Supplementary Material for Point-Unet: A Context-aware Point-based Neural Network for Volumetric Segmentation

Ngoc-Vuong Ho[1], Tan Nguyen[1], Gia-Han Diep[3], Ngan Le[4], and Binh-Son Hua[1,2]

[1] VinAI Research, Vietnam
{v.vuonghn,v.tannh10,v.sonhb}@vinai.io
[2] VinUniversity
[3] University of Science, VNU-HCM, Vietnam
han.diep@ict.jvn.edu.vn
[4] Department of Computer Science at University of Arkansas in Fayetteville
thile@uark.edu

1 Pancreas Results

We provide full comparisons on Pancreas dataset in Table 1, which extends the results reported in the main paper.

Table 1: Dice score comparison on Pancreas dataset.

Method	Average ↑	Method	Average ↑
Zhou et al. [15]	82.37 ± 5.68	Roth et al. [10]	78.01 ± 8.20
Roth et al. [9]	71.42 ± 10.11	Roth et al. [11]	81.27 ± 6.27
Oktay et al. [8]	83.10 ± 3.80	Zhang et al. [13]	77.89 ± 8.52
Cai et al. [1]	82.40 ± 6.70	Zhou et al. [15]	82.37 ± 5.68
Zhu et al [16]	84.59 ± 4.86	Dou et al. [2]	82.25 ± 5.91
Ours	**85.68** ±5.96	Yu et al. [12]	84.50 ± 4.97

2 Ablation study

We conducted an ablation study on the offline validation set of BraTS20. We provided variants of our model to demonstrate the significance of the volumetric saliency attention network, point-based segmentation network, and the GDL loss. The results are shown in Table 2, which demonstrates the importance of our proposed saliency network and point-based segmentation.

Particularly, model A is our reported model in Table 3 in the main paper. By removing the saliency attention network and only use point-based segmentation, we obtain model B that resembles RandLANet, the accuracy of which drops

Table 2: Ablation study. We show dice measurements on variants of our method on the BraTS 2020 offline validation set.

Method	Dice score ↑			
	ET	WT	TC	AVG
E: 3D U-Net [5]	66.92	82.86	72.98	74.25
D: Attention 3D U-Net [14]	69.87	89.68	79.28	79.61
C: Ours without GDL Loss	69.33	89.36	69.51	76.06
B: Ours without saliency (RandLANet)	67.40	87.74	76.85	77.33
A: Ours (Point-Unet)	76.43	89.67	82.97	83.02

significantly (6%). Model B was reported in Table 3 (paper) as RandLANet [3]. To verify the effectiveness of the generalized dice loss (GDL), we remove GDL and use cross entropy in the point-based network (model C), which also results in accuracy loss by almost 7%. Compared to the traditional U-Net (3D U-Net [5], model E) and an attention-based U-Net which we extended from [14] (model D) that employs only volumetric segmentation, which is equivalent to our method without the point-based segmentation, our method also outperforms significantly. This verifies that both components (volumetric saliency attention network and point-based segmentation network) together with GDL contribute significantly in our method.

3 Details on Point Sampling

We provide additional details for the point sampling step in our method. In general, our proposed method has two main steps: saliency attention prediction and point-based segmentation. In our training, we train the saliency network (voxel segmentation) and point segmentation separately. We first train the saliency network with the target labels as binary: foreground that are union of all ground truth tumor regions, and background for remaining voxels. Dice loss is used for training the saliency network. After training the saliency network, we perform point sampling based on thresholding the confidence output from the saliency network to establish point clouds for segmentation.

To generate the point cloud, we threshold the confidence output of the attention network (threshold 0.9), and voxels passing the threshold become foreground (FG) points. We randomly sample remaining voxels to obtain background (BG) points. The union of FG and BG points form an input point cloud for segmentation. Note that the FG already contains tumor regions, and the BG only provides additional context data for learning. The segmentation results of the FG can be simply used as the final tumor segmentation results, and *no resampling* from point cloud to volume is required. We also tested with different thresholds and found that it is quite insensitive to the model performance (1-percent difference when varying the threshold in [0.6, 0.95]) as shown in Table 3.

Table 3: Evaluation with different confidence thresholds for our point sampling scheme. The results are on BraTS 2020 offline set.

Threshold	Dice score ↑			
	ET	WT	TC	AVG
0.5	75.82	85.53	82.05	81.14
0.6	75.90	86.43	82.26	81.53
0.7	75.91	87.45	82.69	82.02
0.8	76.04	88.45	82.88	82.46
0.9	**76.43**	**89.67**	82.97	**83.02**
0.95	76.26	89.39	83.12	82.93
0.975	76.36	89.35	82.57	82.76

3.1 Random Sampling vs. Our Sampling

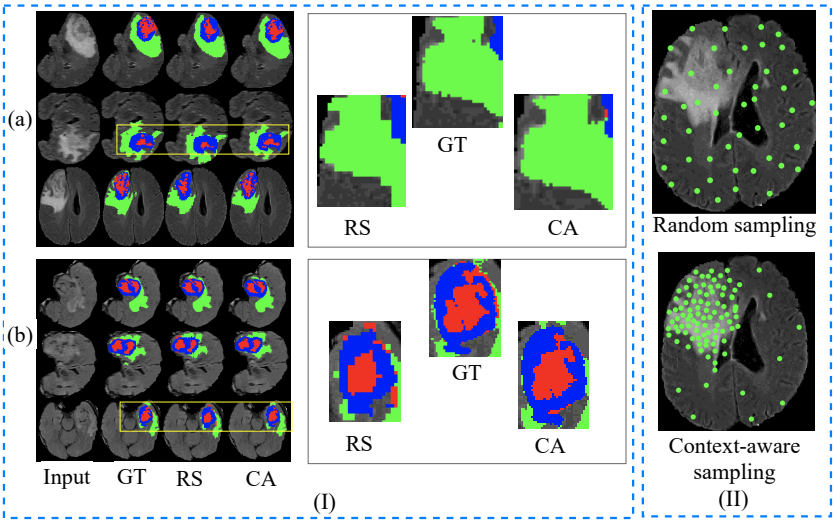

Fig. 1: Illustration on the *small objects*. **RS**: random sampling, **CA**: context-aware, **GT**: groundtruth. (a) and (b) are two different brain subjects. The enlarged view of the lesion is given on the right.

