2020

Deep Learning with Context Encoding for Semantic Brain Tumor Segmentation and Patient Survival Prediction

Linmin Pei
Old Dominion University, lpei@odu.edu

Lasitha Vidyaratne
Old Dominion University, lvidyara@odu.edu

Md Monibor Rahman
Old Dominion University, mrahm003@odu.edu

Khan M. Iftekharuddin
Old Dominion University, kiftekha@odu.edu

Original Publication Citation

Pei, L., Vidyaratne, L., Rahman, M. M., & Iftekharuddin, K. (2020) Deep learning with context encoding for semantic brain tumor segmentation and patient survival prediction. In Hahn, H.K. & Mazurowski, M.A. (Eds.) *Medical Imaging 2020: Computer-Aided Diagnosis, 16-19 February 2020 Houston, Texas, U.S.A.* 11314OH (pp. 1-8) SPIE. https://doi.org/10.1117/12.2550693.

Deep learning with context encoding for semantic brain tumor segmentation and patient survival prediction

Pei, Linmin, Vidyaratne, Lasitha, Rahman, Md Monibor, Iftekharuddin, Khan

Linmin Pei, Lasitha Vidyaratne, Md Monibor Rahman, Khan M. Iftekharuddin, "Deep learning with context encoding for semantic brain tumor segmentation and patient survival prediction," Proc. SPIE 11314, Medical Imaging 2020: Computer-Aided Diagnosis, 113140H (16 March 2020); doi: 10.1117/12.2550693

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Deep Learning with Context Encoding for Semantic Brain Tumor Segmentation and Patient Survival Prediction

Linmin Pei, Lasitha Vidyaratne, Md Monibor Rahman and Khan M. Iftekharuddin

Vision Lab in Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA 23529

Email: {lxpei001, lvidy001, mrahm006, kiftekha}@odu.edu

ABSTRACT

One of the most challenging problems encountered in deep learning-based brain tumor segmentation models is the misclassification of tumor tissue classes due to the inherent imbalance in the class representation. Consequently, strong regularization methods are typically considered when training large-scale deep learning models for brain tumor segmentation to overcome undue bias towards representative tissue types. However, these regularization methods tend to be computationally exhaustive, and may not guarantee the learning of features representing all tumor tissue types that exist in the input MRI examples. Recent work in context encoding with deep CNN models have shown promise for semantic segmentation of natural scenes, with particular improvements in small object segmentation due to improved representative feature learning. Accordingly, we propose a novel, efficient 3DCNN based deep learning framework with context encoding for semantic brain tumor segmentation using multimodal magnetic resonance imaging (mMRI). The context encoding module in the proposed model enforces rich, class-dependent feature learning to improve the overall multi-label segmentation performance. We subsequently utilize context augmented features in a machine-learning based survival prediction pipeline to improve the prediction performance. The proposed method is evaluated using the publicly available 2019 Brain Tumor Segmentation (BraTS) and survival prediction challenge dataset. The results show that the proposed method significantly improves the tumor tissue segmentation performance and the overall survival prediction performance.

Keywords: Brain tumor segmentation, convolutional neural network, context encoding, multimodal magnetic resonance imaging, survival prediction

1. INTRODUCTION

Gliomas are the most common primary brain tumor in adults. These tumors originate from glial cells and infiltrate surrounding tissues. Brain tumors have different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e., peritumoral edematous/invaded tissue, necrotic core, active and non-enhancing core [1]. Even under treatments, the median survival period of patients diagnosed with the most aggressive form of brain tumor still remains two years or less [2]. The accurate brain tumor segmentation is important not only for treatment planning, but also for follow-up evaluation [3]. However, manual tumor segmentation is tedious, time-consuming and suffers from the inter- and intra-reader variability. Consequently, computer-aided diagnosis and automatic brain tumor segmentation is needed to address these needs. Magnetic resonance imaging (MRI) is a non-invasive technique, which offers good soft tissue contrast and very useful to assess gliomas by radiologist in clinical practice [4]. Due to the nature and appearance of brain tumors, a single MRI sequence is insufficient to capture all abnormal tissues in tumors. Mulrimodal MRI (mMRI) sequences are employed for diagnosing and delineating brain tumor segmentation in current clinical routine. The sequences usually include T1-weighted MRI (T1), T1-weighted MRI with contrast enhancement (T1ce), T2-weighted MRI (T2), and T2-weighted MRI with fluid-attenuated inversion recovery (T2-FLAIR) [4]. However, the mMRI intensity-based brain tumor segmentation is still a challenging task. Appearance of brain tumor vary widely in term of size, location, and shape across patients. Each mMRI modality has different biological information, and they must be aligned properly [2]. Figure 1 shows an example of high-grade glioma.

