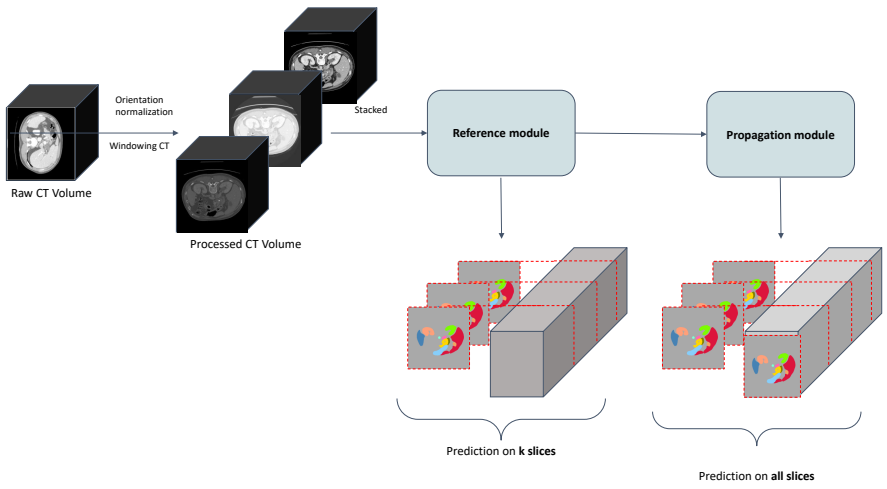

Fig. 2. Our overall proposed pipeline. Firstly, the entire CT Volume is processed using windowing CT to get a stack of three-channel slices. Then the slices progress through the reference modules to obtain minimal number of preliminary masks. Lastly, the Propagation module refine these initial masks to finalize the result.

Our method composes of two main modules: Reference module and Propagation module, as can be seen in Fig 2.

In the beginning, we heuristically select only a small number of slices from the CT Volume to be our initial candidates. Next, these slices are processed by using Windowing technique 2. Afterwards, these slices are put through the Reference module, in which performs the standard multiclass segmentation, then the preliminary k masks can be obtained. Based on these pairs of potential slices and masks as prior knowledge, the Propagation module can utilize them to propagate the objects transformation information to the remaining slices across the CT volume length. The final output of this module is a 3D dense mask prediction, with each voxel indicating a class.

Reference module This module is expected to provide a suggestion of a minimal amount of slices and predicted masks that might contains the most information describing the entire CT Volume. Fig 3 describes the details of this module.

To utilize the enormous number of unlabeled data, we apply the recent semi-supervised method that performs effectively on several other datasets, which is called Cross Pseudo Supervision (CPS) [7] (yellow cube in Fig. 3. CPS enables the usage of unlabeled data by following the dual students technique, where two models are trained simultaneously on labeled data while generating pseudo data for their "peer" to learn. In the testing phase, two models predict on the same image and the result is aggregated by summing up.

