A Memory-efficient Deep Framework for Multi-Modal MRI-based Brain Tumor Segmentation

Hashemi, Nima; Masoudnia, Saeed; Nejad, Ashkan; Nazem-Zadeh, Mohammad-Reza

Published in:
2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)

DOI:
10.1109/EMBC48229.2022.9871726

2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871726

A Memory-efficient Deep Framework for Multi-Modal MRI-based Brain Tumor Segmentation

Nima Hashemi[1], Saeed Masoudnia[2], Ashkan Nejad[3] and Mohammad-Reza Nazem-Zadeh[2]

Abstract— **Automatic Brain Tumor Segmentation (BraTS) from MRI plays a key role in diagnosing and treating brain tumors. Although 3D U-Nets achieve state-of-the-art results in BraTS, their clinical use is limited due to requiring high-end GPU with high memory. To address the limitation, we utilize several techniques for customizing a memory-efficient yet accurate deep framework based on 2D U-nets. In the framework, the simultaneous multi-label tumor segmentation is decomposed into fusion of sequential single-label (binary) segmentation tasks. In addition to reducing the memory consumption, it may also improve the segmentation accuracy since each U-net focuses on a sub-task, simpler than whole BraTS segmentation task. Extensive data augmentations on multi-modal MRI and the batch dice-loss function are also employed to further increase the generalization accuracy. Experiments on BraTS 2020 demonstrate that our framework almost achieves state-of-the-art results. Dice scores of 0.905, 0.903, and 0.822 for whole tumor, tumor core, and enhancing tumor are accomplished on the testing set. Moreover, our customized framework is executable on budget-GPUs with minimum requirement of only 2G RAM.** *Clinical relevance*— **We develop a memory-efficient deep Brain tumor segmentation tool that significantly reduces the hardware requirement of tumor segmentation while maintaining comparable accuracy and time. These advantages make our framework suitable for widespread use in clinical applications, especially in low-income regions. We plan to release the framework as a part of a free clinical brain imaging analysis tool. The code for this framework is publicly available:`https://github.com/Nima-Hs/BraTS`.**

I. INTRODUCTION

Automatic segmentation of brain tumors from Magnetic Resonance Imaging (MRI) images constitutes an essential step in today's clinical diagnosis and treatment planning and assessment. However, it is a challenging task due to not only tumor's inherent heterogeneous tissue but also its small ratio within large volume of brain [1].

Numerous deep learning studies have tried to address the challenges, while U-net architecture is the most successful model among them for Brain Tumor Segmentation (BraTS) [2]. This architecture is designed based on a deep encoder-decoder convolution network with skip connections. Many extensions of U-Net have been proposed based on different losses and architectures. Deep-learning segmentation frameworks rely not only on the choice of network architecture but also on the choice of loss function. Several works introduced different losses in U-net model in order to address the mentioned challenges for BraTS, [3], [4], [5]. Although Cross Entropy (CE) was suggested in the original U-net [2], other losses, e.g., dice loss also used [6], specially, in the unbalanced segmentation tasks [7]. Moreover, different modifications were suggested for dice loss to better handle unbalanced segmentation challenges. Carole H Sudre et al. [3] suggested weighting each class by the inverse of its volume, thus balancing between them. Oldřich Kodym et al. [5] proposed batch dice loss in which loss computation is extended over whole data mini-batches and considered them as a single 4-dimensional tensor. The authors declare that this regularization avoids disturbance in mini-batch optimization if few individual gradients are very different. Moreover, U-Nets based on different architectures were proposed. The first introduced U-nets are based on 2D CNNs, taking a single slice as input, can not take advantage of context from adjacent slices. 3D U-nets address this limitation by using 3D convolution filters on a 3D input volume. This improvement enables 3D U-nets to exploit inter-slice context, which leads to better segmentation. However, it comes at a cost of high computational and memory requirements due to increased number of parameters [8].

Although 3D U-nets achieve state-of-the-art results in BraTS task [9], [10], [4], they only could be runnable on high-end GPUs with high RAM. This heavy computational requirement severely limits the widespread clinical applications of these networks, especially in low-income communities. Due to this consideration, we attempt to provide a feasible solution based on 2D U-net for BraTS. In this paper, we address the mentioned BraTS challenges by combining the idea proposed in [3], [5]. Some technical modifications are also proposed to enhance the performance in order to be comparable with state-of-the-arts, while yet runnable on budget GPUs.

