

# A System for Brain Image Segmentation and Classification Based on Three-Dimensional Convolutional Neural Network

Ahmed Kharrat<sup>1</sup>, Mahmoud Neji<sup>2</sup>

<sup>1</sup> University of Sfax, MIRACL Laboratory ISIMS, Sakiet Ezzeit, Sfax, Tunisia

<sup>2</sup> University of Sfax, MIRACL Laboratory FSEG, Elmatar, Sfax, Tunisia

ahmed.kharrat@isims.usf.tn, mahmoud.neji@fseg.rnu.tn

**Abstract.** We consider the problem of fully automatic brain tumor segmentation in MR images containing glioblastomas. We propose a three Dimensional Convolutional Neural Network (3D-CNN) approach that achieves high performance while being extremely efficient, a balance that existing methods have struggled to achieve. Our 3D-Brain CNN is formed directly on raw image modalities and thus learn a characteristic representation directly from the data. We propose a new cascading architecture with two pathways that each model normal details in tumors. Fully exploiting the convolutional nature of our model also allows us to segment a complete cerebral image in **one minute**. In experiments on the 2013 and 2015 BRATS challenge dataset; we exhibit that our approach is among the most powerful methods in the literature, while also being very effective.

**Keywords.** Brain tumor, segmentation, deep learning, convolutional neural networks.

## 1 Introduction

The goal of brain tumor segmentation is to detect the area of the brain based on texture from information in MRI images. Segmentation methods typically look for active tumor tissue (vascularized or not), necrotic tissue and edema (swelling near a tumor) by exploiting multiple magnetic resonance imaging (MRI) modalities, such as T1, T2, T1-Contrasted (T1C) and Flair. Recently, Convolutional neural networks (CNNs) [25] are a type of deep artificial neural networks widely

used in the field of computer vision. They have been applied to many tasks, including image classification,[25, 19, 27, 15, 12] super-resolution [13] and semantic segmentation [26]. Recent publications report their usage in medical image segmentation and classification [14, 18, 33, 3, 34, 21, 4, 10, 9, 12].

For instance, Kamnitsas et al. [10] introduce a 3D CNN architecture designed for various segmentation tasks involving MR images of brains. The authors benchmark their approach on the BRATS [17] and ISLES [16] challenges. Their approach comprises a CNN with 3D filters and a conditional random field smoothing the output of the CNN. The authors propose dividing the input images into regions in order to address the high memory demand of 3D CNNs.

Notable in Kamnitsas is the usage of an architecture consisting of two pathways. The first receives the subregion of the original image that is to be segmented, while the second receives a larger region that is downsampled to a lower resolution before being fed to the network.

This enables the network to still be able to learn global features of the images. Havaei et al. [9] train and test MRI images of brains from the BRATS and ISLES data sets. The authors process the images slice-by-slice using 2D convolutions.

In addition, Havaei et al. [9] use the second part of their two-path architecture to fulfill the functionality of a Conditional Random Fields (CRF)

by feeding the segmentation output of the first path to the second one. Like Kamnitsas, the first CNN receives a larger portion of the original image than the second, with the purpose of learning both global context and local detail.

To conclude, the variety of CNN-based medical image segmentation methods is largely due to different attempts at addressing difficulties specific to medical images.

These are chiefly the memory demands of storing a high number of 3D feature maps, the scarcity of available data and the high imbalance of classes.

In dealing with the first issue, most researchers have turned to divide images into a small number of regions and stitching together the outputs of different regions [3, 34, 10] and/or using downsampled images [4].

Data augmentation is often used to address the scarcity of data [3, 34, 21, 4, 10, 7]. As for class imbalance, reported methods include using weighted loss functions [34, 4] overlap metrics such as the dice similarity [21, 7] or deep supervision [10, 6].

Recent research has shown that deep learning methods have performed well on supervised machine learning and image segmentation tasks [31, 32, 30]. The purpose of this study is to apply deep learning methods to segment brain tumor.

