Deep Learning Methods for MRI Brain Tumor Segmentation: a comparative study

Ikram Brahim, Dominique Fourer, Vincent Vigneron, Hichem Maaref

Deep Learning Methods for MRI Brain Tumor Segmentation: a comparative study

Ikram Brahim, Dominique Fourer, Vincent Vigneron and Hichem Maaref
IBISC, EA 4526
University of Évry/Paris-Saclay, France
ikrammeki@gmail.com, dominique@fourer.fr

***Abstract*—Brain tumor segmentation from MRI is an important task in biomedical image processing that can help specialists to predict diseases and to improve their diagnoses. Nowadays, most of the state-of-the-art techniques are based on deep learning neural networks for which the choice of the best architecture remains an open question. Hence, this paper aims at providing answers through an intensive and comprehensive comparison between several promising neural network architectures. Our study leads us to three approaches which are respectively based on 2D U-Net, 3D U-Net and cascaded neural networks, that are compared together and with another unsupervised technique based on k-mean clustering. We also consider several enhancement techniques such as data augmentation, curriculum learning and an original boosting method based on majority voting. We achieve to improve the results of the baseline methods in terms of Dice score when the suitable combination of techniques is used.**

***Index Terms*—Brain tumor segmentation, deep learning, MRI, curriculum learning.**

I. Introduction

Nowadays, image processing techniques can provide excellent results in several tasks such as classification, detection or segmentation. In biomedical applications, such tools are useful for specialists to decision aids. More precisely, the segmentation of brain tumors from multiple modalities can produce predictions that are full of interest for surgical planning, postoperative analysis or radiotherapy [1], [2]. Brain tumors are dangerous diseases which require early detection and sometimes a drawn-out treatment. They can be benign or malignant in the presence of cancerous cells. Primary malignant brain tumors are most frequently gliomas, or primary central nervous system lymphomas. In such case, it has a faster growth rate and is more threatening. Benign tumors are slower in growth including low-grade variants and do not require an immediate treatment. Both cases always require neuroimaging prior, during and after treatment where automatic medical imaging processing aids in evaluating the tumor progression, surgical planning, and overall treatment [3].

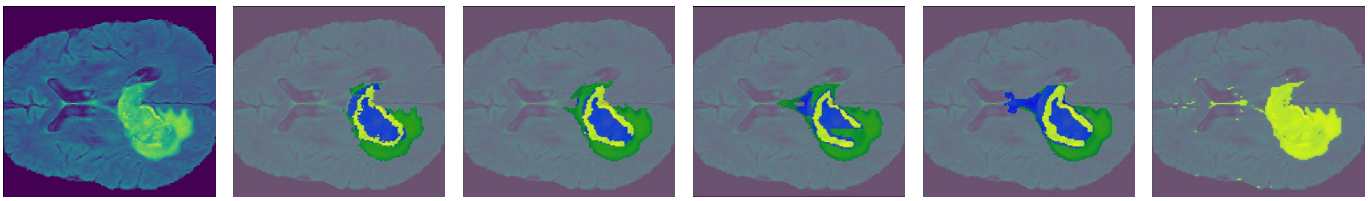
The information acquired from the images provides useful information to physicians. Shape, size, location, and metabolism of brain tumors are relevant in the diagnosis [4]. Thus, brain MRI is a reliable imaging technique with a significant influence on medical image analysis [5]. The challenging tasks for artificial intelligence include detection, regression and classification. To this end, deep learning techniques recently gained interest by achieving excellent results in image recognition [6]. Since then, medical images segmentation, detection and classification also benefits of recent advances based on deep learning. In fact, for various applications, Deep Neural Network (DNN) techniques and more specially those involving Convolutional Neural Networks (CNN) are among the best performing [7]. Some of them include breast ultrasound image segmentation [8], lung segmentation for dynamic chest radiography [9] or detection of complete anterior cruciate ligament tear [10].

