Received: 2 October 2017 | Revised: 14 December 2017 | Accepted: 3 January 2018

DOI: 10.1002/jemt.22994

RESEARCH ARTICLE

WILEY MICROSCOPY RESEARCH & TECHNIQUE

Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)

Sajid Iqbal[1,2] | M. Usman Ghani[1] | Tanzila Saba[3] | Amjad Rehman[4]

[1]Department of Computer Science and Engineering, University of Engineering and Technology, Lahore, Pakistan

[2]Department of Computer Science Bahauddin Zakariya University Multan Pakistan

[3]College of Computer and Information Sciences, Prince Sultan University, Riyadh, 11586, Saudi Arabia

[4]College of Computer and Information Systems, Al Yamamah University, Riyadh, 11512, Saudi Arabia

Correspondence
Amjad Rehman, College of Computer and Information Systems, Al Yamamah University, Riyadh, 11512, Saudi Arabia.
Email: rkamjad@gmail.com

Funding information
Machine Learning Research Group; Prince Sultan University Riyadh; Saudi Arabia, Grant/Award Number: RG-CCIS-2017-06-02

Review Editor: Dr. Peter Saggau

Abstract

A tumor could be found in any area of the brain and could be of any size, shape, and contrast. There may exist multiple tumors of different types in a human brain at the same time. Accurate tumor area segmentation is considered primary step for treatment of brain tumors. Deep Learning is a set of promising techniques that could provide better results as compared to nondeep learning techniques for segmenting timorous part inside a brain. This article presents a deep convolutional neural network (CNN) to segment brain tumors in MRIs. The proposed network uses BRATS segmentation challenge dataset which is composed of images obtained through four different modalities. Accordingly, we present an extended version of existing network to solve segmentation problem. The network architecture consists of multiple neural network layers connected in sequential order with the feeding of Convolutional feature maps at the peer level. Experimental results on BRATS 2015 benchmark data thus show the usability of the proposed approach and its superiority over the other approaches in this area of research.

KEYWORDS

BRATS datasets, convolutional neural networks, deep learning, features mining, tumor segmentation

1 | INTRODUCTION

Artificial neural networks (ANN) and convolutional neural networks (CNN) have shown their usability in various tasks of computer vision such as object recognition, human activity identification, face based person identification, and plethora of applications in the domain of medical imaging (Abbas et al., 2016; Saba, 2017). Various works published in recent years have proved that CNN and other deep learning based approaches are at the forefront of medical image segmentation and analysis related tasks (Husham, Hazim Alkawaz, Saba, Rehman, & Saleh Alghamdi, 2016; Saba, Rehman, & Sulong, 2011). Iterative gradient based optimization is main reason for popularity of deep learning approaches that provide an automatic way of optimal feature extraction used for segmentation and classification tasks. It relieves the researcher from manual feature engineering which is hard and error prone process where type and quality of extracted features mainly depends upon human knowledge and expertise. Manually engineered features are questioned to relevance to the task and need to go through feature selection and feature reduction to avoid over-fitting. It is very easy to come up with nonrelevant or semi-relevant features due to lack of knowledge or domain expertise or designing the features that cause model over-fitting (Brownlee, 2014).

In conventional machine learning approaches a pipeline of processes is adapted to complete the segmentation and classification task. This pipeline consists of required and optional steps that include preprocessing, feature extraction, feature reduction, segmentation, and classification. Brain tumor segmentation is one of the main areas in medical image processing where different neural network based approaches have been successfully applied (Norouzi et al., 2014; Rad, Rahim, Rehman, Altameem, & Saba, 2013; Saba, Rehman, & Elarbi-Boudihir, 2014). Medical images are produced using different type of imaging technologies called modalities that may include CT, MRI, PET, and their different variants (Husham et al., 2016; Iftikhar, Fatima, Rehman, Almazyad, & Saba, 2017; Jamal, Alkawaz, Rehman, & Saba, 2017) that are broadly categorized as invasive and noninvasive imaging methods.

