Accurate Automatic Glioma Segmentation in Brain MRI images Based on CapsNet

M. Jalili Aziz, A. Amiri Tehrani Zade, P. Farnia, M. Alimohamadi, B. Makkiabadi, A. Ahmadian, J. Alirezaie*

***Abstract*— Glioma is a highly invasive type of brain tumor with an irregular morphology and blurred infiltrative borders that may affect different parts of the brain. Therefore, it is a challenging task to identify the exact boundaries of the tumor in an MR image. In recent years, deep learning-based Convolutional Neural Networks (CNNs) have gained popularity in the field of image processing and have been utilized for accurate image segmentation in medical applications. However, due to the inherent constraints of CNNs, tens of thousands of images are required for training, and collecting and annotating such a large number of images poses a serious challenge for their practical implementation. Here, for the first time, we have optimized a network based on the capsule neural network called SegCaps, to achieve accurate glioma segmentation on MR images. We have compared our results with a similar experiment conducted using the commonly utilized U-Net. Both experiments were performed on the BraTS2020 challenging dataset. For U-Net, network training was performed on the entire dataset, whereas a subset containing only 20% of the whole dataset was used for the SegCaps. To evaluate the results of our proposed method, the Dice Similarity Coefficient (DSC) was used. SegCaps and U-Net reached DSC of 87.96% and 85.56% on glioma tumor core segmentation, respectively. The SegCaps uses convolutional layers as the basic components and has the intrinsic capability to generalize novel viewpoints. The network learns the spatial relationship between features using dynamic routing of capsules. These capabilities of the capsule neural network have led to a 3% improvement in results of glioma segmentation with fewer data while it contains 95.4% fewer parameters than U-Net.**

I. Introduction

Gliomas are the most common fatal brain tumors caused by the abnormal growth of glial cells in the brain [1, 2]. It is including different sub-regions, i.e., pre-tumoral edema, necrotic core, enhancing and non-enhancing tumor core. Among the various medical imaging techniques, Magnetic Resonance Imaging (MRI) has become the gold standard for brain tumor diagnosis due to its ability to provide good contrast for soft tissue and the availability of a variety of additional modalities [3, 4]. A clear appearance of the glioma sub-regions, can be found in four MRI sequences, including: T1-weighted (T1), T2-weighted (T2), T1-weighted with gadolinium contrast enhancement (T1-Gd) and Fluid attenuated Inversion Recovery (FLAIR) [5]. Gliomas can appear in any part of the brain and are heterogeneous in shape, size, and appearance with blurred and irregular borders that make identifying the exact boundaries in the image exceedingly difficult [6-8].

The maximal resection of the tumor while preserving the neighboring functional non-tumoral tissue is a pivotal goal in neuro-oncological surgery. Therefore image-guided surgery is nowadays an important part of the modern neuro-oncology practice. Tumor segmentation is an essential step in surgery under image-guidance, where the image visualization and registration depend on the accurate segmentation of the tumoral tissue [9-11]. Over time, different conventional brain image segmentation methods have been developed, including manual segmentation, intensity-based methods [12], surface-based methods [13], and other deformable models. In the manual segmentation method, the segmentation process is performed by trained clinicians based on their skills. Therefore, its results depend on the user's subjective decision. Also, the time-consuming nature of manual segmentation is another limitation of these methods. However, manual segmentation is still necessary as a gold standard approach to evaluate other methods [5]. Despite many efforts that have been made to overcome the limitations of tumor segmentation algorithms, due to the highly heterogeneous nature of gliomas, conventional solutions have not been satisfactory. On the other hand, deep learning approaches has received a great deal of attention for many image processing applications such as image de-noising [14], image segmentation [15, 16], and image reconstruction [17].

Convolutional Neural Networks (CNNs), as one of the most successful deep learning methods, can provide accurate image segmentation results [18]. In this method, the network is able to learn useful features automatically, without the need for manual feature selection. CNNs have been widely used in the literature, with different approaches and have provided acceptable results in the brain tumor segmentation [5, 19-21]. Despite the recent significant achievements reported in the literature, there are four main shortcomings associated with CNNs. The first limitation is that CNNs cannot maintain the dependencies between the object parts and their totality, because of their structure design [22]. Second, the pooling layers use a type of routing which is not based on the human visual system. It routs important information which is extracted from the image to all the neurons of the next layer completely, thus details or small objects in the image are missed. The third limitation is that, in the pooling layer, the information is routed statically from one layer to the next, and the next layer of neurons is selected

This work was supported by the Faculty of Medicine, Tehran University of Medical Sciences under grant number 49513.