In this paper, we propose a successful and very efficient CNN architecture for brain tumor segmentation. Two main contributions were presented: Combining multiple segmentation maps created at different scales and using element-wise summation to forward feature maps from two stages of the network.

The remainder of this paper is structured as follows. We present our proposed methodology in section 2. Section 3 is devoted to experimental setup. In Section 4, we present the results achieved and compare with other existing approaches. Conclusions are finally drawn in Section 5.

## 2 Materials and Methods

The proposed methodology, shown in Figure 1, is applied on multimodal MRI sequences and exploits the inherent pattern recognition capability of CNN to classify tumor pixels. A patch based approach is used for pixel classification, where pre-processed images are passed through a CNN and post-processed to obtain a segmented image highlighting the tumor area.

The architecture proposed in Figure 2, takes as input patches of multiple modalities and predicts the class of center pixel in respective patches. The BRATS dataset [17] lacks resolution in third dimension; so, to extract 2D patches, axial view is used.

In the first convolution layer, input is the extracted patches from the original MR images (size  $128 \times 128$ ) that correspond to various anchors used. The produced feature maps are then taken as input by the cascading layers. A network of six convolution layers is implemented to learn feature maps with various kernel sizes. Rectified linear units (RLUs) activation is used for non-linear representation since it gives a better representation.

To reduce the input dimensionality that is going into the next layers, three max-pooling layers are used. The max-pooling layer selects the max value and discards the rest therefore summarizing the data in a small rectangle. In this way, the irrelevant information is discarded and the next convolution layer only receives the summarized important data. Pooling layers have more beneficial effects like invariance to lightning conditions and position.

To further reduce the complexity, a maxout layer is used, which reduces the number of feature maps by reducing the dimensions in third axis. It is used after convolution layer and selects two adjacent feature maps at maximum; therefore the number of maps produced by convolution layer is reduced to half. This resulted in a small improvement in performance.

The two-pathway convolution layer architecture produces the input for the fusion step of the network. The concatenated input is then fed to the second part of the network and the output layer

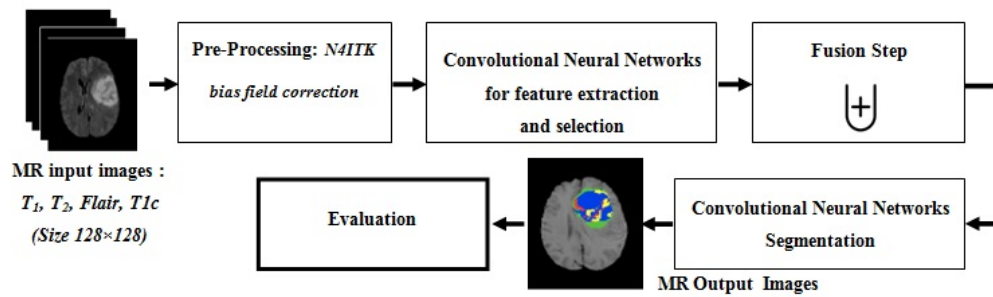


Fig. 1. Generic flow diagram of our proposed method

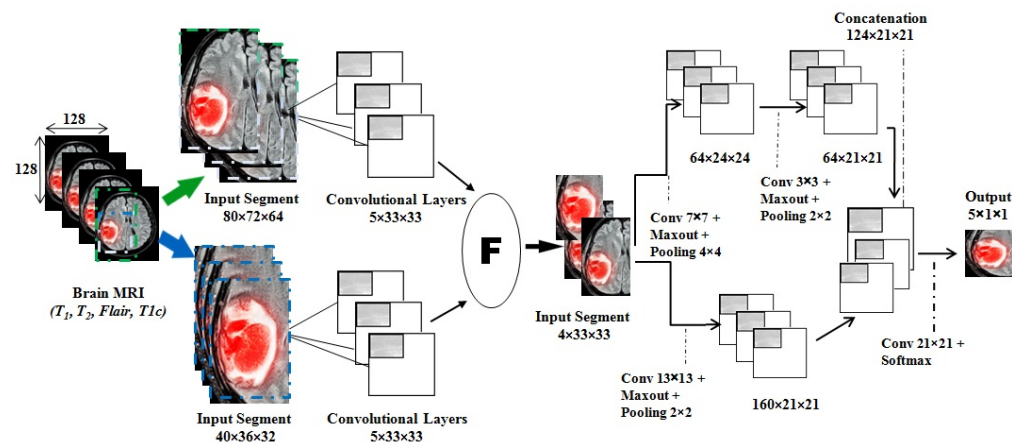


Fig. 2. Architecture 3D CNN for Brain segmentation with two pathways

i.e. softmax activation predicts class probabilities, which are accounted for in the loss function.