Due to the drawbacks of manual brain tumor segmentation, semi/fully-automatic brain tumor segmentation is required. The automatic brain tumor segmentation method can be generally categorized as four groups, threshold-based, region-based, model-based method, and pixel-based [5]. Gibbs *et al.* introduced an unsupervised method for enhancing tumor

Medical Imaging 2020: Computer-Aided Diagnosis, edited by Horst K. Hahn, Maciej A. Mazurowski, Proc. of SPIE Vol. 11314, 113140H · © 2020 SPIE
CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2550693

segmentation from T1ce images. The method requires applying an intensity threshold to a manually selected region of intensity, and segmenting image objects that have different intensity from surroundings [6]. The threshold-based method is simple and computationally fast, but it has limited applicability to tumor sub-regions. Region-based methods include region growing and watershed segmentation. Kaus *et al.* implemented a region growing tumor segmentation method which divides an image into different tissue classes based on signal intensity value [7]. Cates *et al.* proposed a watershed based brain tumor segmentation in [8]. Even though region-based method is simple and capable to correctly segment regions that have similar properties, noise and intensity inhomogeneity can result in over-segmentation. Level set is a kind of model-based method. Xie *et al.* developed a hybrid level set brain tumor segmentation method driven by region and boundary information simultaneously [9]. The expensive computation and wrong boundary converge in case of intensity inhomogeneity are two main disadvantages of using model-based method.

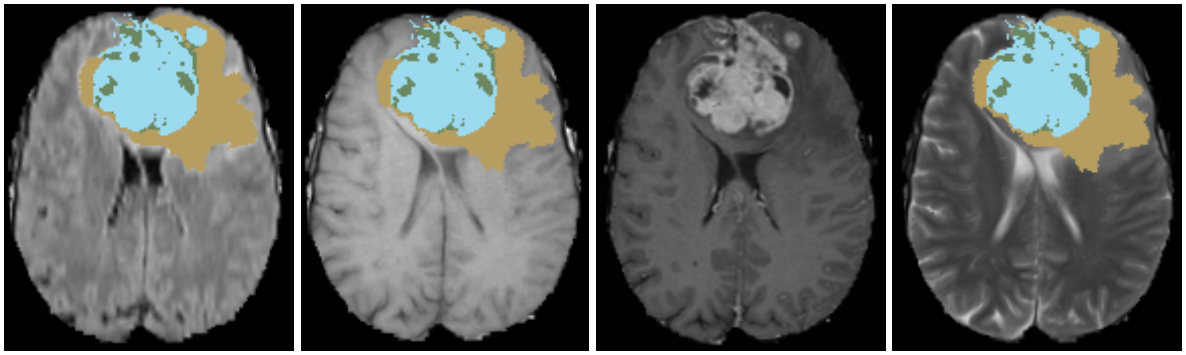

Figure 1. An example of patient with high grade glioma (HGG). From left to right: T2-FLAIR image overlaid with ground truth, T1 image overlaid with ground truth, T1ce image, and T2 image laid with ground truth. Color code: brown for edema, light blue for necrotic, and green for enhancing tumor.