Medical imaging research community is applying CNN for segmentation and classification to extract diseases and interest areas in medical images for different body parts like tissue (Mandal, Chatterjee, & Maitra,

2017; Xue, Xu, Zhang, Long, & Huang, 2017), skin (Chaichulee et al., 2017), bone (Gan, Xia, Xiong, Li, & Zhao, 2017), and their different level of visualizations. Images obtained through noninvasive methods like MRI, are considered better and used commonly as the imaging process does not cause a harmful effect on body tissues. In case of brain, where small damage to any part of tissue can result in an unrecoverable loss, such modalities are highly preferred (Mughal, Muhammad, Sharif, Saba, & Rehman, 2017). Images produced through different imaging modalities could enhance a specific set of features (Muhsin, Rehman, Altameem, Saba, & Uddin, 2014) and medical research community, normally, uses images produced through different modalities in complementary way to extract maximum and highly useful features (Rahim, Norouzi, Rehman, & Saba, 2017a; Rahim, Rehman, Kurniawan, & Saba, 2017b). Gliomas are the most common type of tumors found in human brain. These are grouped into two main categories based on their level of severity: low-grade glioma (LGG) and high-grade glioma (HGG). LGG is considered as benign whereas HGG is malignant.

In this work, proposed models are applied on the dataset provided by Swiss Institute for Computer Assisted Surgery (SICS) that manages brain tumor segmentation challenge yearly and a significant number of high performing algorithms have been published through this platform. These algorithms are based on spatial and volumetric segmentation of human brain scans mainly using deep learning algorithms (Pereira, Oliveira, Alves, & Silva, 2017; Zhao et al., 2018) and conventional machine learning approaches (Mughal et al., 2017). Autoencoders, fully connected neural networks and inception based methods are popular among researchers and their different variants are being explored (Clark, Wong, & Haider, 2017; Szegedy, Ioffe, Vanhoucke, & Alemi, 2017).

Image segmentation in medical images includes two phases; first, detection of unhealthy tissue and second, delineation of different anatomical structures or areas of interest. Recently published work shows that neural network based approaches are winning algorithms (Saba, Al-Zahrani, & Rehman, 2012). There are mainly two types of neural networks being used for this task that is pixel-wise classification (semantic segmentation) and patch wise classification. In patch wise classification process, small spatial or volumetric patches from an image are taken and fed into a network which categorizes the patch to one of the target classes (Saba et al., 2014). In the second approach, known as semantic segmentation, the whole input image is given to network and network provides a label for each pixel (Saba et al., 2011).

Neural network model presented by Badrinarayanan, Kendall, and Cipolla (2015) for scene parsing of colored images is taken as baseline and extended in three different ways. We extend this network in three ways. The network consists of five blocks of an encoder which extracts features and down samples the input images whereas in decoder phase, the same number of blocks are used to upsample input to original size. Rest of the article is organized as follows: Section 2 review the related work, Section 2.1 presents the proposed methodology and implementation details. Section 3 exhibits its experimental results, and in section 3.1 research is concluded along with future work discussion.

2 | RESEARCH BACKGROUND

Brain tumor segmentation techniques are categorized as manual, semi and fully automated approaches. Neural network based classification and semantic segmentation based methods lie in the category of fully automated approaches (Iqbal, Khan, Saba, & Rehman, 2017). Recently performance of deep learning methods has attracted researchers to employ CNN for biological image segmentation. Additionally, for semantic segmentation, CNN and FCNN exhibited prominent segmentation accuracy (Havaei et al., 2017; Shi et al., 2016; Xue et al., 2017; Zhao et al., 2018). Most successful neural networks are based on encoder-decoder based structures. Among these U-net and V-net based networks are popular architectures (Dong et al., 2017). Lun and Hsu (2016) employed SegNet for brain tumor segmentation using BRATS 2015 dataset. Isensee et al. (2017) worked on receptive field size of the neural network to improve segmentation accuracy. To obtain better results, important points to consider include preprocessing, data augmentation, batch normalization, loss function optimization function selection and postprocessing. Dong et al. (2017) has shown the effectiveness of preprocessing and data augmentation in brain tumor segmentation process. Due to BRATS data format, few preprocessing operations are considered to be essential like pixel intensity range standardization. Zhao et al. (2018) have used conditional random fields on predictions achieved by the network to fine-tune their results. Hierarchical network-based segmentation approaches like multi-resolution based loss calculation also exhibited acceptable results (Xue et al., 2017). Besides automatic approaches, interactive approaches also produce good results and being explored (Wang, Wang, Zhang, Xiang, & Pan, 2017).