These steps are discussed in detail in the following subsections.

## 2.1 Preprocessing

The MR images when extracted from volumetric data have artifacts due to different acquisition techniques and systems [5]. Especially in T1 and T1c modality, the same type of tissues has different intensities across the dataset.

N4ITK bias field correction is applied using 3D slicer toolkit [28, 8], to T1 and T1c modalities. Image normalization is performed to ensure zero mean and unit variance.

Finally, patches are normalized with respect to mean and variance. Since fusion architecture is used for the neural network, two types of patches

are extracted: one having  $80 \times 72 \times 64$  pixels and the other having  $40 \times 36 \times 32$  pixels co-centric with the fusion step.

The choice of features maps and batch size is made empirically and it's limited by the quality of materials namely graphic memory, the processor and central memory. These constraints have a bad impact on the execution time as the run time can last several days without convergence.

## 2.2 Convolutional Neural Network (CNN) for Feature Extraction and Selection

CNN has an advantage over other classifiers as kernels used in convolution layers have same weights for all inputs, which detect same characteristics that makes them invariant by translation [24]. Usually, a non-linear activation

function is used to convert features into class probabilities.

Although inherently classifiers, CNNs can address segmentation tasks by throwing them to voxel-wise classification. The network processes a 3D patch around each voxel of an image. It is trained to predict whether the central voxel is pathology or normal brain tissue, depending on the content of the surrounding 3D patch.

During training, kernel parameters are optimized using gradient descent, with the goal of minimizing the error between predictions and true labels. One of the limitations in the above framework is that the segmentation of each voxel is done only by processing the contents of a small patch around it. It is intuitive that the context is more likely to lead to better results. However, a straight-forward increase in the size of the 3D input patch would increase the memory requirement and computational burden.

Our proposed solution is to perform a parallel image processing at multiple scales. Our network architecture consists of two parallel convolutional pathways, where both have receptive fields of the same size. The entrance to the second path, however, is a patch extracted from a subsampled version of an image, thus allowing it to a larger area around each voxel. This architectural design is shown in Figure 2.

Another significant feature of our architecture is its full convolutional nature, which allows its effective application on larger parts of the image. By supplying as input segments of an image larger than the receiver field of the neurons of the final layer, the network can efficiently process the larger input and provide as output predictions for several neighboring voxels. As a result of [7, 6] we also use this feature during training, building our training batches by extracting image segments larger in size than the network's receptive field.

### 2.3 Fusion Step

An earlier version of our system is shown in Figure 2. Our method uses a two-pathway architecture, in which each pathway is responsible for learning about either the local details or the broader context of tissue appearances (for example, whether or not he is close to the skull).

The tracks are connected by their concatenate feature cards immediately before the output layer. The fusion step used the result of the two pathway convolution layer. The result of step fusion is a patch size  $4 \times 33 \times 33$  features. This patch is an input of the two pathway of the final step.

### 2.4 Convolutional Neural Network (CNN) for Segmentation and Classification

Finally, a prediction of the class label is made by stacking a final output layer, which is totally convolutional to the last convolutional hidden layer. The number of feature cards in this layer corresponds to the number of class tags and uses the so-called non-linearity softmax.

3D CNN's perform pixel classification without taking into account the local dependencies of labels, one can model the dependencies of labels by taking into account estimates of wise pixel probability of a first CNN as an additional entry to a second 3D CNN, in formation of a new cascade architecture. Our final network is composed of two parallel pathway. The first pathway uses two layers with 73 and 33 kernels. The second pathway uses one layer with 133 kernels.