This study focuses on the brain tumor segmentation from MRI task for which public annotated datasets allow reproducible research. This enables to comparatively assess the efficiency of three distinct and promising deep learning architectures selected from the top-ten methods of the BraTS'17 challenge [11]. Now we aim at motivating the choice of an architecture in comparison to another by presenting their advantages in an image segmentation problem while considering the latest advances based on deep learning [1], [2], [7]. Our study also investigates several enhancement techniques based on data augmentation and boosting which can improve the results when combined with an existing method. This paper is organized as follows. In Section II, the segmentation problem from MRI is formulated with the considered experimental dataset. Section III describes the investigated MRI brain tumor segmentation methods for which several enhancement techniques are proposed in Section IV. The numerical results are presented in Section V and discussed in the last section.

II. Materials

A. Problem formulation

We aim at designing a method considering an MRI as input and which provides a segmentation mask as output. An MRI is a 3D volume corresponding to a set of several images associated to the slices of the observed patient brain. Each slice can exist in several modalities (e.g. T1, T2, flair) which correspond to the same observation but with different weights applied on each pixel. Hence, the investigated methods automatically compute a segmentation mask from the available information. The investigated task considers several sub-regions of the tumor that should be segmented: whole tumor (W.T), tumor core (T.C) and enhancing tumor (E.T). An illustration of the addressed task is presented in Fig. 1. The investigated supervised methods use a training dataset of fully annotated MRI and are evaluated on a distinct test set.



(a) Reference (FLAIR) (b) Ground Truth (c) 2D U-Net+CL (d) 3D U-Net (e) Cascaded Network (f) K-Means

Fig. 1: Brain tumor segmentation of an MRI slice using different methods. The different sub-regions correspond to whole tumor region (green+yellow+blue), tumor core region (yellow+blue) and enhancing tumor region (blue).

B. The BraTS dataset

The multimodal Brain Tumor Segmentation (BraTS) challenge [11] compares each year the best new methods and uses a publicly available dataset[1] allowing a transparent and reproducible biomedical research. This study uses the BraTS'17 dataset which contains clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG) of 285 patients. The data of each patient contains several modalities native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), Fluid Attenuated Inversion Recovery (FLAIR) volumes. The ground truth is also provided and corresponds to a manual segmentation that is approved by neuro-radiologists. Since, the test set used by the segmentation challenge is not available, the present study uses our own split of the provided training dataset.

III. Methods

A. 2D U-Net

The U-Net was first proposed for biomedical image segmentation by Ronneberger *et al.* in 2015 [12] as an improvement on the fully convolutional neural networks for semantic segmentation [13]. This architecture contains two paths respectively called encoder and decoder made of several convolutional and max-pooling layers at the encoder and transposed convolution (up-conv) layers at the decoder. The encoder is also called the contraction path and the decoder is the expanding path. They follow the idea of an autocoder to find a latent representation of a lower dimension than the input that is used for the segmentation task. Our model follows the U-Net architecture proposed by Dong *et al.* [14] depicted in Fig. 2. Unlike the originally proposed U-Net, our model uses zero-padding to preserve the dimension of the output at each layer that enables more flexibility for the dimension of the input. For the training, we use the Dice Loss Function (DLF) given by Eq. (2) instead of the similarity coefficients suggested in the original method.

B. 3D U-Net

Cicek *et al.* [15] extended the U-Net network for volumetric segmentation. The inputs are the voxels from volumetric images and the resulting output is a 3D segmentation mask. Since medical imaging is frequently presented in the three-dimensional realm, adjusting the network to take 3D input

