A Cascaded Deep-Learning Framework for Segmentation of Metastatic Brain Tumors Before and After Stereotactic Radiation Therapy*

Ali Jalalifar, Hany Soliman, Arjun Sahgal, and Ali Sadeghi-Naini, *Senior Member, IEEE*

Abstract— **Radiation therapy is a major treatment option for brain metastasis. For radiation treatment planning and outcome evaluation, magnetic resonance (MR) images are acquired before and at multiple sessions after the treatment. Accurate segmentation of brain tumors on MR images is crucial for treatment planning, response evaluation, and developing data-driven models for outcome prediction. Due to the high volume of imaging data acquired from each patient at multiple follow-up sessions, manual tumor segmentation is resource- and time-consuming in clinic, hence developing an automatic segmentation framework is highly desirable. In this work, we proposed a cascaded 2D-3D Unet framework to segment brain tumors automatically on contrast-enhanced T1-weighted images acquired before and at multiple scan sessions after radiotherapy. 2D Unet is a well-known structure for medical image segmentation. 3D Unet is an extension of 2D Unet with a volumetric input image to provide richer spatial information. The limitation of 3D Unet is that it is memory consuming and cannot process large volumetric images. To address this limitation, a large volumetric input of 3D Unet is often patched to smaller volumes which leads to loss of context. To overcome this problem, we proposed using two cascaded 2D Unets to crop the input volume around the tumor area and reduce the input size of the 3D Unet, obviating the need to patch the input images. The framework was trained using images acquired from 96 patients before radiation therapy and tested using images acquired from 10 patients before and at four follow-up scans after radiotherapy. The segmentation results for the images of independent test set demonstrated that the cascaded framework outperformed the 2D and 3D Unets alone, with an average Dice score of 0.9 versus 0.86 and 0.88 for the baseline, and 0.87 versus 0.83 and 0.84 for the first follow-up. Similar results were obtained for the other follow-up scans.**

*This research was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada, Terry Fox Foundation, and the Lotte and John Hecht Memorial Foundation.

A. Jalalifar is with the Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, ON, Canada (e-mail: alijfar@yorku.ca).

H. Soliman, and A. Sahgal are with the Department of Radiation Oncology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; also with the Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada (e-mail:hany.soliman@sunnybrook.ca; arjun.sahgal@sunnybrook.ca).

A. Sadeghi-Naini is with the Department of Electrical Engineering and Computer Science, Lassonde School of Engineering York University, Toronto, ON, Canada; also with the Department of Radiation Oncology and Physical Sciences Platform, Odette Cancer Centre and Sunnybrook Research Institute, Sunnybrook Health Sciences Centre; also with the Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada (e-mail: asn@yorku.ca, phone: 416-736-2100 x20590).

I. Introduction

Malignant tumors in the brain can be classified into the primary and secondary (metastatic) tumors. Primary brain tumors initiate in the brain, while metastatic brain tumors originate from a malignancy in another part of the body, e.g. lung, breast, kidney, colon. Up to 40% of cancer patients develop brain metastases [1]. Patients with brain metastasis suffer from poor overall survival even after treatment [2], [3]. Brain metastasis can present as single or multiple tumors. Clinical records show that 50% of brain metastases occur with a single tumor, 20% with two tumors, and 30% with three or more tumors [4]. Early diagnosis and precise treatment of brain metastasis may lead to less brain symptoms and improve patient's quality of life.

Based on a number of diagnostic factors including the origin of primary cancer, number and location of metastases, and symptoms, a treatment regimen is planned for brain metastasis. Current treatment options for brain metastasis include one or a combination of surgery, systemic therapy, whole-brain radiation therapy (WBRT), stereotactic radiosurgery (SRS), and hypo-fractioned stereotactic radiation therapy (SRT). With the advancements of imaging technologies, surgery is a safe but the most invasive method of choice for brain metastasis treatment. The main decisive factors to proceed with surgery include accessibility and size of the tumor, patient's age, and the degree of mass effect [5]. In Whole-brain radiation therapy, radiation is given to the whole brain over a period of several weeks. SRS delivers very high-dose radiation to a precisely-targeted area in a single session (fraction). Sometimes the location of the tumor is too close to a critical structure or the size of the tumor is large. In these cases, high-dose radiation is frequently delivered to tumor in few fractions (hypo-fractionated SRT).

