A Bi-directional, Multi-modality Framework for Segmentation of Brain Structures

Skylar S. Gay(✉), Cenji Yu, Dong Joo Rhee, Carlos Sjogreen, Raymond P. Mumme, Callistus M. Nguyen, Tucker J. Netherton, Carlos E. Cardenas, and Laurence E. Court

Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
sgay1@mdanderson.org

Abstract. Careful delineation of normal-tissue organs-at-risk is essential for brain tumor radiotherapy. However, this process is time-consuming and subject to variability. In this work, we propose a multi-modality framework that automatically segments eleven structures. Large structures used for defining the clinical target volume (CTV), such as the cerebellum, are directly segmented from T1-weighted and T2-weighted MR images. Smaller structures used in radiotherapy plan optimization are more difficult to segment, thus, a region of interest is first identified and cropped by a classification model, and then these structures are segmented from the new volume. This bi-directional framework allows for rapid model segmentation and good performance on a standardized challenge dataset when evaluated with volumetric and surface metrics.

Keywords: MICCAI 2020 · ABCs · Deep learning · Segmentation

1 Introduction

Successful radiotherapy of brain tumors requires careful identification and delineation of normal organs-at-risk (OARs). Standard clinical practice involves manual delineation, a time-consuming process that is subject to intra- and inter-observer variability [1, 2]. Existing automatic segmentation approaches often fall within four categories: atlas based, machine learning based, deformable based, or region-based. Atlas based approaches involve large amount of registration, a computationally expensive process that is also dependent upon atlas anatomical similarity. Machine learning based techniques suffer from lack of global understanding of shape information, especially for complex structures with inhomogeneous textures. Deformable methods are sensitive to initialization, whereas region-based methods often fail at low contrast boundaries [3].

Recently, deep learning-based methods have demonstrated state-of-the-art performance in a variety of medical image segmentation tasks [4]. Deep neural network architectures such as U-Net [5], SegNet [6], and MeshNet [7] are proposed for segmentation of the brain itself and surrounding structures. However, these approaches often underperform for small structures in the brain whose classes are sufficiently underrepresented...
relatively to the background class. In this work, we therefore apply a two-stage approach to segmenting brain structures. Large structures are directly segmented, while smaller ones are first localized by a classification algorithm, and then segmented. This approach shows good results when used as part of the ABCs [8] challenge, a satellite event of the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) 2020 annual conference.

2 Methods

This approach segments structures from the ABCs dataset by combining predictions from a bi-directional framework. Of the T1-weighted MR, T2-weighted MR, and CT images, only the T1w and T2w MR images are used in this approach to segment large structures. Challenge segmentations are divided into two tasks: Task 1 segments brain structures (falx cerebri, tentorium cerebelli, sagittal and transverse brain sinuses, cerebellum, ventricles) critical for defining clinical target volumes (CTVs) for brain radiotherapy treatment, while Task 2 segments structures (brainstem, optic chiasm, cochleas, eyes, lacrimal glands, and optic nerves) used for optimizing the radiotherapy treatment plan. Large structures, defined as all structures from Task 1 along with the brainstem, and optic chiasm from Task 2, are directly segmented using MR images alone. For small structures (eyes, optic nerves, lacrimal glands, and cochleas), Inception-ResNet-v2 is used to localize each VOI using the CT, followed by segmentation of these structures from the co-registered T1 and T2 VOIs.

Fig. 1. Bi-directional framework. Task 1 uses T1 and T2 images to directly segment large structures. For Task 2, the CTs are input into an Inception Res-Net-V2 and outputs the range of slices for each small structure (using a single model per structure). These coordinates are used to select VOIs from the T1 and T2 images and are subsequently segmented as in Task 2. Black circles are convolutional blocks, black circles with down/up arrows are down/up sampling, Cyan circles are $1 \times 1$ convolutional blocks, green circles are convolutional blocks with ReLU pre-activations and sigmoid post activations, pink squares in the Attention U-Net network are attention gates.
2.1 Data

Data from ABCs [8] were used to train segmentation CNNs. This dataset included glioblastoma and low-grade glioma patients, with all provided patient scans consisted of CT, T1-weighted MR, and T2-weighted MR images of the post-operative brain. All scans were rescaled to have the same spatial dimensions. This dataset provided separate training and testing sets of 45 and 15 patients, respectively. The training dataset was further split 70%–30% training-validation. At the conclusion of the challenge, a previously unseen dataset of 15 additional patients was provided and used for the final scoring stage.