Pixel-based method is widely used for brain tumor segmentation. These methods consider segmentation as a classification problem. The conventional machine learning based methods use different types of classifiers such as support vector machine (SVM), random forest (RF), etc. Bauer *et al.* proposed a fully automatic brain tumor segmentation using SVM classification in combination with hierarchical conditional random field regularization [10]. RF classifier is also used for tumor segmentation [11-14]. For these conventional feature-based machine learning methods, hand-crafted feature extraction is very challenging. In recent years, deep learning methods [15] have been successfully used for tumor segmentation in 2D patch [3] and 3D volume [16] at cost of expensive computation.

Inspired by UNet [17] and context encoding model [18], this work uses the context encoding for brain tumor segmentation. In addition, we use the front-end of the network to extract high-dimensional features, then apply a regular machine learning method to these features for overall survival prediction.

2. RELATED WORK

Context encoding network (EncNet) is proposed for semantic segmentation [18]. The EncNet captures global context information. Semantic error (SE) loss considers big and small objects equally. In brain tumor segmentation, data imbalance is a challenge that may result in underfitted model.

UNet architecture is widely used for medical image segmentation [17]. It consists of encoding and decoding paths. In the encoding path, each block has 3x3 convolution layer, followed by a rectified linear unit (ReLU) and 2x2 max pooling layer. After each block in encoding path, the feature map channel is typically doubled, and size shrinks as half. In the decoding path, it has similar structure as in encoding path, but with opposite operation. The channel of feature map is typically shrunk as half, but double its size.

In this work, we propose a novel network for brain tumor segmentation that utilizes the advantages of UNet, and also overcomes data bias using context encoding module. We refer the network as Enc-UNet. Further, to reduce computation cost and save graphics processing unit (GPU) memory, there is no skip connection between encoding and decoding part.

3. METHODS

3.1 Context encoding module

In a convolutional neural network (CNN), the high dimensional feature maps contain rich information. Encoding layer captures the global sematic context features of the scene and outputs the residual encoders with rich contextual information. In a recent work, a set of scaling factors are predicted to selectively highlight the class-dependent feature maps to utilize the context [18]. In training phase, the model is adjusted according to the difference between segmentation prediction from the ground truth. To regularize the training of context encoding module, we use a semantic encoding loss (SE-loss) to force the network understanding the global context information. To calculate the SE-loss, we construct another fully connected layer with sigmoid activation function upon the encoding layer, so that predict object classification in the image [18].

3.2 Enc-UNet

The framework for Enc-UNet is shown in Figure 2.

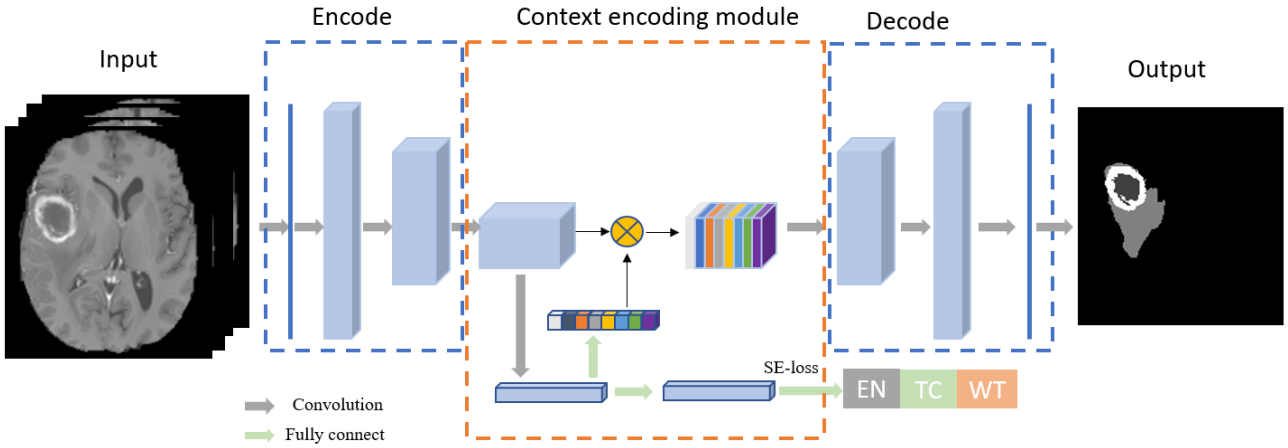

Figure 2. Overview of the proposed Enc_UNet method.