M. Jalili Aziz, A. Amiri Tehrani Zade, P. Farnia, and A. Ahmadian* are with the Image-Guided Surgery Group, Research Centre of Biomedical Technology and Robotics (RCBTR), Tehran University of Medical Sciences, Tehran, Iran and with the Medical Physics and Biomedical Engineering Department, Faculty of Medicine, Tehran University of Medical Sciences (TUMS).

M. Alimohamadi is with the Brain and spinal cord injury research center, Neuroscience Institute, Tehran University of Medical Sciences (TUMS), Tehran, Iran.

B. Makkiabadi is with the Medical Physics and Biomedical Engineering Department, Faculty of Medicine, Tehran University of Medical Sciences.

J. Alirezaie is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada.

*Corresponding author: ahmadian@sina.tums.ac.ir

with no intuition. In comparison, in the human visual system, this is done dynamically and neurons from the next layer can choose what information is important [23]. The fourth and most serious issue with CNNs is the amount of data necessary for training. Tens of thousands of images are required for CNN training, and preparing huge datasets for medical applications is a challenging task [24].

To overcome these limitations of CNNs, a novel class of neural networks was proposed by Sabour et al [25], where a group of neurons represents the existence of features. This group of neurons is called a "Capsule", and the network made up of these blocks is called a "Capsule Neural Network" (also called CapsNets). The vertices on a CNN network are neurons, which have a scalar representation of the output. In CapsNet, an advanced technique is used to connect the capsules between layers, which obtains the network weights based on an iterative optimization strategy. In this routing technique, the output of the previous capsule is given as input to the next capsule and during an iterative process, the similarity between the input and output of the capsule is compared. Finally, the previous capsules are routed to the capsule which has a similar output [25]. Recently, CapsNets have been used for medical image segmentation [26].

To the best of our knowledge, this is the first time that CapsNet has been optimized for glioma segmentation on MR images. In our proposed method, the network is able to train with a smaller dataset in comparison with the commonly used CNNs. This network, by using "routing by agreement", detects the relationship between parts of an object and the whole object by performing an iterative process.

II. METHOD AND MATERIAL

A. Dataset Description

Brain Tumor Segmentation (BraTS) is a well-accepted benchmark for the automatic brain tumor segmentation in multimodal MRI scans of high-grade and low-grade glioma. [27]. BraTS2020 which is used in this experiment, contains MRI scans of 369 patients with ground truth in four modalities (T1, T2, T1-Gd and FLAIR).

B. Capsule Network and routing by agreement algorithm

CapsNet is a network that identifies the spatial and hierarchical relationships between objects in images. The network is composed of several capsule layers, which are trained with an iterative algorithm. Each capsule is made up of a number of neurons and its output is a vector, which is a richer representation of the output. This vector has two general characteristics, including length and orientation. The probability that the entity indicated by the capsule corresponds to the current input is represented by the vector length. The orientation of the vector indicates the state of an object. A simple capsule structure is shown in Fig. 1. Here these weights are determined through routing by agreement iterative algorithm [25]. A temporary variable (b_{ij}) is established for updating the weights in this procedure, and it is set to zero at the start of the loop.

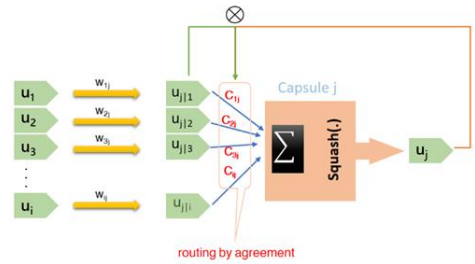

Fig. 1. Capsule structure including inputs ($\hat{u}_i$) and output ($\hat{u}_j$).