2.2 Training Task 1

After preliminary evaluations of various deep learning models, a modified Attention U-Net [9] with batch normalization and a $3 \times 3 \times 3$ convolutional kernel was selected for training all Task 1 segmentation tasks. DSC was used as the loss function as proposed by Milletari et al. [10]. Adam [11] was used as the optimizer, and the learning rate was set to 0.001 and reduced by 15% when loss did not decrease for 70 epochs. Training was performed on a 16 GB NVIDIA-V100 with batch size 2. Each model was trained for approximately 1,000 iterations using early stopping. Each of the five structures in Task 1 was trained using a dedicated model, and individual structure predictions were combined to create the final predictions.

During training, a body mask was found by thresholding all CT values of –600 or higher. T1w and T2w regions within this mask were normalized by subtracting the mean and then dividing by the standard deviation (“z-score normalization”). Values outside the mask were reset to the minimum value for each scan. Only T1w and T2w images were then used by the model for training.

All data was augmented on-the-fly during training by random combinations of rotations in the x-y plane ($\pm 15^\circ$); translations ($\pm 10, \pm 10, \pm 5$) along the x, y, and z axes, respectively; zooms in the x-y plane ($\pm 15\%$), and warps. Additionally, before training, all data were flipped left-right to increase the training set size.

2.3 Training Task 2

The structures in Task 2, apart from the brainstem, are significantly smaller than the structures in Task 1, and the approach used for Task 1 was not optimal for the small structures. The relative volume of these structures compared to the input image can be increased by reducing the field of view of a CT scan. To do so, a CNN-based classification was applied to the CT scan to detect the location of the structures, and the images were cropped around the center of mass. The accuracy of the segmentation models improved by training the model with cropped images (i.e. narrow field of view CT scans).

Classification. Inception-ResNet-v2 [12], a CNN-based 2D classification architecture, was trained to classify the existence of an organ-of-interest in each CT slice. The classification model was trained to detect the eyes and the cochleae in each CT slices. The center of mass of each organ in cranial-caudal direction was determined by taking the average of the results from the classification model. Then, 20 slices above and below
from the central slice were clipped from an original CT scan to cover the entire organ with some margin. In the segmentation phase, the total of the 40 slices were transferred to the segmentation model. The cropped slices for the eyes were also used to segment the optic nerves, the optic chiasm, and the lacrimal glands, as these structures were all located within the slices cropped for the eyes.

**Segmentation.** After preliminary evaluations of various deep learning models, two models were selected for Task 2 segmentation based upon individual structure segmentation performance. Each of the six structures in Task 2 was trained using a dedicated model, and individual structure predictions were combined to create the final predictions. Training was performed on a 32 GB NVIDIA-V100 for all models.

The brainstem, eyes, lacrimal structures, and optic nerves were segmented with a modified Attention U-Net [9] with batch normalization and a $3 \times 3 \times 3$ convolutional kernel. DSC was used as the loss function as proposed by Milletari et al. [10]. Adam [11] was used as the optimizer, and the learning rate was set to 0.001 and reduced by 15% when loss did not decrease for 70 epochs. Batch size was set to 5. Each model was trained for approximately 1,000 iterations using early stopping.

For the remaining Task 2 structures, the optic chiasm and the cochleas, a modified V-Net [10] with batch normalization and a $3 \times 3 \times 3$ convolutional kernel was selected for segmentation. Batch size was set to 2, and all other training hyperparameters followed those described above. The choice of this architecture for these structures, instead of the Attention U-Net architecture used for all other Task 2 segmentations, was based upon visual inspection of preliminary results which revealed better performance for these structures.

During training, a body mask was found by thresholding all CT values of –600 or higher. T1w and T2w regions within this mask were normalized by subtracting the mean and then dividing by the standard deviation ("z-score normalization"). Values outside the mask were reset to the minimum value for each scan. Only T1w and T2w images were then used by the model for training.

All data was augmented on-the-fly during training by random combinations of rotations in the x-y plane (±15°); translations (±10, ±100, ±5) along the x, y, and z axes, respectively; zooms in the x-y plane (±15%), and warps. Additionally, before training all data was flipped left-right to increase the training set size.