[1]https://www.med.upenn.edu/sbia/brats2017/data.html

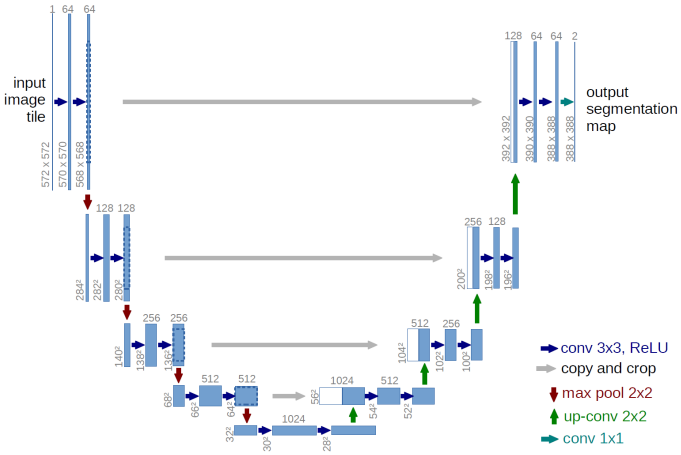


Fig. 2: 2-Dimensional U-Net Architecture

reduces the repetitive information. When fed slice-by-slice, the neighboring slices often contain the same information. The proposed 3D U-Net is based on the original U-Net architecture that also includes the encoder and decoder paths. The whole image is analyzed in the contracting path and subsequent expansions produce the final segmentation. Some of the changes include having all the operations in 3D and using batch normalization which was shown to improve the training convergence [16]. Another difference is the reduction in the number of blocks in each path from five to four. The Dice loss function in Eq. (2) was also used for the training of this network. Fig. 3 illustrates the proposed architecture where the encoder path blocks contain two 3D convolutions followed by a Rectified Linear Unit (ReLU), and a 2 x 2 x 2 max-pooling with strides of two. The decoder path blocks include 2 x 2 x 2 transposed convolution (up-conv) by strides of two in each dimension and two 3D convolutions followed by a ReLU. Shortcut connections between paths enable high-resolution features from encoder to decoder path.

C. Cascaded Anisotropic Network

The cascade of three convolutional neural networks proposed by Wang *et al.* [17] won the second place of MICCAI 2017 BraTS challenge. The contributions include a combination of three CNNs that segment each of three sub-regions sequentially: Whole tumor, Tumor core and Enhancing tumor. Hence, anisotropic convolutions (*i.e.* dependent on the

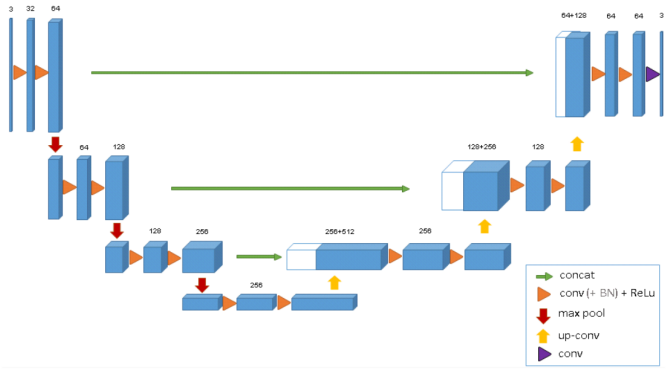


Fig. 3: 3-Dimensional U-Net Architecture

direction) are used to deal with 3D MRI but it results in a higher model complexity and memory consumption. Lastly, the fusion of the CNN outputs in three orthogonal views (*i.e.* axial, sagittal and coronal) is used to enhance the segmentation of the brain tumor. The three CNNs follow the hierarchical structure of the tumor sub-regions as depicted in Fig. 4.

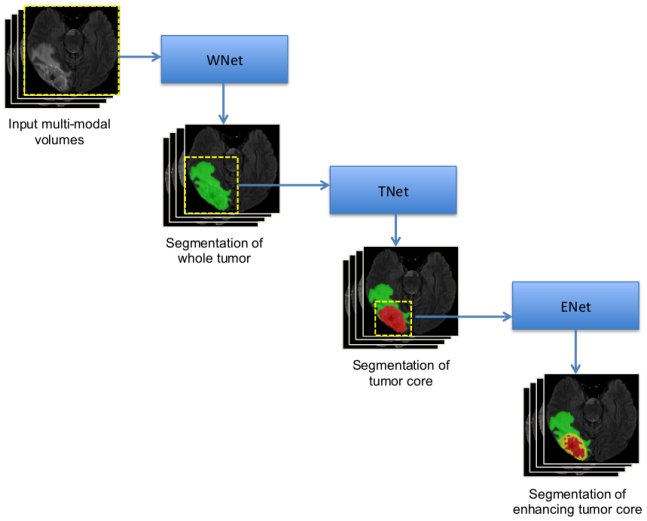


Fig. 4: Triple Cascaded framework for segmentation of brain tumor and its sub-regions [17].