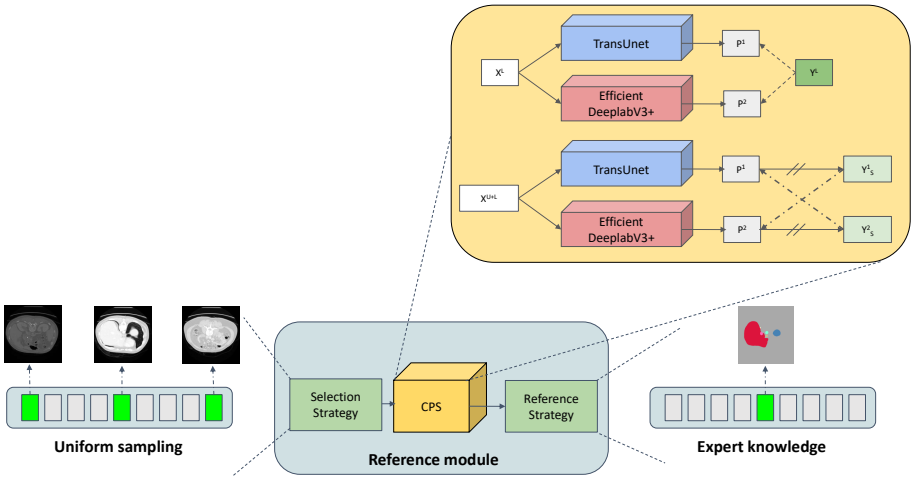

Fig. 3. The reference module. The semi-supervised technique CPS is applied in both training and inference stage to enhance the precision of model prediction. Strategies are used to smartly choose slices that are informative for the next stage.

We adopt two prominent state-of-the-arts 2D segmentation models with highly different learning paradigms for this CPS framework, which is TransUNet [5] and DeeplabV3+ [6].

While DeeplabV3+ traditionally focus more on the local information, transformers model the long-range relation, so the cross training can help to learn a unified segmenter with these two properties at the same time. In short, we choose TransUNet and DeeplabV3+ due to the their ability to compensate each other for better performance. [13]

In addition, we also propose a both logical and specialist-based strategy to choose which slices that can be further used to boost the performance of the Propagation module. The goal of this action is to preserve only some of the most useful information for the refinement stage.

To elaborate on these strategies, prior to being put into the CPS module for prediction, a small number of slices are uniformly sampled from the processed CT volume. After CPS produces segmentation masks for these slices, another selection step is performed to picking only some of the masks that contains the organs having the largest areas.

Propagation module This module aims to utilize prior knowledge of given annotated slices from the Reference module to make prediction on the remaining slices, this mechanism can be referred as mask (or label) propagation.

Intuitively, the conventional 2D CNNs cannot comprehend the third dimension information within a CT volume. Thus, in hope of the ability to capture the "temporal" information along the axial plane, we adapt the Space-Time Correspondence Networks (STCN) [8], which is a semi-supervised segmentation algorithm that has achieved promising results on Video object segmentation problem, to this 3D manner.

Basically, STCN proposes the use of memory bank that stores information of previous frames and their corresponding masks, and uses them later as prior knowledge. To generate the mask for the current frame, a pairwise affinity matrix is calculated between the query frame and memory frames based on negative squared Euclidean distance, then it is used for supporting the current mask generation. [8]

Different from the original STCN, we slightly modify it to match the current problem. As the original work, they use only a single dense mask to propagate through the entire video, therefore for the model to perfectly work, that selected mask must contains information of all available classes. For our case to achieve that, we enable the usage of multiple masks for propagation, so that all of these mask should contain enough information of every organs. We also allow the STCN to work in a bidirectional way to enhance the refinement. Fig 4 illustrates this process.

Specially, STCN can be simply trained in the binary manner, meaning that each of the abdominal organs can be learned separately. Therefore, the knowledge can be transferred well between different organ classes.

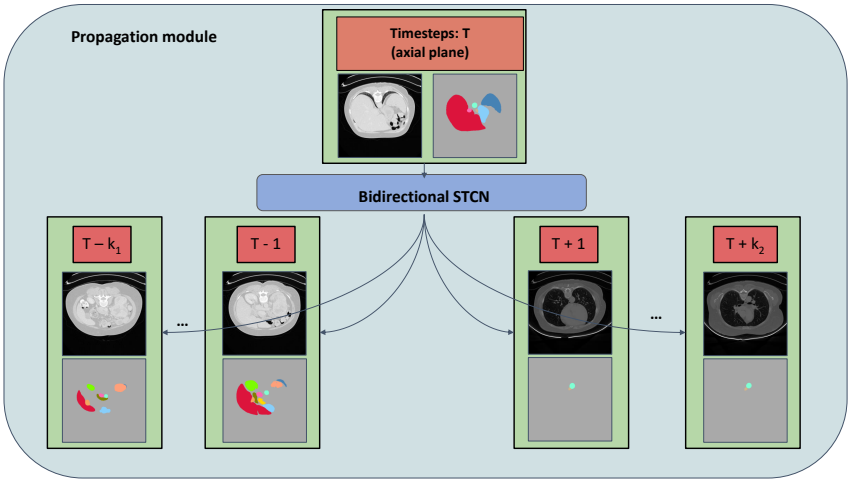

Fig. 4. The propagation module. From an annotated slice of CT, at timestep T, STCN can make use of that to spread the information through the entire defined range $[T - k_1, T + k_2]$.

