Med-DANet: Dynamic Architecture Network for Efficient Medical Volumetric Segmentation – Supplementary Material

Wenxuan Wang[1], Chen Chen[2], Jing Wang[1], Sen Zha[1], Yan Zhang[1], and Jiangyun Li[1,†]

[1] School of Automation and Electrical Engineering, University of Science and Technology Beijing, China, s20200579@xs.ustb.edu.cn, m202120718@xs.ustb.edu.cn, g20198675@xs.ustb.edu.cn, m202110578@xs.ustb.edu.cn, leejy@ustb.edu.cn

[2] Center for Research in Computer Vision, University of Central Florida, USA, chen.chen@crcv.ucf.edu

[3] † Corresponding author: Jiangyun Li

Overview

In this supplementary material, we provide the following items:

1. (Sec. 1) Experiments on the LiTS 2017 dataset [1] (Liver Tumor Segmentation using **CT** scans), *revealing the generalization ability of our proposed dynamic architecture framework (i.e. Med-DANet) on other medical imaging modalities (e.g. Computed Tomography (CT)) for the segmentation task.*
2. (Sec. 2) More ablation study and analysis on BraTS 2019 and 2020 datasets for a comprehensive investigation.
3. (Sec. 3) More visual comparison of brain tumor segmentation for qualitative analysis.
4. (Sec. 4) Discussion about the broader impact and limitation of this work.

1 Experimental Results on LiTS 2017

In order to show that our proposed insight is not just limited to segmentation task for MRI brain tumors but is a common phenomenon in different medial image modalities, we also present the image content distribution along the slice dimension of a 3D **CT** case from the LiTS 2017 dataset [1] in Fig. 1. CT is another widely used imaging modality for various medical applications. Obviously, Fig. 1 shows that the image content also varies significantly across different CT slices, which is similar to the distribution across diverse MRI slices. It is also evident that the segmentation difficulties are different among CT slices. Therefore, it is reasonable to adjust the model complexity according to different inputs (e.g. image slices) for effective accuracy and efficiency trade-offs.

To evaluate the generalization ability of our proposed Med-DANet, we conduct experiments of liver tumor segmentation on CT scans using the LiTS 2017 dataset [1].

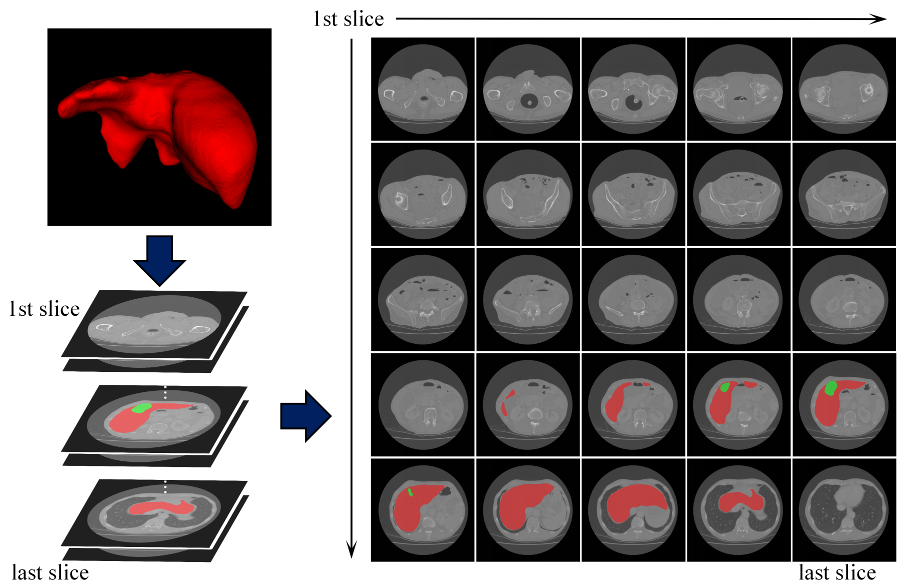

Fig. 1. The illustration of image content distribution along slice dimension of a CT case (Axial View) from the LiTS 2017 dataset [1]. The red regions denote the liver and the green regions denote the tumors.