As shown in Fig. 5, WNet and TNet have the same architecture while ENet only uses one downsampling layer due to the smaller input size. The WNet takes the full MRI as input and segments the first region: whole tumor. A corresponding bounding box (*cf.* Fig. 4) is computed and used as the input of the TNet that segments the tumor core similarly used for the ENet. The bounding boxes allow a restriction of the segmentation region and minimize false positives and false negatives. One of the cascade's drawback is that it is not an end-to-end method and thus training and testing time is longer than with the other methods.

D. K-means : Unsupervised MRI segmentation

K-means clustering algorithm [18] is an unsupervised method that is full of interest due to its simplicity and relatively low computational complexity. K-Means clustering divides a dataset into K (integer) homogeneous clusters. K corresponds to the number of inertia centers (or centroids) assigned to each cluster. Hence, for a set of points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, with $n > K$, the algorithm separates the data points into K subsets $S = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_K\}$ by minimizing the intra-class distance. After a random initialization of the centroids, the algorithm iteratively assigns each point to a cluster $\mathcal{S}_k$ and updates its corresponding centroid C_k until convergence. The result S minimizes the overall distance of each point to its cluster that can be expressed as:

$$S = \underset{S'}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{x_i \in \mathcal{S}'_k} \|x_i - C_k\|^2 \qquad (1)$$

This method was shown suitable for biomedical image segmentation as the number of regions to be segmented is usually known. In our experiments which aim at segmenting three subregions of the tumor, we set $K = 3$ to obtain a segmentation of the whole tumor.

IV. Proposed Enhancement Techniques

A. Data augmentation

Data augmentation is well known technique which aims at improving the robustness of a machine learning model by artificially increasing the size of the reference training dataset [19]. To this end, we should consider a set of data transformations for which we want our trained model to be invariant. In this study, the following geometrical transformations are used with randomly chosen settings:

- Rotation (90 degrees)
- Horizontal flip
- Vertical flip
- Cropping
- Gaussian white noise

B. Curriculum Learning

Curriculum learning [20] is a promising paradigm which was first proposed to deal with non-convex optimization to avoid the local optimum issue. The intuition behind curriculum learning is to mimic the learning of human with a gradual training process with examples sorted in an increasing level of difficulty. Following this idea, we propose to pre-train the considered models from artificially downsampled MRIs by a progressive increasing level of resolution. This enhancement was carried out by downsampling then upsampling by successive factors equal to eight, four and two. Hence, the first model is trained with the data that is downsampled/upsampled by a factor eight. Once saved, it is re-trained with the data that is downsampled/upsampled by a factor four. This process is then repeated with the data downsampled/upsampled by a factor two. Finally, the resulting model is trained with the data in its original resolution.

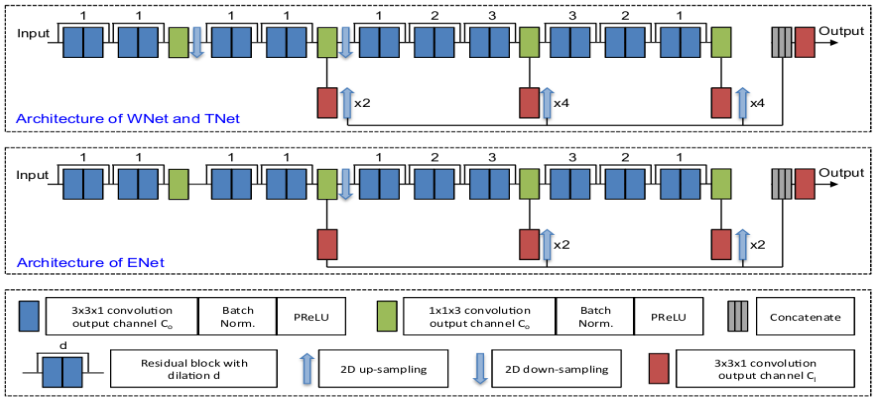


Fig. 5: WNet, TNet and ENet with dilated convolution, residual connection and multi-scale fusion [17].