Given an input image, it is down-sampled in encoding part. The dense feature maps convert to fully connected (FC) feature in context encoding part. One branch of the FC features forms the SE-loss to regularize the model. Another branch generates the selectively highlighted class-dependent feature maps. In decoding phase, the refined feature map is up-sampled for classification. EN, TC, and WT represent segmentation of enhancing tumor, tumor core, and whole tumor, respectively. To keep global context information, we empirically crop all images with size as 160x192x128. We believe the size is large enough to keep all relevant information. The batch size is set as 1 as input data is large. The loss function is computed as $L = L_{dice} + L_{se}$, where L_{dice} is a dice loss applied to the decoder prediction output to compare with the ground truth, and L_{se} is the semantic loss. The network architecture details are listed in Table 1.

Table 1. Enc-Unet architecture details. Conv is for convolutional layer. BN is for batch normalization, and FC is for fully connected layer.

Name	Ops	Output size
Input	-	1x4x160x192x128
En-block1	Conv (4), BN, ReLU, Conv (12), BN, ReLU, Maxpooling	1x12x80x96x64
En-block2	Conv (12), BN, ReLU, Conv (24), BN, ReLU, Maxpooling	1x24x40x48x32
En-block3	Conv (24), BN, ReLU, Conv (48), BN, ReLU, Maxpooling	1x48x20x24x16
Context	Conv (96), BN, ReLU, FC (48)	1x48x1x1x1

SE-loss	FC (3)	1x3
De-block3	Conv (48), BN, ReLU, Conv (24), BN, ReLU, interpolation	1x24x40x48x32
De-block2	Conv (24), BN, ReLU, Conv (12), BN, ReLU, interpolation	1x12x80x96x64
De-block1	Conv (12), BN, ReLU, Conv (3), BN, ReLU, interpolation	1x3x160x192x128
output	-	1x3x160x192x128

4. DATA

In the experiments, mMRI data are obtained from the 2019 BraTS [1, 2, 19, 20] challenge. The training, validation, and testing data consists of 335, 125, and 166 cases, respectively. The training data contains 259 cases with high grade glioma (HGG) brain tumors, and 76 cases with low grade glioma (LGG) brain tumors. All modality sequences are co-registered, skull-stripped, denoised, and bias corrected [20]. Ground truth is also provided in the 2019 BraTS data. Each case has mMRI with size 240x240x155 for each image modality [20]. Tumor is categorized as three sub-regions: enhancing tumor (ET), tumor core (TC), and whole tumor (WT). TC consists of ET and peritumoral edema (ED), and WT is made up of TC and necrotic (NC). Note that only ground truth of training data is available for public. We train the proposed model using training data, then online evaluate using validation data, and finally apply to testing data for competition.

5. EXPERIMENTS

We implement the proposed deep neural network using PyTorch 1.0 and implement the code on high-performance cluster with Nvidia V-100 GPU.

5.1 Pre-processing and augmentation

All mMRIs are pre-processed as follows: co-registered, skull-stripped, denoised and bias corrected following BraTS protocol [20]. However, the intensity of MRI still widely varies across patients, which impacts the tumor segmentation performance using intensity-based deep learning method. Therefore, intensity normalization is required before applying our proposed method. We normalize all input images to have zero mean and unit standard deviation based on brain region only. An example of intensity normalization is shown in **Error! Reference source not found.**.