The quantitative results on LiTS 2017 testing set are presented in Table 1. It can be clearly seen that our method achieves comparable or even higher Dice scores than previous state-of-the-art methods with much less model complexity. Note that most of the comparison methods didn't provide the source codes. Therefore, we can not obtain the computational costs of those methods. In comparison with recently proposed Transformer based method named TransBTS [12]) for medical image segmentation task (the source code of TransBTS is publicly available), our Med-DANet considerably advances the segmentation accuracy with greatly reduced computational costs. Specifically, the computational complexity of TransBTS [12] is **8.89** times that of our Med-DANet, which is similar to the situation on BraTS 2019 and BraTS 2020 datasets. Thus, the results confirm the generalization ability and effectiveness of our adaptive framework with dynamic architecture.

2 More Ablation Study and Analysis

In this section, to further explore the potential of our dynamic framework and justify the rationale of its design choices, more ablation experiments are conducted. (1) We investigate the effect of different training strategies for training the model candidates in our Model Bank. Experiments are carried out using five-fold cross-validation evaluations on the BraTS 2019 training set. (2) We present

Table 1. Performance comparison on LiTS 2017 testing set. “P” refers to pre-trained model. Per case and per slice denote the computational cost of segmenting a 3D patient case and a single 2D slice, respectively.

Method	Dice per case (%) ↑		Dice global (%) ↑		FLOPs (G) ↓	
	Lesion	Liver	Lesion	Liver	Per Case	Per Slice
U-Net [5]	65.00	-	-	-	-	-
3D DenseUNet w/o P [8]	59.40	93.60	78.80	92.90	-	-
2D DenseUNet w/o P [8]	67.70	94.70	80.10	94.70	-	-
2D DenseNet w/ P [8]	68.30	95.30	81.80	95.90	-	-
2D DenseUNet w/ P [8]	70.20	95.80	**82.10**	96.30	-	-
I3D [3]	62.40	95.70	77.60	96.00	-	-
I3D w/ P [3]	66.60	95.60	79.90	96.20	-	-
Han [7]	67.00	-	-	-	-	-
Vorontsov et al. [11]	65.00	-	-	-	-	-
TransUNet [4]	61.70	95.40	77.40	95.60	1200.64	9.38
Swin-UNet [2]	-	92.70	67.60	91.60	249.60	1.95
TransBTS [12]	70.30	96.00	81.50	96.40	330.00	2.58
Ours	**70.50**	**96.10**	81.90	**96.60**	**37.12**	**0.29**

the selection ratio of each candidate model in the Model Bank for the BraTS 2019 and 2020 datasets.

2.1 Effect of Different Training Strategies for Candidate Networks

In order to employ the most suitable and efficient training approach for our Med-DANet, we investigate different strategies to train the candidate models in the model bank of our proposed framework. Since the Model Bank is composed of four candidate networks, the simple and straightforward way of training these candidates would be training each candidate individually (i.e. individual training). However, this individual training scheme is time-consuming. Therefore, in our proposed Med-DANet, we simultaneously train all the candidate networks together in a joint fashion (i.e. joint training). The comparison of the segmentation performance and training time for these two training strategies is shown in Table 2. The joint training method can greatly reduce the training time (i.e. up to **7.36 hours**) while achieving higher model accuracy. However, the individual training scheme yields better computational efficiency in terms of FLOPs.

Table 2. Ablation study on effect of different training strategies for candidate networks.

Training Strategy	Dice Score (%) ↑			FLOPs (G) ↓		Training Time (hour) ↓
	ET	WT	TC	Per Case	Per Slice	
Med-DANet(Individual Training)	75.73	90.25	82.31	**962.87**	**7.52**	21.99
Med-DANet(**Joint Training**)	**78.75**	**90.40**	**83.13**	1,551.485	10.01	**14.63**

2.2 Activation Ratio of Each Candidate Network

We further illustrate the activation ratio (or selection ratio) of each candidate model in the model bank for 3D volumentric segmentation in a slice-by-slice manner. Fig. 2 and Fig. 3 show the activation ratio of each candidate model on BraTS 2019 and BraTS 2020 dataset, respectively. During each inference, a