C. Boosting

In order to simultaneously take benefit of all the investigated methods, this proposal consists in developing an original improved method which combines all the resulting predictions provided by each technique (*i.e.* 2D U-Net, 3D U-Net, Cascaded network). Our approach simply considers an equally-weighted majority voting applied to each pixel of the input MRI. For the prediction, all the methods have the same relevance (or weight) to assign a score to each prediction. The final decision is set to use the prediction which obtains the highest voting score. If several different predictions obtain an identical score, the final prediction is randomly chosen among the best proposed choices.

V. Numerical Experiments

A. Implementation

The investigated methods have been implemented in Python using `Tensorflow` and Keras frameworks except for the cascaded network for which the authors provides an implementation based on Niftynet[2] [17]. Our computation were completed on a Intel Xeon W-2133 CPU @ 3.60 Ghz with 32 GB of RAM and a NVIDIA GTX 1080 GPU. For the sake of reproducibility, the used BraTS'17 dataset can be downloaded at https://fourer.fr/publi/datasets/BRATS17.tar.gz and our produced Python code can be found at: https://github.com/ikramabdel/tumorsegmentation.

B. Experiment Details

In the first experiment, we considered the four methods: 2D U-Net, 3D U-Net, Cascaded Network and K-means for which the training details are presented in Table I. The second experiment combines the best method in terms of the highest Dice (*i.e.* 2D U-Net) with the proposed enhancement methods. Hence, the results of the retrained model are presented following a curriculum learning (CL) as well and through Data augmentation (DA). The third experiment considers the equally weighted majority voting performed using the 3D U-Net, Cascaded Network and the best performing model (*i.e.* 2D U-Net + CL) from the second experiment. When used, all the DA and CL transformation are applied on 25% of the initial training dataset.

1) Data: From the original BraTS dataset, we selected 100 patients for the training-validation set and 25 patients for test set. Within the set of 100 patients, 75 were used for training, and 25 for the validation. To improve the computation efficiency of our evaluation, each MRI of the dataset was cropped from 240x240x155 to 144x160x60, removing background region pixels.

2) Training protocol: All the four supervised methods were trained using the Dice Loss Function (DLF):

$$\mathrm{DLF}(P,T) = 1 - \frac{2\sum_i P_i \times T_i}{\sum_i P_i + \sum_i T_i + \epsilon} \tag{2}$$

where P denotes the set of the predicted labels (P_i being the i-th element) and T the set of the corresponding ground truth. We arbitrary defined $\epsilon = 1$ to deal with the particular case when P and T only contain background values equal to zero. The 2D and 3D U-Net were trained with 300 epochs while the Cascaded Anisotripic Network was only trained for 30 epochs due to time restrictions. The network requires separate training for each region and each of the three views which lengthens training time.

C. Objective Evaluation Metrics

Our results presented in Tables II show the efficiency of each method measured in terms of Sørensen-Dice indice, Sensitivity, Specificity and Hausdorff distance that can be formulated as (card () standing for cardinality):

$$\text{Dice} = \frac{2\,\mathrm{card}\,(P \bigcap T)}{\mathrm{card}\,(P) + \mathrm{card}\,(T)} \tag{3}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

with TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives and Sensitivity = Recall.