The final layer deep network exhibits significantly more accurate segmentation performance. Initial learning rate is set at 0.01 and is gradually reduced during training, and also constant impulse equal to 0.8. The training time requires for final system convergence is about one day using an NVIDIA Tesla K10 GPU with 8 GB of memory. Segmentation 3D brain tumor with four modalities requires 16 seconds.

## 3 Experimental Setup

### 3.1 Dataset

For testing and evaluating of our proposed system, we use the main objective of the annual BRATS challenge: segmenting tumor regions in brain MRI. More concretely, the network is trained using the BRATS 2013 and 2015 training set [1, 2]. They contain four modalities i.e., T1, T1-Contrasted (T1C), T2 and Flair. BRATS 2013 comprises of 30 training images (20 brains with high-grade (HG)

and 10 brains with low-grade (LG) tumors) and 10 brains with high-grade tumors for testing.

The data are rather sparse and preprocessing steps like skull stripping have been performed to improve the data representation. In 2013 dataset, two more data subsets are provided i.e. leaderboard and challenge data. These two subsets comprise of 65 MR images. Manual segmentation is available for training data only.

BRATS 2015 contains a total of 274 images, 220 were classified as high-grade gliomas (HG), while the remaining 54 were classified as low-grade gliomas (LG), with no inclusion of images depicting healthy brains. The classes involved in the segmentation task are: (1) necrosis, (2) edema, (3) non-enhancing and (4) enhancing tumor. Several examples are depicted in Figure 3.

All of the images have a size of  $240 \times 240 \times 155$ , which can be cropped to a region of the size  $160 \times 144 \times 128$ , while still containing the entire brain. For some of the experiments, these cropped images were further down sampled to a size of  $80 \times 72 \times 64$  for training and test.

### 3.2 Implementation Details

The algorithm is implemented in Theano with CUDA/GPU and CuDNN acceleration library in python. Hyper-parameters are tuned using grid search and the parameters on which model performed best on validation data are selected. Parameters such as learning rate and momentum are varied during training. Momentum is initially set to 0.6 and is gradually increased to 0.8. Learning rate, on the other hand, is initially set to 0.01 and then gradually decreased to  $0.1 \times 10^{-3}$ . Hence, a dropout value of 0.5 is used in the network to avoid over-fitting.

Segmenting brain tumor is an unbalanced classification problem where, most of the pixels are of healthy tissues.

Experiments have been performed on BRATS 2013 dataset that has two types of tumors, HG and LG glioma, divided into four tumor classes. There are 30 volumetric images in 2013 dataset containing slices varying in the range of 150 to 220. The dataset is divided randomly into training and testing sets with 80:20 ratios. The

dataset also contains synthetic data with low variance in intensity values of a similar class that is comparatively easy to classify. Therefore, only real patient data are used for evaluating the model. Evaluation metrics are determined for three tumor regions namely a) the complete tumor area (all four tumor labels), b) the core tumor area, and c) the enhancing tumor region.

The following describes the general training set-up used throughout the experiments: Of the 274 MRI volumes in the BRATS 2015 data set, 220 are used for training, while the remaining 54 are reserved for the validation and test sets, with 27 images each. We use the cross-validation approach to generate our training model. The Adam optimizer is used to optimize the network settings. Training takes place on an NVIDIA Tesla k10 for 200 epochs (around 120 hours). The trained network takes roughly 4 seconds to segment one  $80 \times 72 \times 64$  sized image.