Magnetic resonance imaging (MRI) is performed before and at several follow-up sessions after radiotherapy as part of standard of care for treatment planning and therapy outcome evaluation. Local response of brain metastasis to radiation therapy is conventionally assessed based on changes in physical dimensions of tumor on follow-up imaging [6]. Therefore, for each follow-up assessment, the tumors are required to be segmented on magnetic resonance (MR) images. Manually segmenting tumors on several 2D planes of a 3D MRI is a laborious task, especially in the presence of multiple follow-up scans for a single patient. An automatic framework for segmentation of brain tumors is highly desirable and can streamline radiation therapy workflow

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

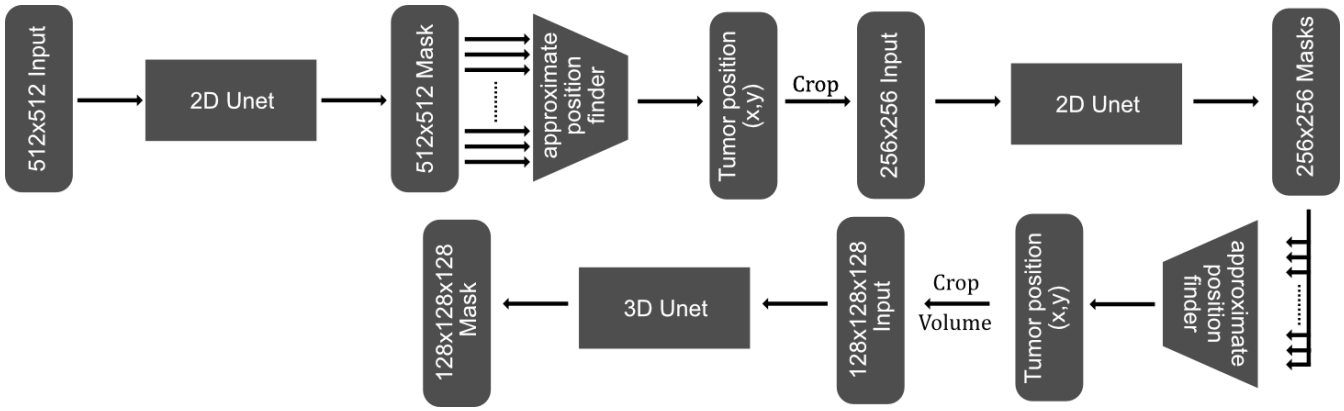

Figure 1 - System Overview. For a volumetric input $512 \times 512 \times 128$ voxel, all slices are initially fed to a 2D Unet one by one. The generated masks from the 2D Unet are used to find an approximate tumor position (x, y). The volumetric input image is then cropped around (x, y) into a $256 \times 256 \times 128$ voxel volume. A similar procedure is performed to reduce the size of volumetric image containing the tumor to 128×128×128 voxel. This volume is then fed into a 3D Unet for final segmentation with no patching.

considerably. Development of such methods have been the focus of intense research efforts, and various algorithms have been proposed including the morphological methods [7], region-based methods [8], deformable models [9], and models based on deep learning [10].

Not only an automatic brain tumor segmentation framework can streamline the radiation therapy workflow, but also it can help to develop radiomics-based predictive models of therapy response and outcome. Radiomics aims to extract minable high-dimensional features from clinical images for diagnostic and prognostic applications [11]. Radiomic features quantify the shape, or spatial arrangement of the voxel intensity values (texture) within an abnormality and can be applied on the original or filtered images [12]. Radiomic features have shown a good potential to predict the outcome of therapy in different cancer sites [13]–[17]. In such applications, the radiomic features are extracted from the tumor region and its surrounding areas. Therefore, to adapt radiomic models for outcome prediction in cancer therapeutics, tumor segmentation is an initial necessity. An automatic segmentation framework can facilitate the adaptation and development of radiomic models for brain tumor management.

In this study, we proposed an automatic framework for accurate segmentation of brain tumors on contrast-enhanced T1-weighted images acquired before (baseline) and at several follow-up sessions after radiation therapy. We applied a cascade of 2D and 3D Unets to guide the focus of the framework to the tumor regions and subsequently segment it with high accuracy. 2D Unet [18] is a well-known and powerful network, widely used for medical image segmentation. 3D Unet [19] is an extension of 2D Unet which aims to incorporate more spatial information for segmentation by using 3D voxels instead of pixels. One limitation of 3D Unet is that it is memory demanding. In order to meet the memory requirements, one has to patch the input volume into smaller volumes which leads to loss of context and spatial information for segmentation. To address this problem, we trained two 2D Unets one after another which find the approximate position of the tumor and then crop the image around the tumor. Specifically, the framework first crops the input volume ($512 \times 512 \times 128$ voxel) to a $256 \times 256 \times 128$ voxel volume, and then to a volume with size of $128 \times 128 \times 128$ voxel that contain the entire tumor. Subsequently, this volume is fed to the 3D Unet for final segmentation with no volume patching. The framework was trained and validated using images acquired from 96 patients diagnosed with brain metastasis, and tested independently using images acquired from 10 patients. The results show that the cascaded network outperformed the 2D and 3D Unets alone, with a Dice score of 0.9 for the baseline and 0.87, 0.84, 0.82 and 0.82 for the first to forth follow-up scans, respectively.