The comparison between random sampling and our proposed context-aware sampling is given in Figure 1(II). Random sampling treats every pixel in the same manner, thus there is no mechanism to pay attention to boundaries as well as small objects to address the above limitations. As given in Figure 1(I), random sampling produces the results in zigzag artifact at boundaries while the object surface plays an important role in medical analysis which aims to

understand the topologial strucutre. Furthermore, random sampling scheme may not always capture every sample in a population. This causes the missing of small objects during sampling and segmenting as given in Figure 1(I). Not only unable to address the aformentioned difficulties, random sampling also has some other restrictions. The limitations of random sampling is summarized as follows:

- Results in zigzag artifact at boundaries;
- Unable to capture small objects;
- Inference is costly: In order cover entire brain space, it requires performing random sampling serval times;
- Cannot guarantee the tumor area will be completely covered after many iterations at inference.

By contrast, our context-aware sampling addresses such issues by placing more samples at the region of interests while maintaining samples in the background regions.

4 Visualization Result on BraTS

Figure 2 illustrates volumetric segmentation at three planes (sagittal, coronal, axial) on different methods given an input. The proposed Point-Unet (in the last column) provided a better segmentation results, resulting in better Dice score and Hausdorff95 distance. As can be seen, our Point-Unet segmentation provides an improved segmentation along the tumor boundary (indicated by the pink arrows) than the existing SOTA methods.

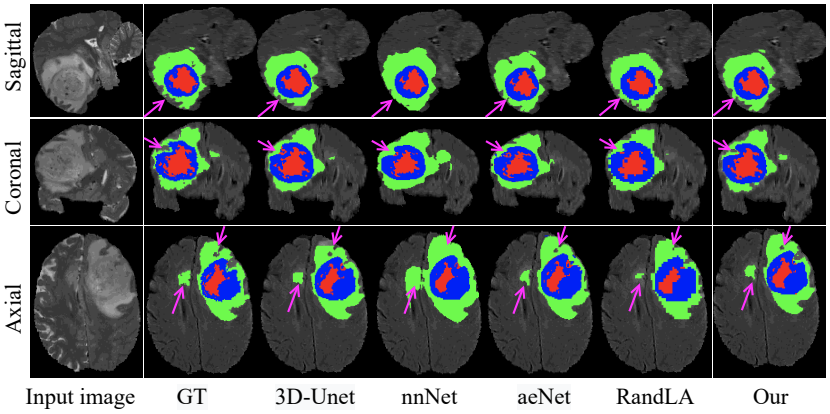

Fig. 2: Segmentation results on different planes (1^{st} row: Sagittal, 2^{nd} row: Coronal, 3^{rd}: Axial) generated by (a) source image (Flair modality); (b) groundtruth; (c) baseline 3DUnet[5]; (d) nnNet [4]; (e) aeUnet [7]; (f) RandLA-Net [6]; (g) our Point-Unet. Our method provides a good boundary on the tumor areas (pink arrows $\rightarrow$) compared with existing methods.

References

1. Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L.: Improving deep pancreas segmentation in ct and mri images via recurrent neural contextual learning and direct loss function. arXiv preprint arXiv:1707.04912 (2017)
2. Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.A.: 3d deeply supervised network for automatic liver segmentation from ct volumes. In: MICCAI. pp. 149–157. Springer (2016)
3. Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A.: Randla-net: Efficient semantic segmentation of large-scale point clouds. In: CVPR (2020)
4. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H., van Walsum, T.: No new-net. In: MICCAI Brainlesion (2019)
5. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al.: Ensembles of multiple models and architectures for robust brain tumour segmentation. In: MICCAI Brainlesion. Springer (2017)
6. Kuijf, H.J., Bennink, E.: Grand challenge on mr brain segmentation at miccai 2018 (2018), http://mrbrains18.isi.uu.nl/
7. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: MICCAI Brainlesion (2019)
8. Oktay, O., Schlemper, J., Folgoc, L.L., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
9. 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: MICCAI. pp. 556–564. Springer (2015)
10. Roth, H.R., Lu, L., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In: MICCAI. pp. 451–459. Springer (2016)
11. Roth, H.R., Lu, L., Lay, N., Harrison, A.P., Farag, A., Sohn, A., Summers, R.M.: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Medical image analysis **45**, 94–107 (2018)
12. Yu, Q., Xie, L., Wang, Y., Zhou, Y., Fishman, E.K., Yuille, A.L.: Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In: CVPR (2018)
13. Zhang, Y., Ying, M.T., Yang, L., Ahuja, A.T., Chen, D.Z.: Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: International Conference on Bioinformatics and Biomedicine (BIBM). pp. 443–448. IEEE (2016)
14. Zhao, T., Wu, X.: Pyramid feature attention network for saliency detection. In: CVPR (2019)
15. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L.: A fixed-point model for pancreas segmentation in abdominal ct scans. In: MICCAI (2017)
16. Zhu, Z., Xia, Y., Shen, W., Fishman, E., Yuille, A.: A 3d coarse-to-fine framework for volumetric medical image segmentation. In: 3DV. pp. 682–690. IEEE (2018)