### 3.3 Evaluation Parameters

The experimental results are evaluated based on one metric, namely dice similarity coefficient (DSC). Dice score is calculated by overlapping predicted labels with actual labels and the intersection of two contributors determine the dice score. Dice score is calculated for three categories i.e. the whole tumor, enhancing tumor, and core tumor and is given by (1):

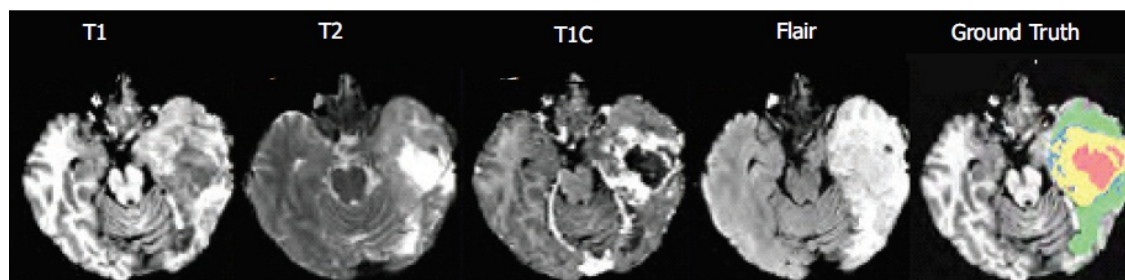
$$DSC = \frac{2 \times (|L \cap P|)}{|L| + |P|}, \quad (1)$$

where L and P stand for actual labels for tumor region and predicted tumor regions respectively.

## 4 Results and Discussion

A comparative analysis is presented in Table 1 to evaluate the effectiveness of the proposed model. In fact, Table 1 details the results obtained in our proposed model. It was found that, when each modality is used on its own, T1C produces the best results on every class but edema, for which Flair, followed by T2, produces better predictions.

Using this as a starting point, the network was then trained on Flair and T1C images together,



**Fig. 3.** The four images on the left show the modalities of the MRIs used as input channels for the CNN models and the one on the right displays the truth labels on the ground, with the following color code: edema (yellow), enhanced tumor (orange), necrosis (green), non-enhanced tumor (red) [9]

which lead to dice scores very close to the ones achieved by a network that had access to all four modalities as input channels. Since T2 is the next best-performing modality, the network was then trained on images in the Flair, T1C and T2 modalities.

Comparing the results of this run with the performance of the network when all modalities are available shows that, interestingly, the network achieves a much higher dice score on necrotic regions when the T1 modality is discarded.

This suggests that some benefit could be achieved by training different networks on different combinations of MRI modalities.

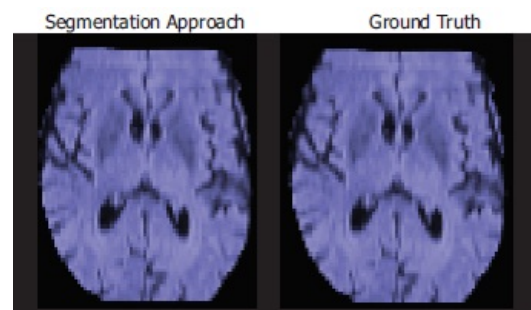
**Table 1.** Segmentation results on BRATS 2015 training data, where Class 1: Necrosis, Class 2: Edema, Class 3: Enhancing, Class 4: Non-enhancing

| Type  | DSC         |             |             |             |
|-------|-------------|-------------|-------------|-------------|
|       | Class 1     | Class 2     | Class 3     | Class 4     |
| Flair | 0.21        | <b>0.82</b> | 0.31        | 0.39        |
| T1    | 0.35        | 0.53        | 0.26        | 0.48        |
| T1C   | <b>0.56</b> | 0.67        | <b>0.52</b> | <b>0.83</b> |
| T2    | 0.38        | 0.71        | 0.31        | 0.46        |

The Table 2 and 3 show how our proposed model compares to the currently published state-of-the-art methods from BRATS 2013 and BRATS 2015 respectively.

It is evident from these Tables that our implemented architecture outperforms state-of-the-art methods [9, 29, 22, 20, 23] in terms

of dice similarity coefficient (DSC). Some of the segmentation results generated using the trained neural networks are shown in Figure 4 and 5.



**Fig. 4.** Model outputs for brain MRI, depicted alongside the ground truth. Colors correspond to: necrosis (green), non-enhanced (red) and enhanced tumor (orange) and edema (yellow)

Figure 4 shows visual segmentation produced by our model from BRATS 2013 and BRATS 2015 respectively. The larger receiver field in the two-path process allows the model to have more contextual information about the tumor and thus provides better segmentations.