Pseudo labeling with Uncertainty Estimation Given a vast amount of unlabeled CT volumes, we apply a uncertainty estimation technique to effectively maximize the utilization of the data.

Firstly, several CPS models are trained on the provided labeled data. Then, we use these trained CPS models to obtain pseudo masks on the unlabeled set. Inspired from [20], we calculate the dice scores between these pseudo masks and the aggregated one. The mean of these dice scores will be compared with a threshold to determine whether the pseudo masks are qualified. Simply speaking, consensus-based assessment is used to evaluate the quality of pseudo labels.

All samples that have high certainty are then reused for the next supervised training cycle. And after the training finishes, the same labeling process is repeated until all aforementioned models achieve satisfied performance or every unlabeled data has been used.

Loss function For the Reference module, we use the prevalent combination of dice loss and cross entropy loss with smoothing value to alleviate the imbalanced number of the small organs, which occurs due to our splitting into slices process. The same settings are used for CPS in its supervised branch whereas only the dice loss is setup for the unsupervised branch.

For the Propagation module, we implement the online hard example cross entropy (OhemCE or Bootstrapping CE) [21] and also calculate the Lovasz loss [3] at the same time. OhemCE can help reduce the contribution of background label to the final loss, Since STCN training on binary task, OhemCE can di-

rect the model to focus on visible difficult objects. Meanwhile, Lovasz loss is commonly used in the past.

2.3 Post-processing

We do not use any post-processing techniques because no complex pre-processing ones are used, and we conduct all our experiment on the nearly-original image volumes apart from the orientation settings. Thus, before submitting to the evaluation system, the mask must be transformed back to the original orientation.

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 [19], KiTS [10,11], AbdomenCT-1K [15], and TCIA [9]. 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 development environments and requirements are presented in Table 1.

Training protocols Currently, we find that using only simple 2D transform functions such as horizontal/vertical flipping or rotating might be enough for both modules to generalize.

4 Results & Discussion

Validation results

Qualitative results later

Table 1. Development environments and requirements.

Windows/Ubuntu version	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz
RAM	1×32GB;
GPU (number and type)	One Quadro RTX 5000 16G
CUDA version	11.6
Programming language	Python 3.10
Deep learning framework	Pytorch (Torch 1.11.0, torchvision 0.12.0)
Specific dependencies	
(Optional) Link to code	

Table 2. Training protocols for Reference module: CPS of TransUnet and Efficientnet DeeplabV3+

Network initialization	Random initialization
Batch size	2 (labeled) + 2 (unlabeled)
Patch size	None
Total iterations	50000
Optimizer	AdamW
Initial learning rate (lr)	0.0001
Lr decay schedule	multiplied by 0.5 for every iteration at $[40000, 45000]$
Training time	48 hours
Number of model parameters	117,597,362 [4]
Number of flops	[5]
CO_2eq	κg[6]

Table 3. Training protocols for Propagation module: STCN with Resnet backbone

Network initialization	Random initialization
Batch size	8
Patch size	None
Total iterations	50000
Optimizer	AdamW
Initial learning rate (lr)	0.0001
Lr decay schedule	multiplied by 0.5 for every iteration at $[40000, 45000]$
Training time	48 hours
Number of model parameters	54,416,065 [7]
Number of flops	[8]
CO_2eq	κg[9]

Table 4. Ablation experiment on proposed modules and techniques

No.	CPS	Windowing CT	UE	MP	Mean DSC
1					0.547
2	✓				0.762
3	✓	✓	✓		0.770
4	✓	✓	✓	✓	0.784