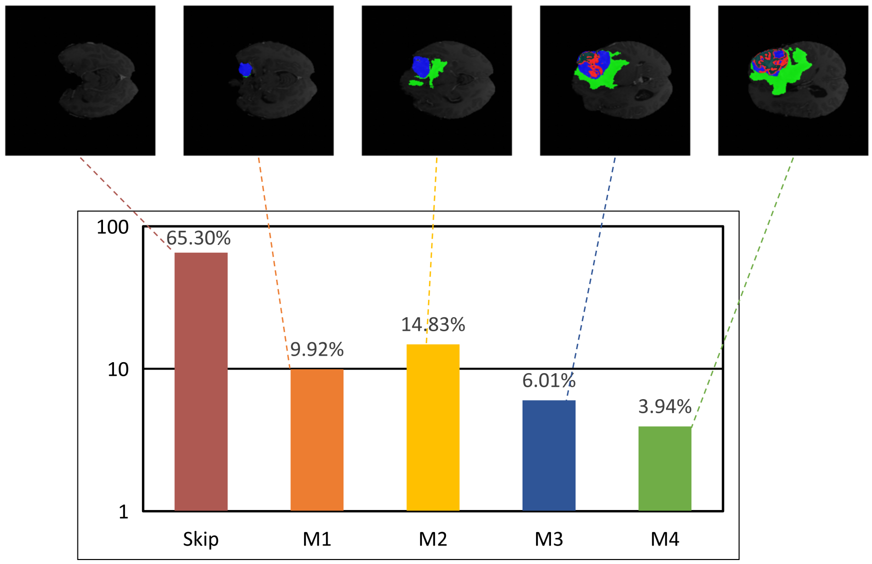

Fig. 2. The activation ratio of each candidate model for different medical image slices in BraTS 2019 dataset. Skip, M1, M2, M3, M4 denote the operation of directly skip, candidate 1, candidate 2, candidate 3, and candidate 4, respectively.

single candidate in the Model Bank is activated according to the segmentation difficulty of the current slice. We can observe that the direct skipping operation accounts for a large portion for the MRI slices (i.e. more than half of the total number of slices). Moreover, since there are relatively more simple slices (e.g. those only contain one or two types of tumors and with small tumor regions) than the difficult slices, lightweight models (i.e. M1 and M2) are activated more over the large models such as M3 and M4. Through the proposed dynamic selection mechanism, an highly efficient and powerful architecture is achieved by our Med-DANet to reach a good balance between accuracy and computational efficiency.

3 More Visual Comparison for Brain Tumor Segmentation

To further demonstrate the advantage of our proposed dynamic framework, we present more visualization of brain tumor segmentation results on BraTS 2019 for qualitative analysis in Fig. 4. The different methods utilized for visual comparison consist of 3D U-Net [6], V-Net [9], Attention U-Net [10], and our Med-DANet. It is clear from Fig. 4 that our framework can segment different kinds of brain tumors more precisely and generate much better fine-grained segmentation masks.