In addition, with its two pathways, the model is flexible enough to recognize the fine details of the tumor rather than making a very smooth segmentation as in a one trajectory process.

By allowing a second phase of training and learning by the true class distribution, the model corrects most of the classification errors produced in the first phase.

**Table 2.** Segmentation results on BRATS 2013 training data compared with the state-of-the-art methods, where Class 1: Complete tumor area, Class 2: Core tumor area, Class 3: Enhancing tumor area

| Method               | DSC          |              |              |
|----------------------|--------------|--------------|--------------|
|                      | Class 1      | Class 2      | Class 3      |
| <b>Proposed</b>      | <b>0.92</b>  | <b>0.83</b>  | <b>0.86</b>  |
| Havaei et al. [9]    | 0.88         | 0.79         | 0.73         |
| Tustison et al. [29] | 0.87         | 0.78         | 0.74         |
| Pereira et al. [22]  | 0.88         | 0.83         | 0.77         |
| Dvorak et al. [20]   | 0.72         | 0.66         | 0.67         |
| Rao et al. [23]      | Not reported | Not reported | Not reported |

**Table 3.** Segmentation results on BRATS 2015 training data compared with the state-of-the-art method, where Class 1: Necrosis, Class 2: Edema, Class 3: Enhancing, Class 4: Non-enhancing

| Method            | DSC         |             |             |            |
|-------------------|-------------|-------------|-------------|------------|
|                   | Class 1     | Class 2     | Class 3     | Class 4    |
| <b>Proposed</b>   | <b>0.52</b> | <b>0.88</b> | <b>0.68</b> | <b>0.9</b> |
| Baris et al. [11] | 0.49        | 0.84        | 0.5         | 0.8        |

Based on experiments results show in Figure 5, we conclude that our system can easily detect the normal brain without tumor compared with ground truth image. The dice similarity coefficient (DSC) value is 0.98 similar in the ground truth image.

The proposed algorithm performs well in specifying tumor region as is evident from the lack of false positives in detections. It also detects enhancing tumor better than most state-of-the-art techniques and gives comparable results on other metrics.

It is observed that the trained model faces difficulty, predicting minority classes. But by increasing training data, this problem can be dealt with. Two variations of the proposed architecture are also tested where a number of features in the fully connected layer is varied. It has been observed that too many features in the fully connected layers lead to over-fitting and if features are reduced enormously, the model does not learn significantly leading to under fitting. It is also observed that fully connected layers are time-consuming compared to convolutions and

thus a trade-off between segmentation time and accuracy is achieved in the fully connected layer.

## 5 Conclusion and Future Work

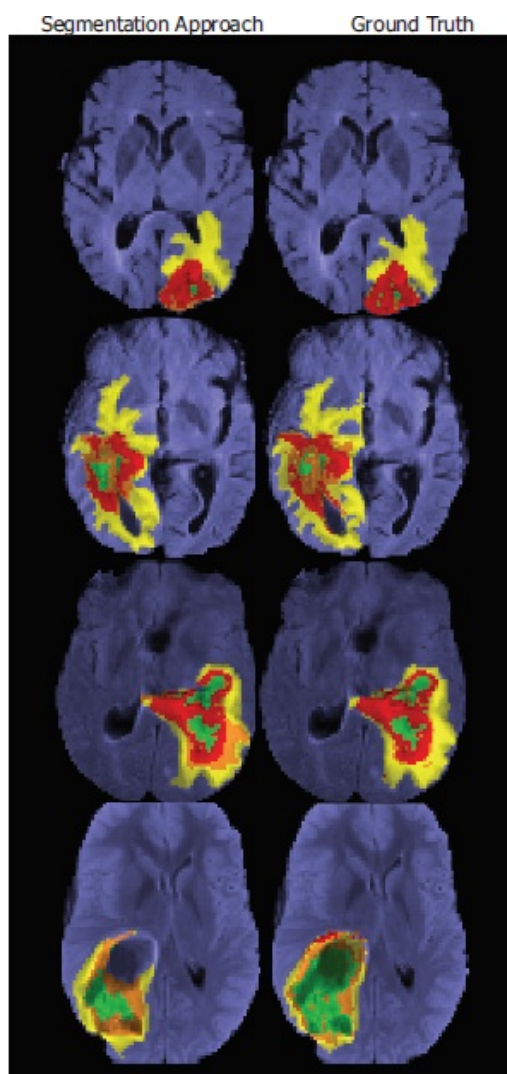
Brain tumor segmentation has a very important role in diagnostic procedures. With accurate segmentation, clinical diagnostic not only becomes easy, but also the chances of subjects survival increase tremendously.