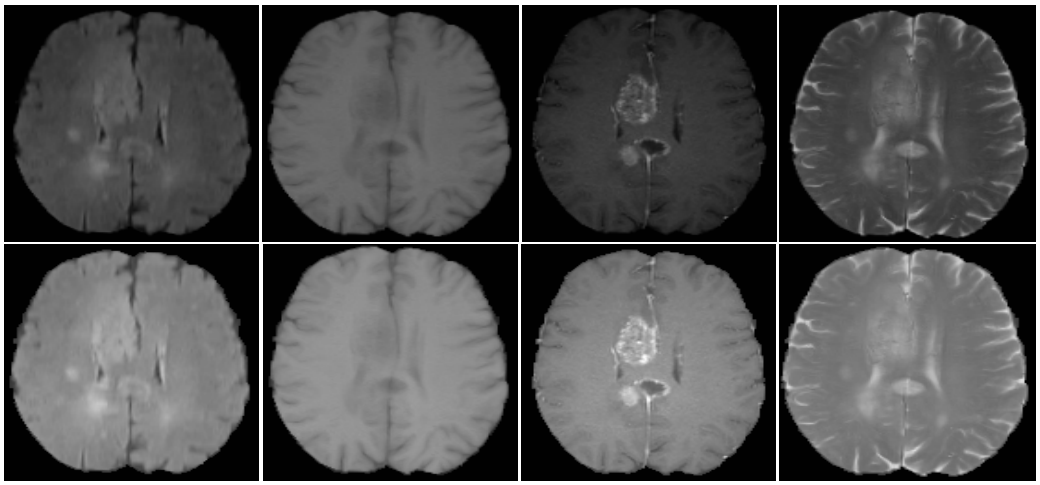

Figure 3. A case for intensity normalization. Top row: original images. Bottom row: normalized images. From left to right: T2-FLAIR, T1, T1ce, and T2 image.

In order to generate synthetic image samples, we apply intensity shift and scale (0.9~1.1) on input image channels for data augmentation.

5.2 Batch size and loss function

We empirically crop all images with size as 160x192x128. The loss function consists of 2 terms:

$$\boldsymbol{L = L_{dice} + L_{se}}, \quad , \tag{1}$$

where L_{dice} is a dice loss applied to the decoder prediction output to compare with the ground truth. Dice loss is computed as:

$$\boldsymbol{L_{dice} = 1 - DSC}, \tag{2}$$

where DSC is dice similarity coefficient [21]. The DSC is defined as,

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

where TP, FP and FN are the numbers of true positive, false positive and false negative, respectively.

5.3 Optimization

We use Adam optimizer [22] with initial learning rate of $lr_0 = 0.001$ in training phase, and the learning rate (lr_i) is gradually reduced by following:

$$\mathbf{lr_i = lr_0 * (1 - \frac{i}{N})^{0.9}}, \tag{4}$$

where i is epoch counter, and N is a total number of epochs in training.

6. RESULTS

The proposed Enc-UNet model is evaluated using brain tumor mMRI data obtained from the 2019 BraTS [1, 20] challenge datasets. The DSC and loss changes in training is shown in Figure 4. It is shown that the training loss drops from 0.9489 down to 0.1076, and the average DSC is ~0.844. In addition, we also implement UNet with variational auto-encoder regularization, known as UNet-VAE [16]. We compared the loss and training average DSC between Enc-UNet and UNet-VAE in Figure 4. Figure 5 shows an example comparison between segmentation using the proposed method and the corresponding ground truth.

The following post-processing steps are implemented following the segmentation steps. The performance of Dice Score Coefficient (DSC), and Hausdroff distance with and without post-processing are shown in Table 2. In Testing phase, there are 166 cases without grading information. The online evaluation shows that the average DSC for ET and TC in testing phase is better than that of in validation phase.