[2]https://niftynet.io

The Hausdorff distance is commonly used to measure how far can be the extrema of two compared sets and provides an information about the presence of outliers:

$$\text{Haus}(P,T) = \max\left\{\sup_{p\in P}\inf_{t\in T} d(p,t), \sup_{t\in T}\inf_{p\in P} d(p,t)\right\} \quad (6)$$

where $d(p,t)$ is defined as the Euclidean distance between the pixel coordinates of p and t. Additional results presented in III and IV detail the metrics of the confusion matrix for each segmented region expressed as:

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}$$

$$\text{F1-Score} = 2\frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (7)$$

VI. Results & Discussion

According to Table II, the 2D U-Net obtains the highest Dice scores for the three sub-regions during the first experiment: whole tumor = 0.78, tumor core = 0.65 and enhancing tumor = 0.46. Moreover, Table III also shows that although the 2D U-Net does not have the lowest false negatives, and false positives for all three regions since scores relative to the Cascaded, 3D U-Net and Unsupervised methods are not drastic. Given this result the 2D U-Net was chosen for the enhancement experiments: Curriculum Learning (CL) and Data Augmentation (DA). As for the equally weighted majority voting the prediction of the first three methods were used to obtain the final prediction. The enhancement of the 2D U-Net results shows that the three proposed enhancement methods improve the Dice score and the sensitivity but the 2D U-Net trained with curriculum learning outperforms the others. However, the Hausdorff distance was not significantly reduced by the proposed enhancement methods. The detailed results presented in Table III confirm the superiority of the 2D U-Net in terms of confusion matrix metrics when trained with curriculum learning.

VII. Conclusion

We have presented and compared together several deep neural network-base methods applied to MRI brain tumor segmentation. Our results confirmed the superiority of the deep learning approach over a unsupervised approach. Moreover, we show that the proposed curriculum learning is an efficient strategy to easily improve the training efficiency of a deep neural network model. In fact, this approach provides the best results of this study when combined with a classical 2D U-Net architecture. Hence, we conclude that 2D U-Net remains a suitable choice for the segmentation of brain tumor with its sub-regions in comparison to more complicated architectures that are longer to train. Future work will consider semi-supervised and weakly-labelled configurations which are more challenging and require different strategies and architectures for obtaining acceptable results.

References

[1] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain MRI segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017.

[2] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[3] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[4] Ali Işın, Cem Direkoğlu, and Melike Şah. Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102:317–324, 2016.

[5] Lawrence O Hall, Amine M Bensaid, Laurence P Clarke, Robert P Velthuizen, Martin S Silbiger, and James C Bezdek. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *IEEE transactions on neural networks*, 3(5):672–682, 1992.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[8] Min Xian, Yingtao Zhang, Heng-Da Cheng, Fei Xu, Boyu Zhang, and Jianrui Ding. Automatic breast ultrasound image segmentation: A survey. *Pattern Recognition*, 79:340–355, 2018.

[9] Yuki Kitahara, Rie Tanaka, Holger R Roth, Hirohisa Oda, Kensaku Mori, Kazuo Kasahara, and Isao Matsumoto. Lung segmentation based on a deep learning approach for dynamic chest radiography. In *Proc. SPIE 10950, Medical Imaging: Computer-Aided Diagnosis*, 2019.

[10] Peter D Chang, Tony T Wong, and Michael J Rasiej. Deep learning for detection of complete anterior cruciate ligament tear. *Journal of digital imaging*, pages 1–7, 2019.

[11] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2015.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proc. IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[14] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.

[15] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Int. conf. on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[16] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[17] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. *Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks*, volume 10670 LNCS of *Lecture Notes in Computer Science*, pages 178–190. Springer Verlag, Germany, 2 2018.

[18] J. MacQueen. Some methods for classification and analysis of multivariate observations. 1967.

[19] David A Van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50, 2001.

[20] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proc. 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

TABLE I: Experiment Details.

Method	Loss Function	Training Set Size	# of Trainable Parameters	Test Set Size	Epochs	Training duration (hours)
Cascaded Network	Dice Loss	100	n/a	25	30	9
2D U-Net	Dice Loss	100	31,032,451	25	300	**13**
3D U-Net	Dice Loss	100	14,491,619	25	300	24

TABLE II: Overall comparative results.