In this paper, a 3D CNN architecture for brain tumor segmentation is presented. This algorithm incorporates both global and local features since context is important when it comes to tumor segmentation task. The use of max-pooling, max-out, and drop-out complement the learning process, improving training and testing speed by reducing features in the fully connected layer as well as reducing a number of parameters, which in turn reduce the chances of over-fitting.

Evaluation results show that the proposed network architecture is promising and performs particularly well in detecting an enhancing tumor as well as specifying tumor to actual tumor region only.

Our architecture illustrates promising performance, with capabilities for delicate segmentations. The difficulties are observed in the segmentation of the lesions of particularly low size.

The separation of the lesions in various categories, for example according to their size and their treatment by different classifiers could simplify the task for every learner and help to limit the problem.



**Fig. 5.** Detection of normal Brain, depicted alongside the ground truth

## Acknowledgment

The authors would like to thank Prof. Xin Yao<sup>1</sup> for his validation of our architecture for Brain segmentation with two pathways and all his pertinent remarks during *Internet of Things (IoT) Workshop*<sup>2</sup>.

<sup>1</sup><http://www.cs.bham.ac.uk/xin/>

<sup>2</sup><http://devfest.ieee.tn/>

## References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., & Davatzikos, C. (2017). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive*.
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., & Davatzikos, C. (2017). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Archive*.
3. Chen, H., Dou, Q., Lequan, Y., Qin, J., & Pheng, A. H. (2017). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, Vol. 161, pp. 135–146.
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *CoRR*, Vol. abs/1606.06650, pp. 1–8.
5. Collewet, G., Strzelecki, M., & Mariette, F. (2004). Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic Resonance Imaging*, Vol. 22, pp. 81–91.
6. Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., & Heng, P. (2016). 3D deeply supervised network for automatic liver segmentation from CT volumes. *CoRR*, Vol. abs/1607.00582, pp. 1–8.
7. Drozdal, M., Vorontsov, E., Chartrand, G., Cadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. *CoRR*, Vol. abs/1608.04117, pp. 1–9.
8. Fedorov, A., Beichel, R., Finet, J. K.-C. J., Fillion-Robin, J.-C., Pujol, S., & et al (2012). 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, Vol. 30, pp. 1323–1341.
9. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P., & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, Vol. 35, pp. 18–31.
10. Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3d CNN with fully connected CRF for