In each iteration (r) for each capsule (i) in the layer l, this variable will be updated. In Fig. 1, $u_0, u_1, u_2, ..., u_i$ represent the output vectors of previous layer capsules (layer l) that encode the existence and the state of features in low level objects. Prediction vectors ($\hat{u}_{j|i}$) were obtained by multiplying $\hat{u}_i$ by the corresponding weights. These weights show the relationship between the features obtained from lower-level capsules and higher-level capsules. C_{ij} is a scalar weighing of inputs, called "coupling coefficient", stored in C_{ij} after passing the soft-max function. Initially, all routing weights are equal, this denotes the greatest degree of uncertainty in routing the previous layer capsules' output vectors. In the next step, for the capsule j in layer $l+1$, the input vector is calculated as (1):

$$S_j = \sum_i C_{ij} . \hat{u}_{j|i} \tag{1}$$

S_j denotes the sum of weighted inputs. Next, for all capsules in layer $l+1$, to determine capsule output vector (V_j), S_j is passed through a vector-to-vector function as below (2):

$$V_j = \frac{\| S_j \|^2}{1 + \| S_j \|^2} \frac{S_j}{\| S_j \|} \tag{2}$$

This nonlinear function compresses the vector's length to the interval between zero and one, which is true in the sense that the output vector is probabilistic without affecting its direction. Then, for all capsules in the layer l and for all capsules in the layer $l+1$, the temporary variable is updated as follows (3):

$$b_{ij(new)} = b_{ij(old)} + V_i \cdot \hat{u}_{j|i} \tag{3}$$

In this equation, the correspondence between the current output and $\hat{u}_{j|i}$, is shown by a dot product operation (Fig. 1). This routing algorithm calculates the output of the capsule j during an iterative process.

In this section, the basic concepts of capsule network architecture and its routing algorithm were stated. In the next section, we will describe the segmentation capsule network (SegCaps) architecture, which is an extension of the capsule network for segmentation tasks.

C. SegCaps architecture

Due to the complexity of the CapsNet, the network encounters runtime and memory constraints for image segmentation, for the first time, LaLonde et al [28] introduced a new architecture called SegCaps based on CapsNet to address segmentation problems. In SegCaps framework, routing of lower layer capsules to the next layers is done only in a specific spatial window and the transformation matrixes are also shared between the capsules of each layer. The routing algorithm in SegCaps differs from CapsNet proposed in [25] in details.

V_j in (2) is rewritten as $V_{x,y}$; the output of the capsule at spatial coordinates (x, y) as follows (4):

$$V_{x,y} = \frac{\| P_{x,y} \|^2}{1+\| P_{x,y} \|^2} \frac{P_{x,y}}{\| P_{x,y} \|} \tag{4}$$

$P_{x,y}$ is the capsule input. The agreement is defined as dot product of output vector and corresponding prediction vectors (5):

$$a_{x,y|t_i^l} = V_{x,y} \cdot \hat{u}_{x,y|t_i^l} \tag{5}$$

Also, in this architecture, a deeper network was introduced than CapsNet, by developing the concept of de-convolutional capsules. Therefore, the dimensions of the input image will be increased to 512 × 512, in which case the segmentation task is possible.

III. Experiments and Results

A. Experimental Configuration

We implemented the proposed networks with Keras and TensorFlow framework using NVIDIA GeForce RTX 1080 TI GPU. All training is performed using dice loss and Adam optimizer. The data included MRI scans of 369 patients, of which 70 slices were used for each patient scan. Then 15% of this data is extracted randomly as a test dataset and the remaining 85% are used as training and validation set. We used the same test subset of the dataset to validate our results on SegCaps and U-Net. The image slices are cropped to 224× 224 and fed 4 channel images (including flair, T1, T1-gd, and T2) to network with random augmentation and the batch size of one. Implementation details and specific settings for each network are as follows:

- **SegCaps**: In this experiment, we randomly selected 20% of the total slices and trained the network in two steps. 80% of this subset is used for training and the remaining 20% is used as validation data. In the first phase, initially, we trained the network with a learning rate of 0.001, step decay of 1e-6 and reconstruction weight of 20. The weights of this phase are used as an initial weight for the second phase of training, via a learning rate of 1e-7 and reconstruction weight of zero.
- **U-Net:** We trained the U-Net network on 85% of the whole dataset as a training subset of the dataset. 80%of this subset is used as training and 20% as validation. The network is trained using a learning rate of 1e-7 and step decay of 1e-6 for 200 epochs.