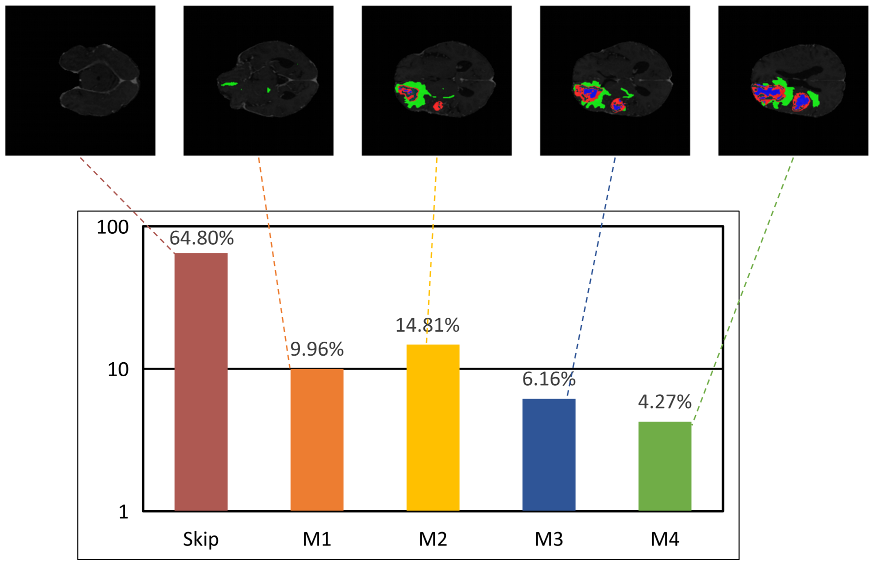

Fig. 3. The activation ratio of each candidate model for different medical image slices in BraTS 2020 dataset. Skip, M1, M2, M3, M4 denote the operation of directly skip, candidate 1, candidate 2, candidate 3, and candidate 4, respectively.

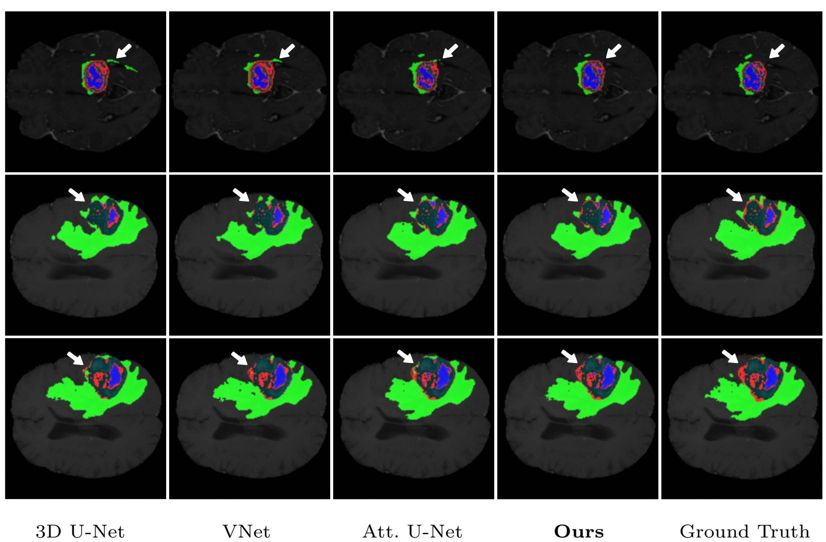

Fig. 4. More visual comparison of MRI brain tumor segmentation results on BraTS 2019. The blue regions denote the enhancing tumors, the red regions denote the non-enhancing tumors, and the green ones denote the peritumoral edema.

4 Broader Impact and Limitation

Our approach provides a novel solution to efficient volumetric segmentation for medical applications, which inspires new research in this direction. Moreover, it is generalizable to other volumetric data (e.g. CT). One potential limitation could be the increased training cost due to several networks in the Model Bank, as compared to single network training. However, our joint training strategy and the multi-GPU training paradigm can greatly alleviate this issue. It also provides a future research direction to develop more efficient training schemes to match the cost of single network training.

References

1. Bilic, P., Christ, P., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C., Han, X., Heng, P., Hesser, J., et al.: The liver tumor segmentation benchmark (lits). arxiv 2019. arXiv preprint arXiv:1901.04056 (2019) 1, 2
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv preprint arXiv:2105.05537 (2021) 3
3. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017) 3
4. 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. arXiv preprint arXiv:2102.04306 (2021) 3
5. Chlebus, G., Meine, H., Moltz, J.H., Schenk, A.: Neural network-based automatic liver tumor segmentation with random forest-based candidate filtering. arXiv preprint arXiv:1706.00842 (2017) 3
6. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016) 4
7. Han, X.: Automatic liver lesion segmentation using a deep convolutional neural network method. arXiv preprint arXiv:1704.07239 (2017) 3
8. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE transactions on medical imaging **37**(12), 2663–2674 (2018) 3
9. 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) 4
10. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018) 4
11. Vorontsov, E., Tang, A., Pal, C., Kadoury, S.: Liver lesion segmentation informed by joint liver segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1332–1335. IEEE (2018) 3
12. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: Transbts: Multimodal brain tumor segmentation using transformer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–119. Springer (2021) 2, 3