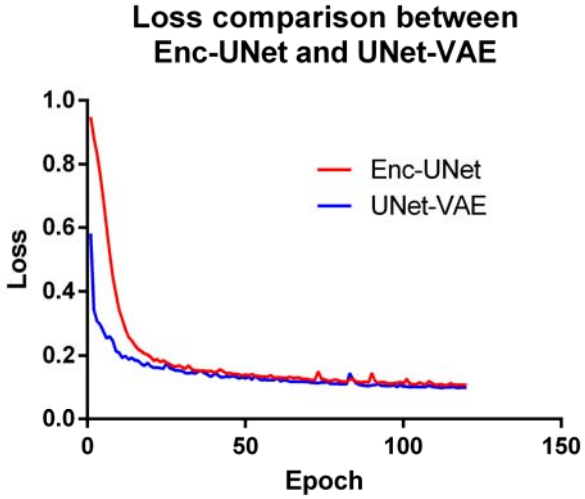

Figure 4. Comparison of loss and DSC changes in training.
Left: loss changes, right: DSC changes.

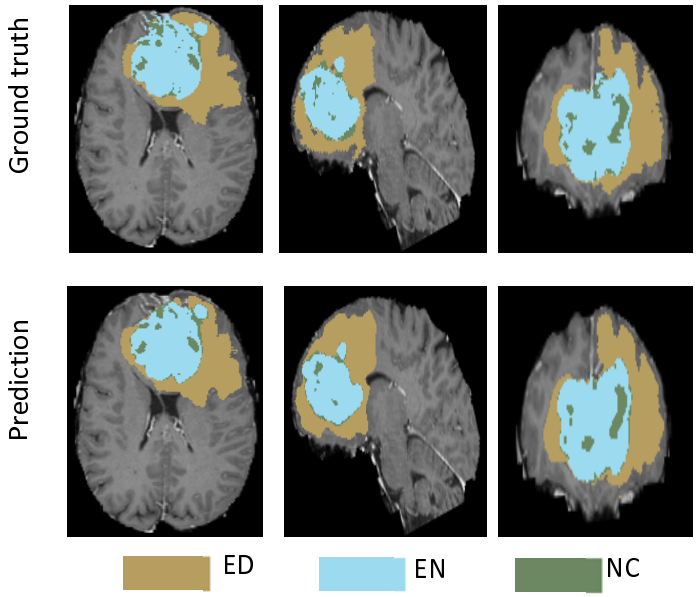

Figure 5. Comparison between tumor segmentation using the proposed method and ground truth. Top row: T1ce image overlaid with ground truth, and bottom row: T1ce image overlaid with our segmentation. From left to right: axial view, sagittal view, and coronal view, respectively.

Table 2. 2019 BraTS validation and testing online evaluation.

	DSC			Hausdorff (mm)		
2019 BraTS validation	ET	WT	TC	ET	WT	TC
Without post-processing	0.69595	0.89563	0.79531	5.99258	6.92109	8.19885
With Post-processing	0.74895	0.89478	0.79669	3.84836	5.86472	7.90552
2019 BraTS Testing	0.8133	0.8867	0.84031	2.22904	4.78833	4.14805

We use two different methods for survival prediction: hand-crafted feature-based and feature fusion-based. For hand-crafted feature-based method, we use conventional machine learning based method to extract texture features, such as gray level co-occurrence matrix (GLCM). For the feature fusion-based method, we fuse the hand-crafted features with high dimensional features by CNN. After feature extraction, feature selection is applied, then followed by linear regression for survival prediction is obtained. In the both cases, age is used as an additional feature. We applied both methods to the 2019 BraTS validation dataset, and the performance is reported in Table 3. The performance comparison shows promising performance by using feature fusion-based method which provides a higher accuracy and less mean square error (MSE). Note this survival prediction result is currently among the top three groups as posted on the 2019 BraTS competition web (https://www.cbica.upenn.edu/BraTS19//lboardValidationSurvival.html.) [23]. In testing phase, our online evaluation of total 107 case has accuracy 0.439 with mean square error (MSE) 449009.785.

Table 3. Survival prediction comparison between using hand-crafted and fused features.

Method	Accuracy (%)	MSE
Hand crafted	48.3	110496.3
Fusion	51.7	70590.75

7. NOVEL CONTRIBUTION

The novelty of this work lies in the introduction of context encoding in a state-of-the-art deep CNN based U-Net model, Enc-UNet, for semantic tumor segmentation with mMRI. An additional contribution is the utilization of the resulting context augmented DL features to significantly improve the performance of a state-of-the-art survival prediction pipeline.