II. Methods

A. Dataset

This study was conducted in accordance with institutional research ethics approval from Sunnybrook Health Sciences Centre (SHSC), Toronto, Canada. The imaging data were collected from 106 patients diagnosed with brain metastasis and planned for hypo-fractionated stereotactic radiation therapy (SRT). The dataset included the gadolinium contrast-enhanced T1-weighted images acquired before SRT for treatment planning, and at up to 4 follow-up sessions after treatment. The dataset also included the tumor contours for each patient, delineated by expert oncologists and neuroradiologists. These tumor contours were used as ground truth to train and evaluate the framework. The baseline images of 90 and 6 patients were used for training and validation of the framework, respectively. The baseline and follow-up images of 10 patients (6 patients for the fourth follow-up) were used as an independent test set for framework evaluation

B. Segmentation Framework

Figure 1 shows a scheme of the proposed automatic segmentation framework. The framework consists of three separate networks, two 2D Unets and one 3D Unet, each of them was trained separately. The purpose of 2D Unets is to find an approximate position of the tumor. Once the approximate position is found, the image is cropped around the tumor position to make the size of input for the next

network smaller. Specifically, the size of the input images for the first 2D Unet was 512×512 pixel. The output of this network was a 512×512 pixel tumor mask. The second 2D Unet was trained on cropped 256×256 pixel slices. For training this network, cropping was done manually using ground truth data. It should be noted that manual cropping was only done on the training data to generate training samples. All the test images were cropped automatically as follows. For the images of the independent test set, the 512×512 pixel tumor masks generated by the first 2D Unet were used to find an approximate location of the tumor and crop a 256×256 pixel window around the tumor area. In the next step, this cropped window was fed to the second 2D Unet. The output of the second 2D Unet was a 256×256 pixel tumor mask. The final approximate position of the tumor was found using these 256×256 pixel tumor masks and used to crop the input volume ($512 \times 512 \times 128$ voxel) to a $128 \times 128 \times 128$ voxel volume containing the whole tumor. The cropped volume was subsequently fed to the 3D Unet for final segmentation. The 3D Unet was trained on manually cropped $128 \times 128 \times 128$ voxel volumes from the training set.

1) 2D Unet

Unet is a well-known network structure for medical image segmentation consists of 19 fully-connected layers. It exploits the power of skip connections to tackle the trade-off between context and localization. The U-like structure of the network consists of two paths: the analysis path and the synthesis path. The analysis path consists of a series of convolution and max-pooling. The synthesis path consists of up-sampling and deconvolution layer. Skip connections connect the layers of the same resolution from the analysis path and synthesis path. The connection from the analysis path to the synthesis path provides high-resolution contextual features to deconvolution layers [20].

As described previously, two 2D Unets were applied in the proposed cascaded structure to guide the focus of the framework to the tumor area. The second 2D Unet potentially outputs a better approximation of the tumor region compared to the first 2D Unet because its smaller-size input image includes less redundant information and a higher propositional area of tumor.

The output of each 2D Unet for a patient is a set of masks. To find the approximate position of the tumor from these masks, a logical OR operation was applied on all the masks to create a single mask presenting an upper-bound of the tumor areas in different slices. Subsequently, the connected components were identified in this single mask and the center of each connected component was regarded as the approximate center of the corresponding tumor. The approximated centers were used to crop the image around the tumor areas.

2) 3D Unet

3D Unet is an extension of 2D Unet that replaced its 2D components with 3D components to provide richer spatial information through volumetric segmentation. In an architecture similar to 2D Unet, 3D Unet also contains contracting and expanding path. In the contracting path, each layer contains two $3 \times 3 \times 3$ convolutional layers followed by a rectified activation unit (ReLu) and max-pooling. In the expanding path, each layer consists of $2 \times 2 \times 2$ upconvolution followed by two $3 \times 3 \times 3$ convolutions. In this study, 3D Unet was trained on $128 \times 128 \times 128$ voxel volumetric T1-weighted MRI images.