II. DATASET

Multimodal 3T MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG) are provided in BraTS2020 dataset [11]. The multimodal scans are native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes (Fig. 1). These volumes were acquired through different clinical protocols and various scanners from multiple institutions. All Images have been manually segmented to GD-enhancing

[1] Nima Hashemi is a graduate student at school of Electrical and Computer Engineering, University of Tehran.`nimahashemi57@gmail.com`

[2] Saeed Masoudnia and Mohammad-Reza Nazem-Zadeh are with Medical Physics and Biomedical Engineering Department and the research center for molecular and cellular imaging (RCMCI), Tehran University of Medical Sciences (TUMS), Tehran, Iran `s.masoudnia@gmail.com` and `mnazemzadeh@tums.ac.ir`

[3] Ashkan Nejad is with the Machine Learning Lab, Data Science Center in Health (DASH), University Medical Center Groningen, University of Groningen, The Netherlands `a.nejad@rug.nl`

978-1-7281-2782-8/22/$31.00 ©2022 IEEE

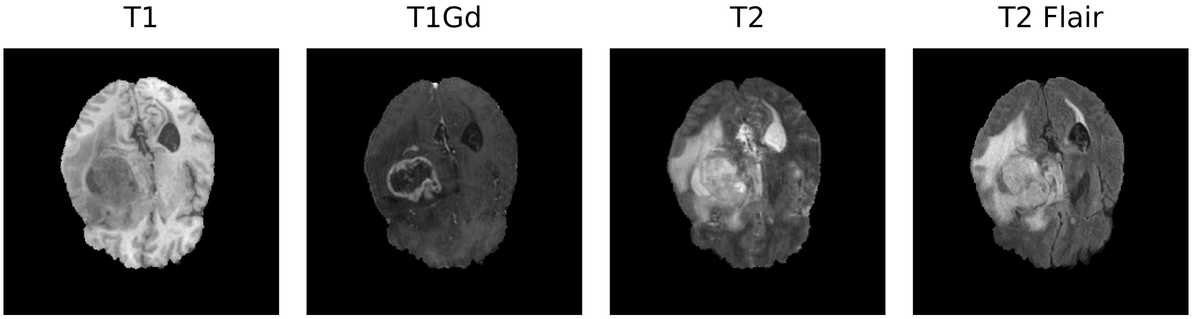

Fig. 1. Different MRI modalities in the BraTS dataset. There are 155 axial slices in each volume, and each slice is 256×256. Each subject has four MRI modalities: T1, T1Gd, T2, and T2 Flair.

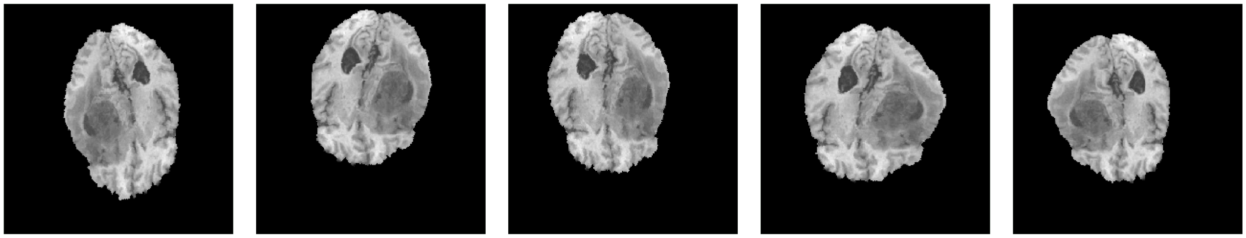

Fig. 2. Five samples of ten augmented images generated from a single slice. The augmentations and their parameters are in Table I.

tumor (ET – label 4), the peritumoral edema (ED – label 2), and the necrotic and non-enhancing tumor core (NCR/NET – label 1). The volumes are co-registered, interpolated to the same resolution($1mm^3$) and skull-stripped.