B. Analysis of Experiments

To evaluate our proposed approach quantitatively, BraTS2020 was used. Quantitative results are presented by the performance measurement in the form of Dice Similarity Coefficient (DSC) for validation data by U-Net and SegCaps in the Table.1. The DSC is calculated as (6) where P_1 is the segmented region of the tumor and T_1 is related area in the ground truth mask [3]:

$$Dice(P,T) = \frac{| P_{1\otimes} T_1 |}{(| P_1 | + | T_1 |)/2} \tag{6}$$

For qualitative evaluations, we have shown segmentation results using two approaches, U-Net and SegCaps, in comparison with Ground Truth, from three different patients with glioma tumors on multimodality MR images in Fig. 2. The three columns contain three patients' MRI data, including T1, T1-Gd, FLAIR, and T2 sequences, in which the ground truth masks, U-Net, and SegCaps segmentation results are marked on the T2 images, respectively.

TABLE I. COMPARISION BETWEEN SEGCAPS AND U-NET ON BRATS2020 DATASET

Method	**Parameters**	**DSC %**
U-Net	31 M	85.56
SegCaps	1.5M	87.96

IV. Conclusion

To achieve accurate glioma segmentation in MR images without the need for a huge number of training datasets, we have optimized the SegCaps architecture introduced by LaLonde et al. [36]. The results of our proposed approach are compared to the results of U-Net, a commonly used network for medical image segmentation. At first, we tried to train the SegCaps using the entire dataset. However, by using different hyper-parameters settings, the DSC did not improve more than 0.7 (70%) and the network could not converge due to the complexity of the BraTS dataset. Also, the complex architecture of capsules and routing algorithms make it very challenging to achieve convergence. Then, we used a subset of the dataset and trained the network in two steps. Since the capsule network has the intrinsic capability to generalize novel viewpoints, SegCaps learn the spatial relationship between features using dynamic routing of capsules. Therefore, it is expected that the SegCaps network, which is trained using a randomly selected subset of the dataset will have comparable results to the U-Net. As we have shown in Table. 1, by using the proposed two-step training method, our experimental results show that SegCaps has about 3% improvement compared to U-Net in DSC on validation data, while using fewer data for training and containing 95.4% fewer parameters than U-Net.

It can be concluded that the main advantage of the SegCaps is overcoming the problem of data amount limitation, which is a common issue in medical datasets. This can be qualitatively observed in Fig.2 that the SegCaps has been

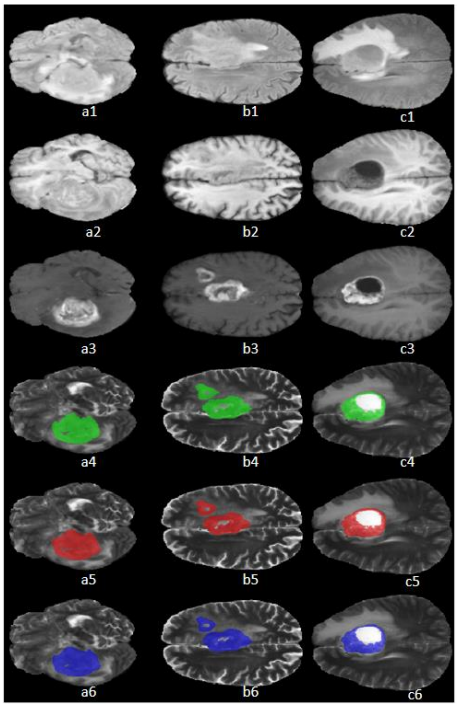

Fig. 2. Qualitative results on the BraTs2020 dataset for three different patients: **a**, **b** and **c**. For each patient, from top to bottom: FLAIR, T1, T1-Gd, T2 with true mask (Green), T2 with U-Net mask (Red) and T2 with SegCaps mask (Blue), respectively.

successful in segmenting the enhancing tumor core area. Despite capabilities of SegCaps, its routing algorithm is much slower than backpropagation. Also, its computational complexity, in addition to being time-consuming, will involve many GPU storage resources. However, since the SegCaps uses convolutional layers as a basic component, it has the potential to be optimized for complex segmentation tasks in challenging medical images, to achieve excellent results without data size limitation.