III. RESULTS

The tumor segmentation results obtained for the T1-weighted images acquired form five representative patients at the baseline and four follow-up sessions are shown in Figure 2. This figure presents the final segmentation masks generated by the framework overlaid on the ground truth masks. Table I presents the average Dice similarity coefficient between the generated and ground truth masks over all tumor slices for these five patients. The results presented in Figure 2 and Table I imply that whereas the network generated accurate segmentation masks for the baseline and first follow-up, the accuracy of the segmentation gradually decreased on the subsequent follow-up images. This trend may be due to the fact that in some of these cases, the tumor shrunk after radiation therapy, made it harder to segment the remaining tumor. Another possible reason is that the texture of tumor may change at many months after multiple session of radiation therapy, resulting in test samples with no similar case in the training set that makes the segmentation challenging for the network.

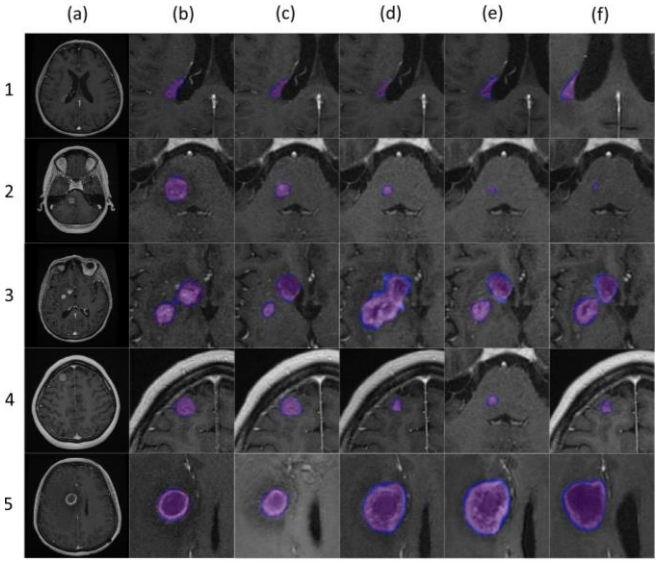

Figure 2 – An axial cross-section image of five representative tumors of the independent test set (first to fifth rows) at the baseline (a) and (b), first (c), second (d), third (e), and forth (f) follow-up scans. The blue, red, and purple overlays represent the ground truth mask, automatic segmentation mask, and the overlap region, respectively.

Table II compares the segmentation result of the cascaded 2D-3D Unet with 2D and 3D Unets alone in terms of average Dice similarity coefficient and Hausdorff distance for all patients of the independent test set. The results of this table show that the cascaded structure improved the segmentation results considerably compared to the 2D or 3D Unets alone.

IV. Discussion

In this study, a cascaded deep-learning framework was proposed for automatic segmentation of brain tumors on T1-weighted MR images before and after radiation therapy. Two back to back 2D Unets cropped the volumetric input image of the 3D Unet in the framework in order to tackle the trade-off between context and localization. Brain tumor segmentation is crucial for follow-up assessment and response evaluation after radiation therapy as well as developing radiomics-based models for therapy outcome prediction. Results from this study showed that the cascaded 2D-3D Unet outperformed the 2D 3D Unets alone in brain tumor segmentation. Further investigation is required to evaluate whether the proposed framework can provide enough accuracy for SRT response evaluation based on the measured changes in the physical dimensions of tumor on follow-up imaging. Such assessments have been planned as a future direction of this research.

TABLE I. THE DICE SIMILARITY COEFFICIENT FOR THE BASELINE AND FOUR FOLLOW-UPS (FU) FOR THE FIVE PATINETS OF FIGURE 2.

Network	Baseline	FU1	FU2	FU3	FU4
2D + 3D Unets	0.90 ± 0.08	0.87 ± 0.07	0.84 ± 0.08	0.82 ± 0.9	0.82 ± 0.08
	1.12 ± 0.3 mm	1.36 ± 0.4 mm	1.38 ± 0.3 mm	1.37 ± 0.3 mm	1.35 ± 0.2 mm
2D Unet	0.86 ± 0.09	0.83 ± 0.1	0.81 ± 0.09	0.79 ± 0.08	0.78 ± 0.09
	1.43 ± 0.5 mm	1.89 ± 0.6 mm	1.88 ± 0.4 mm	1.91 ± 0.4 mm	1.97 ± 0.3 mm
3D Unet	0.88 ± 0.08	0.84 ± 0.08	0.82 ± 0.08	0.8 ± 0.09	0.8 ± 0.09
	1.25 ± 0.4 mm	1.69 ± 0.4 mm	1.75 ± 0.4 mm	1.73 ± 0.3 mm	1.73 ± 0.3 mm

TABLE II. THE DICE SIMILARITY COEFFICIENT (UP IN EACH CELL) AND HAUSDORFF DISTANCE (DOWN IN EACH CELL) FOR THE BASELINE AND FOUR FOLLOW-UPS (FU) FOR ALL PATIENTS OF THE INDEPENDENT TEST SET.