Methods	Dice Score			Sensitivity			Specificity			Hausdroff Dist.		
	W.T	T.C	E.T	W.T	T.C	E.T	W.T	T.C	E.T	W.T	T.C	E.T
Cascaded Network	0.75	0.59	0.58	0.75	0.66	0.73	0.78	0.74	0.80	**4.91**	4.39	3.61
2D U-Net	0.78	0.65	0.65	0.80	0.74	0.77	0.83	0.81	0.81	5.18	**4.23**	3.46
3D U-Net	0.73	0.61	0.46	0.81	0.65	0.72	**0.87**	0.77	0.86	5.77	4.64	4.13
K-Means	0.55	-	-	0.53	-	-	0.61	-	-	5.42	-	-
2D U-Net + CL	**0.81**	**0.73**	**0.68**	0.81	**0.79**	0.77	0.84	**0.84**	0.80	5.18	4.35	**3.38**
2D U-net+DA	0.80	0.71	0.67	0.80	0.78	0.77	0.83	**0.84**	0.81	5.10	4.27	3.40
2D U-Net + DA + CL	0.78	0.67	0.65	**0.82**	0.77	**0.78**	0.85	0.83	**0.82**	5.79	4.61	3.61
Majority Voting	**0.81**	0.69	0.66	**0.82**	0.73	0.76	0.86	0.80	0.82	5.06	4.32	3.44

with CL = Curriculum Learning, DA = Data Augmentation

TABLE III: Additional Results : Confusion Matrix metric

Methods	False Negatives			False Positives			True Positives			True Negatives		
	W.T	T.C	E.T	W.T	T.C	E.T	W.T	T.C	E.T	W.T	T.C	E.T
Cascaded Network	0.28	0.35	0.25	**0.0095**	0.0079	0.0031	**0.83**	0.68	0.67	0.98	**0.99**	**0.99**
2D U-Net	0.20	0.24	0.24	0.011	0.0080	0.0020	0.80	0.67	0.78	**0.99**	**0.99**	**0.99**
3D U-Net	**0.15**	0.30	**0.16**	0.021	0.011	0.0054	0.63	0.55	0.42	**0.99**	**0.99**	**0.99**
K-Means	0.61	-	-	0.018	-	-	0.69	-	-	0.94	-	-
2D U-Net + CL	0.19	**0.19**	0.24	0.011	**0.0072**	**0.0018**	0.81	**0.70**	**0.81**	**0.99**	**0.99**	**0.99**
2D U-Net + DA	0.21	**0.19**	0.23	0.010	0.0077	0.0020	0.82	0.68	0.79	**0.99**	**0.99**	**0.99**
2D U-Net + DA + CL	0.17	0.20	0.21	0.013	0.0078	0.0023	0.77	0.68	0.75	**0.99**	**0.99**	**0.99**
Majority Voting	0.16	0.25	0.22	0.013	0.0079	0.0026	0.77	0.67	0.72	**0.99**	**0.99**	**0.99**

TABLE IV: Additional Results

Methods	Precision			Accuracy			F1 Score		
	W.T	T.C	E.T	W.T	T.C	E.T	W.T	T.C	E.T
Cascaded Network	**0.99**	**0.99**	**0.99**	0.86	0.82	0.87	0.85	0.79	0.84
2D U-Net	**0.99**	**0.99**	**0.99**	0.90	0.87	0.88	0.88	0.85	0.87
3D U-Net	0.97	0.98	0.98	0.90	0.83	**0.90**	0.88	0.78	0.83
K-Means	0.97	-	-	0.72	-	-	0.69	-	-
2D U-Net + CL	**0.99**	**0.99**	**0.99**	0.90	**0.90**	0.88	0.89	**0.88**	0.87
2D U-Net + DA	**0.99**	**0.99**	**0.99**	0.90	0.89	0.89	0.88	0.87	0.87
2D U-Net + DA + CL	0.98	**0.99**	**0.99**	0.90	0.89	0.89	0.89	0.87	**0.90**
Majority Voting	0.98	**0.99**	**0.99**	**0.91**	0.87	0.88	**0.90**	0.84	0.86

with W.T= Whole Tumor, T.C = Tumor Core, E.T=Enhancing Tumor.