There are three main aspects of neural networks which are being explored by researchers: First, extraction of useful features for classification or segmentation, second, finding relative importance of extracted features and third, developing right loss function for underlying problem. For the first task, different network architectures are being designed and tested like VGG, Inception, DeepLab, and GoogleNet (Szegedy et al., 2017). To give right importance to extracted features, different weighting schemes have been explored. This includes the weighing of features in loss layer computed at different granularity level like weighted cross entropy or weighted categorical cross entropy. The weighting schemes include the dataset based feature weighting/class weighting using average, median, and maximum frequency based weights. Another approach is to calculate the features weights based on input batch similar to the proposed work. Wang et al. (2017) have used a gated Convolutional network where relative weights for each pixel are generated and applied in last Convolutional layer. Squeeze and Excitation block (employed in this work) provides another view to weight the computed feature maps in different phases of the network instead of applying them in the last Convolutional layer or loss layer. Defining or selecting right loss function is very important to boost up neural network performance. Normal loss functions like categorical cross entropy and least square methods provide good results for balanced to semi-balanced datasets however the datasets having large imbalance among the classes require special attention and specialized loss functions can perform better. Sudre, Li, Vercauteren, Ourselin, and

Cardoso (2017) have provided a comparative study of different loss functions for highly unbalanced segmentation task and they argue that generalized dice similarity coefficient (DSC) function is better than others.

Pereira, Pinto, Alves, and Silva (2016) employed CNN with small (3×3) filters for deeper architecture to segment brain tumor in MRI and claimed 0.88, 0.93, 0.74 segmentation accuracy for whole tumor, core tumor and active tumor respectively on BRATS dataset. Similarly, Havaei et al. (2017) used Cascaded Two-pathway CNNs for simultaneous local and global processing of brain tumor identification and segmentation. They achieved 0.88, 0.79, 0.73 segmentation accuracy for whole tumor, core tumor and active tumor respectively on BRATS dataset. Finally, Rao, Sarabi, and Jaiswal (2015) fused four CNNs, one for each modality, with their outputs concatenated as an input into a RF to segment the brain tumor in MRI. However, no results are reported.

2.1 | BRATS dataset

All imaging data used in this study is obtained from BRATS 2015 dataset which consists of two major types of brain tumors called HGG and LGG. The dataset consists of images provided by BRATS 2012 and BRATS 2013 and extra cases taken from Cancer Imaging Archive (TCIA). In training dataset, there are 220 cases of HGG and 54 cases of LGG where each case consists of images taken through four modalities T1, T1c, T2, and FLAIR along with the corresponding ground truth. The size of each 3D image is 155 × 240 × 240. Voxels of each image are divided among five classes using numeric labels that are 1 for enhancing core, 2 for the necrotic cystic region, 3-enhancing core, 4 for edema and 0 for everything else (background and noise if any). The test data set consists of 110 cases with mixed grades with no distinction between two types (Menze et al., 2015). Multimodal brain tumor segmentation challenge (BRATS) has provided data consisting of scans taken through four imaging modalities T1, T1-contrast, T2, and FLAIR. Each modality emphasizes different set of features which are used in combination, as they provide complementary information, to find the area and type of tumor-like tumor core, edema, and enhancing tumors.

3 | PROPOSED APPROACH

The proposed approach consists of three major steps: data preprocessing, network training, and predictions generation based on the test dataset.

3.1 | Preprocessing

Preprocessing is necessary as the images in the dataset have varying intensity ranges and are beyond standard intensity value of 255. To make the network training smooth and quantifiable, it is required to do a particular set of preprocessing operations. It has been shown that preprocessing helps in improving the segmentation accuracy (Dong et al., 2017). A series of preprocessing techniques is applied on the data prior to feed CNN for segmentation. Table 1 lists the preprocessing steps and their output is shown in Figure 1.