III. METHOD

A. Pre-processing

The 369 subjects are divided into three parts for training, validation, and testing. The training set has 295 subjects (80%), 37 subjects (10%) for the validation set, and 37 subjects (10%) for the test set. Each volume has 155 axial slices, which are separated to be used for the 2D networks. On average, out of the 155 slices, 17.2±5.82 slices are empty. For every modality, the average and standard deviation are calculated to measure the z-score of the volume. The empty slices are removed from the normalized training set but not from the validation and test sets.

The remaining slices of the training set are then augmented to increase the size of the training set and increase the variability of the data to improve the results. The augmentation transformations and their parameters are provided in Table I. Each slice is augmented ten times with these transforms to increase the size of the dataset (Fig. 2).

The image transformations use interpolation to map the voxels to the calculated coordinate. This process usually reduces the quality of the image. Calculating the mapped coordinates after each transformation and finally applying a single interpolation can prevent further image quality reduction.

TABLE I
AUGMENTATION PARAMETERS. NINE DIFFERENT TRANSFORMATIONS ARE APPLIED TO EACH SLICE WITH RANDOM PARAMETERS.

Parameter	Value
Horizontal flip	$p = 0.5$
Vertical Translation	$\|\Delta y\| < 0.1y$
Horizontal Translation	$\|\Delta x\| < 0.1x$
Vertical Scale	$0.9 \le s_y \le 1.1$
Horizontal Scale	$0.9 \le s_x \le 1.1$
Vertical Shear	$\|r_y\| \le 0.05y$
Horizontal Shear	$\|r_x\| \le 0.05x$
2D rotation	$\|\theta\| < 15°$
Elastic Transform[12]	$\alpha = 600,\ \sigma = 30$

B. Network Architecture

The used network architecture is similar to the base 2D U-Net network [2] which is a fully convolutional neural network. The model has four input channels corresponding to 4 MRI modalities in the dataset. Three separate networks are trained for three different labels, which are Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). Each of these three networks has two output channels, one for the foreground and one for the background. A softmax is applied to these two channels so that the values are between zero and one. Another network is also trained, which has three output channels for three different labels. A sigmoid function is applied to these three channels so that the values are between zero and one.

TABLE II

TEST SET DICE COEFFICIENT COMPARED TO STATE-O-THE-ART. OUR PROPOSED METHODS TEST SET DICE COEFFICIENT AND GPU ARE COMPARED TO BRATS 2020 AND BRATS 2019 RANKS. (U-NET 3C: A SINGLE U-NET WITH THREE OUTPUT CHANNELS. U-NET 2C: THREE SEPARATE U-NET NETWORKS EACH WITH TWO OUTPUT CHANNELS. U-NET 2C AU: SAME AS U-NET 2C BUT TRAINED WITH AUGMENTED DATA.)

	Method	WT	TC	ET	GPU	Memory
Proposed	U-Net 3C	0.884	0.867	0.813	NVIDIA GTX 1050TI	4GB
	U-Net 2C	0.900	0.877	0.813	NVIDIA GTX 1050TI	4GB
	U-Net 2C AU	**0.905**	**0.903**	0.822	NVIDIA GTX 1050TI	4GB
BraTS 2020 Ranks	nnU-Net[4]	0.890	0.851	0.820	RTX 2080 ti	17GB
	H^2NF-Net[10]	0.889	0.854	0.828	NVIDIA Tesla P40 GPU	11GB
	Modality-Pairing[13]	0.891	0.842	0.816	NVIDIA Tesla V100	32GB
	SA-Net[14]	0.883	0.843	0.818	NVIDIA GTX 1080 TI	11GB
BraTS 2019 Ranks	Two-stage Cascaded U-Net[9]	0.888	0.837	**0.833**	Nvidia Titan V	12GB
	Bag of Tricks[15]	0.883	0.861	0.810	Nvidia Titan V	12GB
	DeepSCAN[16]	0.89	0.83	0.81	-	-

C. Loss Function

The loss function is based on the dice coefficient. Since the GPU memory is limited, all 155 slices cannot be passed through the networks simultaneously, and a batch size of 10 is used. The dice coefficient can be calculated for each slice differently and then use the average of these dice values to calculate the loss (1c). Since the output has multiple channels and the size of the ROI in each region is different, each channel must have a different contribution to the loss (1a).