Limitation and future work Apparently, although our proposed method has yet to achieve the high result, we believe it can be further improved if these limitation that we identify here are solved. First of all, the problem of imbalanced dataset has arisen because we perceive this as a 2D problem. Due to the slices splitting process, small organs (such as pancreas, gallbladder or adrenal glands) only appear in a small amount of slices, while larger objects have wider range of appearance. Therefore, it leads to the problem of imbalanced dataset. We tried some ways to tackle the problem, for instance: smart sampling, or imbalanced loss, however only slightly improvement was seen. Secondly, the proposed approach is a two-stage method, the second stage is undoubtedly dependent of the first one. If there are any organs that are missed by the reference module, it definitely cannot be recovered in the propagation phase. In the future, it is encouraged to focus on boosting the performance of the reference module by fully exploiting the temporal information.

5 Conclusion

In summary, we present a two-stage pipeline, which can leverage the strength of many state-of-the-art 2D deep learning algorithms and techniques in videos and images, into the task of 3D object segmentation. Our proposal aims to introduce a novel and inspirational approach in solving one of the most common problem in the medical field.

6 Acknowledgements

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2022 challenge has not used any pretrained models nor additional datasets other than those provided by the organizers.

Furthermore, no manual intervention has been made in the contribution to the results of the proposed method

References

1. NIH Pancreas. https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT (2020), [Online; Accessed: Aug. 2020]

2. Baba, Y., Murphy, A.: Windowing (ct) (Mar 2017). https://doi.org/10.53347/rid-52108, http://dx.doi.org/10.53347/rID-52108 3
3. Berman, M., Triki, A.R., Blaschko, M.B.: The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 4413–4421. Computer Vision Foundation / IEEE Computer Society (2018). https://doi.org/10.1109/CVPR.2018.00464, http://openaccess.thecvf.com/content_cvpr_2018/html/Berman_The_LovaSz-Softmax_Loss_CVPR_2018_paper.html 6
4. Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056 (2019)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. CoRR **abs/2102.04306** (2021), https://arxiv.org/abs/2102.04306 5
6. Chen, L., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VII. Lecture Notes in Computer Science, vol. 11211, pp. 833–851. Springer (2018). https://doi.org/10.1007/978-3-030-01234-2_49, https://doi.org/10.1007/978-3-030-01234-2_49 5
7. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. pp. 2613–2622. Computer Vision Foundation / IEEE (2021), https://openaccess.thecvf.com/content/CVPR2021/html/Chen_Semi-Supervised_Semantic_Segmentation_With_Cross_Pseudo_Supervision_CVPR_2021_paper.html 4
8. Cheng, H.K., Tai, Y., Tang, C.: Rethinking space-time networks with improved memory coverage for efficient video object segmentation. In: Ranzato, M., Beygelzimer, A., Dauphin, Y.N., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual. pp. 11781–11794 (2021), https://proceedings.neurips.cc/paper/2021/hash/61b4a64be663682e8cb037d9719ad8cd-Abstract.html 5
9. 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) 7
10. 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) 7
11. 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) 7

12. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
13. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-supervised medical image segmentation via cross teaching between CNN and transformer. CoRR **abs/2112.04894** (2021), https://arxiv.org/abs/2112.04894 5
14. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021)
15. 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 7
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241 (2015)
17. Roth, H., Farag, A., Turkbey, E., Lu, L., Liu, J., Summers, R.: Data from pancreas-ct. the cancer imaging archive (2016)
18. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 556–564. Springer (2015)
19. 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) 7
20. Wang, J., Chen, Z., Wang, L., Zhou, Q.: An active learning with two-step query for medical image segmentation. In: 2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE). pp. 1–5. IEEE (2019) 6
21. Wu, Z., Shen, C., van den Hengel, A.: High-performance semantic segmentation using very deep fully convolutional networks. CoRR **abs/1604.04339** (2016), http://arxiv.org/abs/1604.04339 6