3.2 | Deep learning models

This work extends the neural network architecture proposed by Badrinarayanan et al. (2015) for contextual image segmentation. Three different network architectures named (1) Interpolated Network (2) SkipNet (3) SE-Net are proposed.

All three networks consist of decoder and encoder architecture where 4 sub-blocks are used in each phase. Combination of convolution, batch normalization and ReLU layers is used as basic unit in encoder. It is to mention that Convolutional gradients may easily explode or vanish after few training iterations. To keep them stable and training smooth, we added batch normalization layers after each convolution.

TABLE 1 Preprocessing techniques

Preprocessing	Description
Intensity Normalization	Voxel values of data set images lie beyond the 0–255 boundary and are in variable range. We used intensity rescaling to convert each image voxel into standard range i.e., 0–255. ITK toolkit Rescale Intensity Filter () method is used for this purpose.
Bias Correction	Usually, a smooth and low-frequency noise signal is added to the image by the imaging camera, N4ITK bias correction is used to remove this effect.
Mean subtraction	Mean across each modality is calculated and then subtracted from rescaled voxel values
2D slicing	Images provided by BRATS are in 3D mha format. We sliced them as 2D numpy arrays. So each 3D image generated 155 numpy arrays each of size 240 × 240.
Cropping	To make training process faster and feed a larger image batch, we calculated the largest-shortest boundary for all images in such a way that the only background is cropped and all foreground pixels remain intact. The calculated boundary is 184 × 184.
Data Augmentation	Different augmentation techniques are being used by researchers like flipping, introducing noise, image enhancement and rotation. However, many of such techniques change the pixel intensity values in original images and some of them also require to change the pixel values in corresponding ground truth like rotation at 20°. To keep the actual image pixel values and related ground truth values unchanged we only used noninvasive preprocessing methods. These are listed below: • Rotation at 90° (0,90,180,270) of original image. • Rotation at 90° (0,90,180,270) of left-right flipped image

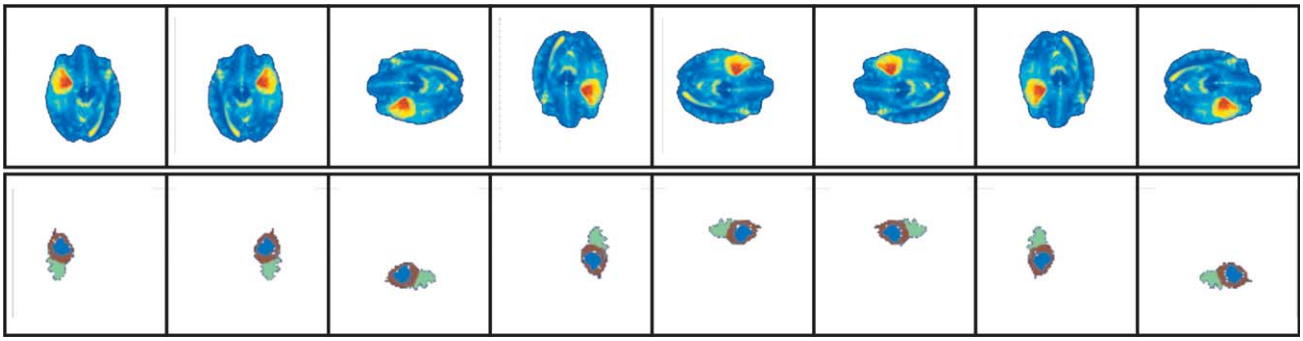


FIGURE 1 Top: (rotated and flipped images taken from T2 modality and Bottom) corresponding ground truths [Color figure can be viewed at wileyonlinelibrary.com]