$$w_{bc} = 1 - \frac{\sum\limits_{x,y} g(b,c,x,y)}{\sum\limits_{k,x,y} g(b,k,x,y)} \tag{1a}$$

$$D_{bc} = \frac{2\sum\limits_{x,y} p(b,c,x,y)g(b,c,x,y) + s}{\sum\limits_{x,y} p(b,c,x,y) + \sum\limits_{x,y} g(b,c,x,y) + s} \tag{1b}$$

$$\mathcal{L}_{\text{dice}} = \sum_{b,c} w_{bc}(1 - D_{bc}) \tag{1c}$$

The weights, w_{bc}, are used to assign a different weight for each channel. D_{bc} is the dice coefficient calculated for each channel and sample in the batch. g and p are the ground truth and predicted image. The smoothing factor, s, is used to prevent division by zero. b, c, x, and y represent the index of samples, index of channels, index of spatial direction x, and index of spatial direction y.

The other approach is to assume that different samples are from a 3D image and calculate the weights and dice coefficients for this 3D volume. (2c).

$$w_c = 1 - \frac{\sum\limits_{n,x,y} g(n,c,x,y)}{\sum\limits_{n,k,x,y} g(n,k,x,y)} \tag{2a}$$

$$D_c = \frac{2\sum\limits_{n,x,y} p(n,c,x,y)g(n,c,x,y) + s}{\sum\limits_{n,x,y} p(n,c,x,y) + \sum\limits_{n,x,y} g(n,c,x,y) + s} \tag{2b}$$

$$\mathcal{L}_{\text{batch dice}} = \sum_{c} w_c(1 - D_c) \tag{2c}$$

D. Evaluation Metrics

The voxels of the prediction, p, are between zero and one, and the ground truth image, g, is a binary image. In the test set, the dice coefficient is calculated for each subject separately. The true positives, false positives, and false negatives are calculated for each slice and summed up. Finally, the dice coefficient for a single subject is calculated using:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3}$$

E. Training, Validation, and Testing

Three different methods are experimented:

- **U-Net 3C:** U-Net with three output channels corresponding to WT, TC, and ET regions. This is our baseline method. In this method, data augmentation is removed from the pre-processing step. A sigmoid activation function is applied to the output channels.
- **U-Net 2C:** Three separate U-Net networks with two output channels. The two output channels are the foreground and background voxels, and each of these three U-Nets segments one of the WT, TC, and ET labels. A softmax is applied to the last two channels of each network so that the voxel values show the probability of being in the background or foreground class. In this method, data augmentation is disabled to examine the effect of separating the baseline U-Net into three U-Nets.
- **U-Net 2C AU:** Three separate U-Net with enabled data augmentation. In this method, the effects of data augmentation are examined.

Each method is trained for 80 epochs using the ADAM optimizer and a batch size of 10. All experiments were performed on PyTorch 1.9.0. An NVIDIA GTX 1050 TI was used to train the networks. The validation and test sets are not augmented and are the same for all the methods examined.

IV. EXPERIMENTAL RESULTS

In this section, the results of our experiments are discussed. The dice coefficients of the examined methods and state-of-the-art are compared in Table II. The baseline U-Net (U-Net 3C) achieved dice coefficients close to BraTS 2019 ranks. It can be concluded that the frequency weighted dice