8. CONCLUSION

The work proposes a new deep learning neural network Enc-UNet for brain tumor segmentation. The model takes advantage of context encoding, which captures global context features. Moreover, we use SE-loss to regularize the model. Unlike pixel-wise loss, it treats large and small class objects equally that can overcome the sample bias issue. The concept of SE-loss is very suitable for mMRI brain tumor segmentation dataset, which usually suffers from sample bias. We apply the proposed Enc-UNet to 2019 BraTS dataset. The result suggests that the performance of brain tumor segmentation and patient survival prediction using the proposed method is competitive as shown in the Leaderboard results on the BraTS web. In future, we plan to improve the proposed model by adding skip connections between encoding and decoding part.

9. ACKNOWLEDGEMENTS

We acknowledge partial funding of this work by a NIBIB/NIH grant# R01 EB020683.

REFERENCES

[1] S. 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 arXiv:1811.02629,* 2018.

[2] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging,* vol. 34, no. 10, pp. 1993-2024, 2014.

[3] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE transactions on medical imaging,* vol. 35, no. 5, pp. 1240-1251, 2016.

[4] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine & Biology,* vol. 58, no. 13, p. R97, 2013.

[5] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *Magnetic resonance imaging,* vol. 31, no. 8, pp. 1426-1438, 2013.

[6] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, "Tumour volume determination from MR images by morphological segmentation," *Physics in Medicine & Biology,* vol. 41, no. 11, p. 2437, 1996.

[7] M. R. Kaus, S. K. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, "Automated segmentation of MR images of brain tumors," *Radiology,* vol. 218, no. 2, pp. 586-591, 2001.

[8] J. E. Cates, R. T. Whitaker, and G. M. Jones, "Case study: an evaluation of user-assisted hierarchical watershed segmentation," *Medical Image Analysis,* vol. 9, no. 6, pp. 566-578, 2005.

[9] K. Xie, J. Yang, Z. Zhang, and Y. Zhu, "Semi-automated brain tumor and edema segmentation using MRI," *European journal of radiology,* vol. 56, no. 1, pp. 12-19, 2005.

[10] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field

regularization," in *international conference on medical image computing and computer-assisted intervention*, 2011, pp. 354-361: Springer.
[11] S. Reza and K. Iftekharuddin, "Multi-class abnormal brain tissue segmentation using texture," *Multimodal Brain Tumor Segmentation,* vol. 38, 2013.
[12] L. Pei, S. Bakas, A. Vossough, S. M. Reza, C. Davatzikos, and K. M. Iftekharuddin, "Longitudinal brain tumor segmentation prediction in MRI using feature and label fusion," *Biomedical Signal Processing and Control,* vol. 55, p. 101648, 2020.
[13] L. Pei, S. M. Reza, W. Li, C. Davatzikos, and K. M. Iftekharuddin, "Improved brain tumor segmentation by utilizing tumor growth model in longitudinal brain MRI," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017, vol. 10134, p. 101342L: International Society for Optics and Photonics.
[14] L. Pei, S. M. Reza, and K. M. Iftekharuddin, "Improved brain tumor growth prediction and segmentation in longitudinal brain MRI," in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2015, pp. 421-424: IEEE.
[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, p. 436, 05/27/online 2015.
[16] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, 2018, pp. 311-320: Springer.
[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241: Springer.
[18] H. Zhang *et al.*, "Context encoding for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7151-7160.
[19] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)," ed, 2017.
[20] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific data,* vol. 4, p. 170117, 2017.
[21] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology,* vol. 26, no. 3, pp. 297-302, 1945.
[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.
[23] M.-B. 2019, "https://www.cbica.upenn.edu/BraTS19/lboardValidation.html," *MICCA-BraTS 2019 Leaderboard,* 2019.