- accurate brain lesion segmentation. *Medical Image Analysis*, Vol. 36, pp. 61–78.
11. **Kayalibay, B., Jensen, G., & van der Smagt, P. (2017).** CNN-based segmentation of medical imaging data. *CoRR*, Vol. abs/1701.03056, pp. 1–24.
  12. **Kharrat, A. & Neji, M. (2018).** Classification of brain tumors using personalized deep belief networks on MRImages: PDBN-MRI. *11th International Conference on Machine Vision (ICMV 2018)*, pp. 1–9.
  13. **Kim, J., Lee, J. K., & Lee, K. M. (2016).** Accurate image super-resolution using very deep convolutional networks. *CoRR*, Vol. abs/1511.04587, pp. 1–9.
  14. **Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017).** Residual and plain convolutional neural networks for 3d brain mri classification. *CoRR*, Vol. abs/1701.06643, pp. 1–4.
  15. **Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).** Imagenet classification with deep convolutional neural networks. *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, pp. 1097–1105.
  16. **Maier, O., Menze, B. H., von der Gabelentz, J., & et al (2017).** ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis*, Vol. 35, pp. 250–269.
  17. **Menze, B., Jakab, A., Bauer, S., & et al (2015).** The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, Vol. 34, pp. 1993–2024.
  18. **Novikov, A. A., Major, D., Lenis, D., Hladuvka, J., Wimmer, M., & Bühler, K. (2017).** Fully convolutional architectures for multi-class segmentation in chest radiographs. *CoRR*, Vol. abs/1701.08816, pp. 1–9.
  19. **Paul, J. S., Plassard, A. J., Landman, B. A., & Fabbri, D. (2017).** Deep learning for brain tumor classification. *Proc. SPIE 10137, Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, SPIE., pp. 129–138.
  20. **Pavel, D. & Bjoern, M. (2015).** Structured prediction with convolutional neural networks for multimodal brain tumor segmentation. *MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS)*, IEEE, pp. 13–24.
  21. **Pei, Y., Qin, H., Ma, G., Guo, Y., Chen, G., Xu, T., & Zha, H. (2017).** Multi-scale volumetric ConvNet with nested residual connections for segmentation of anterior cranial base. *International Workshop on Machine Learning in Medical Imaging*, Springer, Cham, pp. 123–131.
  22. **Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016).** Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans Med Imaging*, Vol. 35, pp. 1240–1251.
  23. **Rao, V., Sarabi, M. S., & Jaiswal, A. (2015).** Brain tumor segmentation with deep learning. *MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS)*, IEEE, pp. 56–59.
  24. **Sainath, T. N., r. Mohamed, A., Kingsbury, B., & Ramabhadran, B. (2013).** Deep convolutional neural networks for LVCSR. *Acoustics, speech and signal processing (ICASSP), 2013 IEEE international conference*, IEEE, pp. 8614–8618.
  25. **Schmidhuber, J. (2015).** Deep learning in neural networks: An overview. *Neural Networks*, Vol. 61, pp. 85–117. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].
  26. **Shelhamer, E., Long, J., & Darrell, T. (2017).** Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 39, No. 4, pp. 640–651.
  27. **Simonyan, K. & Zisserman, A. (2016).** Very deep convolutional networks for large-scale image recognition. *CoRR*, Vol. abs/1409.1556, pp. 1–14.
  28. **Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., A.Yushkevich, P., & et al (2010).** N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, Vol. 29, pp. 1310–1320.
  29. **Tustison, N. J., Shrinidhi, K. L., Wintermark, M., Durst, C. R., Kandel, B. M., GeeMurray, J. C., Grossman, C., & Avants, B. B. (2015).** Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation. *Neuroinformatics*, Vol. 13, pp. 209–225.
  30. **Wang, Y., Widrow, B., Zadeh, L. A., Howard, N., Wood, S., Bhavsar, V. C., & Shell, D. F. (2016).** Cognitive intelligence: Deep learning, thinking, and reasoning by brain-inspired systems. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, Vol. 10, pp. 1–20.

- 31. Wang, Y., Zadeh, L. A., Widrow, B., Howard, N., Beaufays, F., Baciú, G., & Raskin, V. (2017).** Abstract intelligence: Embodying and enabling cognitive systems by mathematical engineering. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, Vol. 11, pp. 1–15.
- 32. Wang, Y. & Zatarain, O. A. (2017).** A novel machine learning algorithm for cognitive concept elicitation by cognitive robots. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, Vol. 11, pp. 31–46.
- 33. W.Gao, X., Hui, R., & Tian, Z. (2017).** Classification of CT brain images based on deep learning networks. *Computer Methods and Programs in Biomedicine*, Vol. 138, pp. 49–56. Elsevier.
- 34. Zhen, X., Chen, J., Zhong, Z., Hrycushko, B., Zhou, L., Jiang, S., Albuquerque, K., & Gu, X. (2017).** Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study. *Physics in Medicine & Biology*, Vol. 62, No. 21, pp. 8246–8263.

Article received on 12/11/2018; accepted on 06/02/2020.  
Corresponding author is Ahmed Kharrat.