### 2.4 Evaluation Metrics

All submissions to the challenge were evaluated with two metrics: the Dice similarity coefficient (DSC) as well as the surface Dice similarity coefficient (SDSC) with a tolerance of 2 mm [13]. DSC and SDSC scores for individual structures were computed separately, and then averaged for the overall Task 1 and Task 2 scores. Finally, the overall challenge ranking was based upon the unweighted average of the Task 1 and Task 2 scores [8].
3 Results

3.1 Challenge Results

In the final challenge submission, relatively good results were achieved for large structure segmentations, with mean DSC above 0.9 for three of the seven large structures, and above 0.7 for three of the remaining. Similarly, mean SDSC with a 2mm tolerance was above 0.9 for six of these. For both metrics, sagittal and transverse brain sinus segmentation scored the lowest, at 0.64 and 0.89 for mean DSC and mean SDSC, respectively.

Conversely, small structure scores were low in the final submission set. This was due to an error in the inference and data conversion scripts, which assigned the left structure class to right structures, and vice-versa. Post-challenge, these errors were corrected, with significant improvement to model performance.

4 Post-challenge Processing

After results were reported at the conclusion of the challenge, poor performance was observed in the segmentation scores for the cochleas, eyes, lacrimal glands, and optic nerves. This was identified as an error in the inference pipeline, which led to organs being incorrectly predicted on the wrong side of the body (i.e. the left eye segment was predicted on the right side of the body, and vice-versa). Performance improved
Table 1. DSC scores for structures in the final challenge submission. The first five structures belong to Task 1, while brainstem and chiasm belong to Task 2. Small structures from Task 2 are excluded as scoring failed due to technical error. See Table 3 for corrected small structure scores.

|        | Cerebellum | Falx  | Sinuses | Tentorium | Ventricles | Brainstem | Chiasm |
|--------|------------|-------|---------|-----------|------------|-----------|--------|
| Min    | 0.996      | 0.919 | 0.244   | 0.623     | 0.602      | 0.999     | 0.85   |
| Mean   | 0.997      | 0.941 | 0.643   | 0.726     | 0.704      | 0.999     | 0.886  |
| Max    | 0.998      | 0.957 | 0.751   | 0.805     | 0.819      | 0.999     | 0.921  |

Table 2. SDSC scores for structures in the final challenge submission. The first five structures belong to Task 1, while brainstem and chiasm belong to Task 2. Small structures from Task 2 are excluded as scoring failed due to technical error. See Table 4 for corrected small structure scores.

|        | Cerebellum | Falx  | Sinuses | Tentorium | Ventricles | Brainstem | Chiasm |
|--------|------------|-------|---------|-----------|------------|-----------|--------|
| Min    | 0.969      | 0.879 | 0.469   | 0.834     | 0.865      | 0.988     | 0.808  |
| Mean   | 0.982      | 0.934 | 0.894   | 0.92      | 0.942      | 0.993     | 0.9    |
| Max    | 0.99       | 0.967 | 0.979   | 0.987     | 0.984      | 0.996     | 0.966  |

**Fig. 3.** Individual structure scores after correcting post-processing. DSC is volumetric dice similarity coefficient, and SDSC + 2 mm is surface dice similarity coefficient with 2 mm tolerance. Upon reassigning the structures to the correct orientation. Corrected scores for small structures are listed in Tables 3 and 4.
Table 3. DSC scores for structures corrected post-challenge. These structures, all belonging to Task 2, were unable to be scored during the challenge due to technical error.

| Structure       | Cochlea L | Cochlea R | Eye L | Eye R | Lacrimal L | Lacrimal R | Optic Nerve L | Optic Nerve R |
|-----------------|-----------|-----------|-------|-------|------------|------------|---------------|---------------|
| Min             | 0.127     | 0.417     | 0.251 | 0.587 | 0.736      | 0.107      | 0.154         | 0.349         |
| Mean            | 0.408     | 0.529     | 0.488 | 0.84  | 0.839      | 0.504      | 0.491         | 0.547         |
| Max             | 0.564     | 0.634     | 0.696 | 0.922 | 0.914      | 0.703      | 0.671         | 0.686         |

Table 4. SDSC scores for structures corrected post-challenge. These structures, all belonging to Task 2, were unable to be scored during the challenge due to technical error.

| Structure       | Cochlea L | Cochlea R | Eye L | Eye R | Lacrimal L | Lacrimal R | Optic Nerve L | Optic Nerve R |
|-----------------|-----------|-----------|-------|-------|------------|------------|---------------|---------------|
| Min             | 0.466     | 0.826     | 0.773 | 0.474 | 0.606      | 0.384      | 0.497         | 0.684         |
| Mean            | 0.77      | 0.932     | 0.911 | 0.844 | 0.84       | 0.773      | 0.778         | 0.872         |
| Max             | 0.893     | 0.999     | 0.995 | 0.993 | 0.987      | 0.952      | 0.923         | 0.972         |

5 Discussion

Differences in modality make direct comparison between the approach of this study and others’ difficult. In particular, published results often focus on MR [14, 15] or CT [13, 16, 17] exclusively instead of the multi-modality approach described here. However, our results generally compare favorably to those published in the literature. For example, this work finds mean brainstem DSC and SDSC as 0.999 and 0.993, respectively, which consistently ranks higher than other reported values [13–17]. Similar improvements for DSC are observed for cerebellum, optic chiasm, and left lacrimals [13–15]. However, SDSC scores for cochleas, lacrimals, and optic nerves were often lower than reported values [13].