Patient Number	Baseline	FU1	FU2	FU3	FU4
1	0.91 ± 0.06	0.9 ± 0.05	0.88 ± 0.04	0.85 ± 0.05	0.84 ± 0.06
2	0.88 ± 0.07	0.87 ± 0.04	0.83 ± 0.03	0.79 ± 0.08	0.79 ± 0.05
3	0.87 ± 0.05	0.86 ± 0.05	0.82 ± 0.03	0.81 ± 0.04	0.8 ± 0.04
4	0.89 ± 0.03	0.86 ± 0.05	0.84 ± 0.06	0.80 ± 0.04	0.81 ± 0.03
5	0.90 ± 0.04	0.88 ± 0.06	0.85 ± 0.05	0.84 ± 0.04	0.84 ± 0.06
Mean ± SD	0.89 ± 0.08	0.87 ± 0.08	0.84 ± 0.07	0.81 ± 0.09	0.81 ± 0.08

References

[1] I. T. Gavrilovic and J. B. Posner, "Brain metastases: epidemiology and pathophysiology," *J. Neurooncol.*, vol. 75, no. 1, pp. 5–14, Oct. 2005.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA. Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, Jan. 2019.

[3] L. Gaspar *et al.*, "Recursive partitioning analysis (RPA) of prognostic factors in three radiation therapy oncology group (RTOG) brain metastases trials," *Int. J. Radiat. Oncol.*, vol. 37, no. 4, pp. 745–751, Mar. 1997.

[4] J. Y. Delattre, G. Krol, H. T. Thaler, and J. B. Posner, "Distribution of Brain Metastases," *Arch. Neurol.*, vol. 45, no. 7, pp. 741–744, Jul. 1988.

[5] T. K. Owonikoko *et al.*, "Current approaches to the treatment of metastatic brain tumours," vol. 11, no. 4, pp. 203–222, 2014.

[6] N. U. Lin *et al.*, "Response assessment criteria for brain metastases: proposal from the RANO group," *Lancet Oncol.*, vol. 16, no. 6, pp. e270–e278, Jun. 2015.

[7] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, "Tumour volume determination from MR images by morphological segmentation," *Phys. Med. Biol.*, vol. 41, no. 11, pp. 2437–2446, Nov. 1996.

[8] A. Stadlbauer *et al.*, "Improved delineation of brain tumors: an automated method for segmentation based on pathologic changes of 1H-MRSI metabolites in gliomas," *Neuroimage*, vol. 23, no. 2, pp. 454–461, Oct. 2004.

[9] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, and F. R. Murtagh, "Automatic segmentation of non-enhancing brain tumors in magnetic resonance images," *Artif. Intell. Med.*, vol. 21, no. 1–3, pp. 43–63, Jan. 2001.

[10] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[11] S. Rizzo *et al.*, "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.*, vol. 2, no. 1, 2018.

[12] S. S. F. Yip and H. J. W. L. Aerts, "Applications and limitations of radiomics," *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–R166, 2016.

[13] M. J. Gangeh, A. Sadeghi-Naini, M. Diu, H. Tadayyon, M. S. Kamel, and G. J. Czarnota, "Categorizing Extent of Tumor Cell Death Response to Cancer Therapy Using Quantitative Ultrasound Spectroscopy and Maximum Mean Discrepancy," *IEEE Trans. Med. Imaging*, vol. 33, no. 6, pp. 1390–1400, Jun. 2014.

[14] M. Vallières *et al.*, "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer," *Sci. Rep.*, vol. 7, no. 1, p. 10117, Dec. 2017.

[15] E. Karami, M. Ruschin, H. Soliman, A. Sahgal, G. J. Stanisz, and A. Sadeghi-Naini, "An MR Radiomics Framework for Predicting the Outcome of Stereotactic Radiation Therapy in Brain Metastasis*," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 1022–1025.

[16] E. Karami *et al.*, "Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis.," *Sci. Rep.*, vol. 9, no. 1, p. 19830, Dec. 2019.

[17] H. J. W. L. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, no. 1, p. 4006, Sep. 2014.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8, 2015.

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," Jun. 2016.

[20] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," *J. Digit. Imaging*, vol. 3, pp. 582–596, 2019.