REFERENCES

1. Mellinghoff, I.K. and R.J. Gilbertson, *Brain tumors: challenges and opportunities to cure.* J Clin Oncol, 2017. **35**(21): p. 2343-2345.
2. Cui, S., et al., *Automatic Semantic Segmentation of Brain Gliomas from MRI Images Using a Deep Cascaded Neural Network.* Journal of Healthcare Engineering, 2018. **2018**: p. 4940593.
3. Gordillo, N., E. Montseny, and P. Sobrevilla, *State of the art survey on MRI brain tumor segmentation.* Magnetic resonance imaging, 2013. **31**(8): p. 1426-1438.
4. Iqbal, S., et al., *Computer-assisted brain tumor type discrimination using magnetic resonance imaging features.* Biomedical Engineering Letters, 2018. **8**(1): p. 5-28.
5. Işın, A., C. Direkoğlu, and M. Şah, *Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods.* Procedia Computer Science, 2016. **102**: p. 317-324.
6. Hoseini, F., A. Shahbahrami, and P. Bayat, *An Efficient Implementation of Deep Convolutional Neural Networks for MRI Segmentation.* J Digit Imaging, 2018. **31**(5): p. 738-747.
7. Najafzadeh, E., et al., *Application of multi-wavelength technique for photoacoustic imaging to delineate tumor margins during maximum-safe resection of glioma: A preliminary simulation study.* Journal of Clinical Neuroscience, 2019. **70**: p. 242-246.
8. Najafzadeh, E., et al., *Evaluation of multi-wavelengths LED-based photoacoustic imaging for maximum safe resection of glioma: a proof of concept study.* International Journal of Computer Assisted Radiology and Surgery, 2020. **15**: p. 1053-1062.
9. Despotović, I., B. Goossens, and W. Philips, *MRI Segmentation of the Human Brain: Challenges, Methods, and Applications.* Computational and Mathematical Methods in Medicine, 2015. **2015**: p. 450341.
10. Farnia, P., et al. *An efficient point based registration of intra-operative ultrasound images with MR images for computation of brain shift; A phantom study*. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011. IEEE.
11. Farnia, P., et al. *On the performance of improved ICP algorithms for registration of intra-ultrasound with pre-MR images; a phantom study*. in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012. IEEE.
12. Sezgin, M. and B. Sankur, *Survey over image thresholding techniques and quantitative performance evaluation.* Journal of Electronic Imaging, 2004. **13**: p. 146-168.
13. Mojtabavi, A., et al., *Segmentation of GBM in MRI images using an efficient speed function based on level set method*. 2017. 1-6.
14. Gholizadeh-Ansari, M., J. Alirezaie, and P. Babyn, *Deep learning for low-dose CT denoising using perceptual loss and edge detection layer.* Journal of digital imaging, 2019: p. 1-12.
15. Pham, D.L., C. Xu, and J.L. Prince, *Current Methods in Medical Image Segmentation.* Annual Review of Biomedical Engineering, 2000. **2**(1): p. 315-337.
16. Heidarian, S., et al., *COVID-FACT: A Fully-Automated Capsule Network-based Framework for Identification of COVID-19 Cases from Chest CT scans*. 2020.
17. Farnia, P., et al., *High-quality photoacoustic image reconstruction based on deep convolutional neural network: towards intra-operative photoacoustic imaging.* Biomed Phys Eng Express, 2020. **6**(4): p. 045019.
18. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology.* Insights into Imaging, 2018. **9**(4): p. 611-629.
19. Iqbal, S., et al., *Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN).* Microsc Res Tech, 2018. **81**(4): p. 419-427.
20. Liu, J., et al., *IOUC-3DSFCNN: Segmentation of Brain Tumors via IOU Constraint 3D Symmetric Full Convolution Network with Multimodal Auto-context.* Scientific Reports, 2020. **10**(1): p. 6256.
21. Wang, G., et al., *Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation.* Front Comput Neurosci, 2019. **13**: p. 56.
22. Bonheur, S., et al. *Matwo-CapsNet: A Multi-label Semantic Segmentation Capsules Network*. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. 2019. Cham: Springer International Publishing.
23. Dombetzki, L.A., *An overview over Capsule Networks.* 2018.
24. Jiménez-Sánchez, A., S. Albarqouni, and D. Mateus. *Capsule Networks Against Medical Imaging Data Challenges*. in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. 2018. Cham: Springer International Publishing.
25. Sabour, S., N. Frosst, and G. Hinton, *Dynamic Routing Between Capsules.* 2017.
26. LaLonde, R., et al., *Capsules for biomedical image segmentation.* Medical Image Analysis, 2021. **68**: p. 101889.
27. Menze, B.H., et al., *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).* IEEE Transactions on Medical Imaging, 2015. **34**(10): p. 1993-2024.
28. Rodney LaLonde, U.B., *Capsules for Object Segmentation.* 2018.