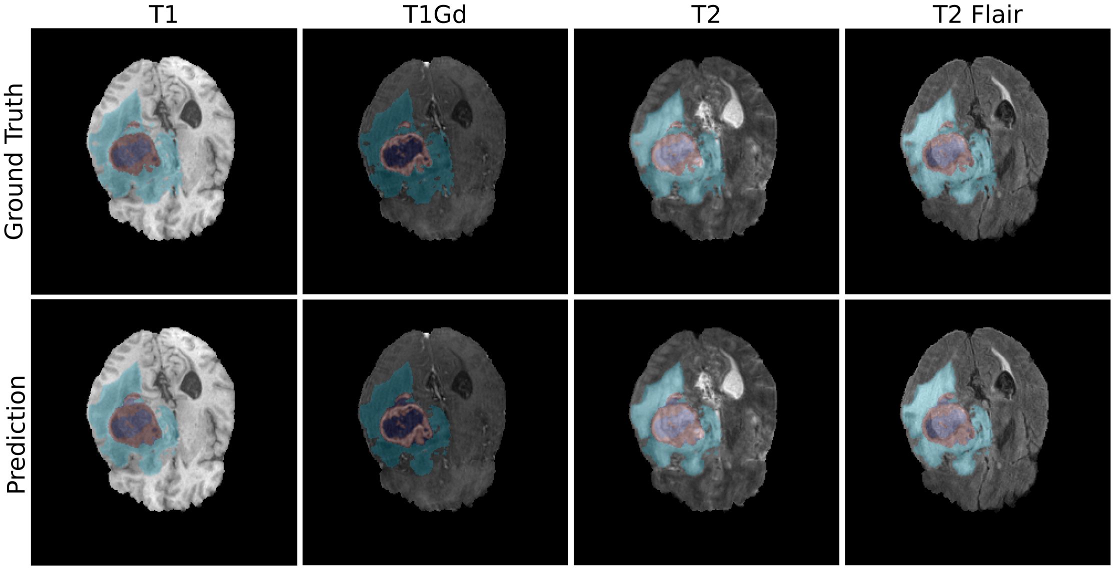

Fig. 3. Segmentation of a single axial slice using U-Net 2C AU. The predicted segmentation and ground truth masks are overlayed on four different modalities of a single axial slice.

loss has improved the results, and the baseline U-Net reached state-of-the-art accuracies.

The separated version of U-Net showed better results in segmenting whole tumor and tumor core, while the enhancing tumor segmentation dice coefficient did not change compared to baseline. The dice coefficients of WT and TC were better than state-of-the-art methods.

The third method, U-Net 2C AU, achieved the best score compared to the two other methods. This method reached the highest accuracy in segmenting WT and TC compared to the state-of-the-art. The ET segmentation dice coefficient increased with data augmentation, but the H^{2}FN-Net[10] achieved a better score. Fig. 3 shows a segmented slice from a test subject compared to the ground truth labels.

Regardless of the training time that is not important in clinical applications, all three methods have the same speed in segmenting the test data and can segment a $256 \times 256 \times 155$ MRI image in about 2 seconds. All three methods have a peak memory usage of 630MB, which is very good compared to other state-of-the-arts with much higher memory usages.

V. Discussion

The need for high-end and expensive GPUs for running state-of-the-art deep frameworks in diagnosing neurological disease, e.g. brain tumors, has limited their wide applications in clinics. This challenge is much more restrictive in low-income communities (as our local challenge). We provide a memory-efficient 2D-Unet framework for BraTS, executable on a budget GPU in a few seconds, but achieves high accuracy comparable with state-of-the-arts. We plan to release the framework as a free clinical brain imaging analysis tool.

References

[1] Florian Kofler et al., "Brats toolkit: translating brats brain tumor segmentation algorithms into clinical and scientific practice," *Frontiers in neuroscience*, 2020.

[2] Olaf Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015*, pp. 234–241.

[3] Carole H Sudre et al., "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 240–248. Springer, 2017.

[4] Fabian Isensee et al., "nnu-net for brain tumor segmentation," in *MICCAI Brainlesion Workshop*, 2020.

[5] Oldřich Kodym et al., "Segmentation of head and neck organs at risk using cnn with batch dice loss," in *German conference on pattern recognition*, 2018.

[6] Fausto Milletari et al., "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016.

[7] Nahian Siddique et al., "U-net and its variants for medical image segmentation: A review," *IEEE Access*, 2021.

[8] Kamnitsas et al., "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[9] Zeyu Jiang, "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task," in *MICCAI*, 2019.

[10] Haozhe Jia et al., "H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task," 2020.

[11] Spyridon Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint:1811.02629*, 2018.

[12] Eduardo Castro et al., "Elastic deformations for data augmentation in breast cancer mass detection," in *IEEE EMBS*, 2018, pp. 230–234.

[13] Yixin Wang et al., "Modality-pairing learning for brain tumor segmentation," *arXiv preprint:2010.09277*, 2020.

[14] Yading Yuan, "Automatic brain tumor segmentation with scale attention network," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Alessandro Crimi and Spyridon Bakas, Eds. 2021, Springer International Publishing.

[15] Yuan-Xing Zhao et al., "Bag of tricks for 3d mri brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, 2019.

[16] Richard McKinley et al., "Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation," in *MICCAI Brainlesion Workshop*, 2019, pp. 379–387.