There are a few limitations associated with this work. The DSC and SDSC reported for small structures are often, though not always, lower than those generated by CT-only approaches. This may be due to the poorer visibility of these structures in MR imagery, which were the only inputs into the segmentation models of this work. Future research will incorporate CT images into the segmentation models for additional context. In addition, the relatively small test set (n = 15) may be considered a limitation; however, the similarity of our results to those with significantly larger test sets indicates that the size of this test set did not significantly impact the work.

6 Conclusion

In this work, we develop a multi-modality framework for rapid autosegmentation of brain structures as part of an international segmentation challenge. The results of this
study showed good agreement with or improvement upon similar values reported in the literature. Future work will include incorporating CT images directly into segmentation models for challenging small structures. This approach could contribute to the Radiation Planning Assistant [18], a fully automated treatment planning tool aimed at improving access to high quality radiation therapy across the globe.

Acknowledgements. The authors acknowledge the support of the High Performance Computing facility at the University of Texas MD Anderson Cancer Center and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing computational resources that have contributed to the research results reported in this paper.

References

1. Jeanneret-Sozzi, W., et al.: The reasons for discrepancies in target volume delineation: a SASRO study on head-and-neck and prostate cancers. Strahlenther. Onkol. 182(8), 450–457 (2006). https://doi.org/10.1007/s00066-006-1463-6
2. Brouwer, C.L., et al.: 3D variation in delineation of head and neck organs at risk. Radiat. Oncol. 7, 32 (2012). https://doi.org/10.1186/1748-717X-7-32
3. González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., Lladó, X.: A review on brain structures segmentation in magnetic resonance imaging (2016). https://doi.org/10.1016/j.artmed.2016.09.001
4. Cardenas, C.E., Yang, J., Anderson, B.M., Court, L.E., Brock, K.B.: Advances in Auto-Segmentation (2019). https://doi.org/10.1016/j.semradonc.2019.02.001
5. 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. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
6. De Brébisson, A., Montana, G.: Deep neural networks for anatomical brain segmentation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 20–28 (2015). https://doi.org/10.1109/CVPRW.2015.7301312.
7. McClure, P., et al.: Knowing What You Know in Brain Segmentation Using Bayesian Deep Neural Networks (2019). https://doi.org/10.3389/fninf.2019.00067. https://www.frontiersin.org/articles/10.3389/fninf.2019.00067/full
8. Shusharina, N., Bortfield, T., Cardenas, C.E.: MICCAI 2020 ABCs Challenge. https://abcs.mgh.harvard.edu/index.php.
9. Oktay, O., et al.: Attention U-Net: Learning where to look for the pancreas. arXiv (2018)
10. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016, pp. 565–571 (2016). https://doi.org/10.1109/3DV.2016.79
11. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization (2014)
12. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017, pp., 4278–4284 (2017)
13. Nikolov, S., et al.: Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. arXiv (2018)
14. Moeskops, P., et al.: Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI. NeuroImage Clin. 17, 251–262 (2018). https://doi.org/10.1016/j.nicl.2017.10.007
15. Luna, M., Park, S.H.: 3D Patchwise U-Net with transition layers for MR brain segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 394–403. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_40

16. Wang, Y., Zhao, L., Wang, M., Song, Z.: Organ at risk segmentation in head and neck CT images using a two-stage segmentation framework based on 3D U-net. IEEE Access. 7, 144591–144602 (2019). https://doi.org/10.1109/ACCESS.2019.2944958

17. Rhee, D.J., et al.: Automatic detection of contouring errors using convolutional neural networks. Med. Phys. 46, 5086–5097 (2019). https://doi.org/10.1002/mp.13814

18. Court, L.E., et al.: Radiation planning assistant - a streamlined, fully automated radiotherapy treatment planning system. J. Vis. Exp. 2018, e57411 (2018). https://doi.org/10.3791/57411