3.3 | Training settings

A batch of 20 images of 4 channel 184 × 184 images is fed to the network as input. In the first two proposed networks, feature maps in encoder are reduced to half spatial size using max pooling and in decoder up-sampling is performed using Caffe deconvolution layer. Following, deconvolution, upsampled feature maps are concatenated with feature maps taken from corresponding encoder block. The loss function in all three networks is categorical cross entropy. Cross-entropy is the measure of the actual distribution using estimated distribution and is given equation 1. If p belongs to true distribution and q belongs to a predicted probability distribution. The measure of similarity between p and q is computed as per Equation 1:

$$H(p,q) = -\sum p_i \log(q_i) \tag{1}$$

The data provided by BRATS is heavily imbalanced and to overcome this issue we used batch wise class weighting strategy in calculating loss function.

$$\text{avg} = \frac{1}{5}\sum \text{classFrequency} \text{ and } \text{weigt}_i = \text{classFrequency}_i/\text{avg} \tag{2,3}$$

The computed loss of each class is multiplied by loss function with corresponding class weight as calculated above.

$$H(p,q) = -\sum w_i p_i \ \log(q_i) \tag{4}$$

Finally, the loss is calculated by taking the average over all samples as shown in equation 5.

$$L(w) = (1/N)\left(-\sum w_i p_i \ \log(q_i)\right) \tag{5}$$

The optimization is done by Stochastic gradient descent (SGD) where weights are updated using following formula.

$$w = (1-\lambda')w + \eta'(y-y')x \tag{6}$$

To avoid over-fitting, dropout layer and L2 regularization are employed.

3.3.1 | Model 1: SkipNet

Original Segnet uses five blocks in encoder and decoders. In this model (Figure 2), one inner most block from encoder and decoder is removed and skip connections are introduced from corresponding layers in encoder to decoder.

3.3.2 | Model 2: Interpolated network (IntNet)

Second model (Figure 3) extends the original VGG architecture in two ways: (1) by propagating Convolutional feature maps from encoder phase to corresponding Convolutional layers, as shown in Figure 4, in decoding phase along with propagation of pooling maps. In addition to that, after two level of down-sampling, intermediate prediction maps are taken and by applying binary linear interpolation (BLI), resulting in 184 × 184, interpolated feature maps are concatenated with input in first convolution layer of final block of the decoder. BLI is performed in both x and y-direction. Although a linear operation is performed in each direction, however, the combined effect is quadratic instead of linear. If value of four neighboring points $Q_{11} = (x_1, y_1),\ Q_{12} = (x_1, y_2),\ Q_{21} = (x_2, y_1),\ Q_{22} = (x_2, y_2)$ is known, the value of a point $P(x, y)$ could be interpolated using formula.

$$\begin{aligned} P \approx & \frac{(x_2-x)(y_2-y)}{(x_2-x_1)(y_2-y_1)}Q_{11} + \frac{(x-x_1)(y_2-y)}{(x_2-x_1)(y_2-y_1)}Q_{21} \\ & + \frac{(x_2-x)(y-y_1)}{(x_2-x_1)(y_2-y_1)}Q_{12} + \frac{(x-x_1)(y-y_1)}{(x_2-x_1)(y_2-y_1)}Q_{22} \end{aligned} \tag{7}$$

To avoid over-fitting, a dropout layer is added between encoder and decoder block. To further weaken the over-fitting L2 regularization is used.

3.3.3 | Model 3: SENet

In this model, two recent approaches are employed, first, sub-sample pixel based up-Quine, sampling and down-sampling; second, Squeeze-and-Excitation block extension of SkipNet.

Quine, Valery, Henok, and Richard (2007) introduced the idea of enhanced version of sub-pixel based interpolation which has been recently used in neural networks by Shi (2016) in their under publication work. Jie, Shen, and Sun (2017) have introduced a squeeze-and-excitation (SE) block in networks using inception blocks or residual blocks. The SE block is used at the end of decoder and its output is fused with the output of inner most encoder block. The fused feature maps are given as input to the first block of decoder as shown in Figure 4.

During upsampling process, convolution with a fractional stride of 1/r is used where r is the factor of upsampling. The process of upsampling is given by

$$I^{SR} = fL\left(I^{LR}\right) = PS\left(W_L * f^{L-1}\left(I^{LR}\right) + b_L\right) \tag{8}$$

Here, low resolutions (LR) image is converted to super-resolution (SR) image using periodic shuffling (PS) that rearranges the elements of

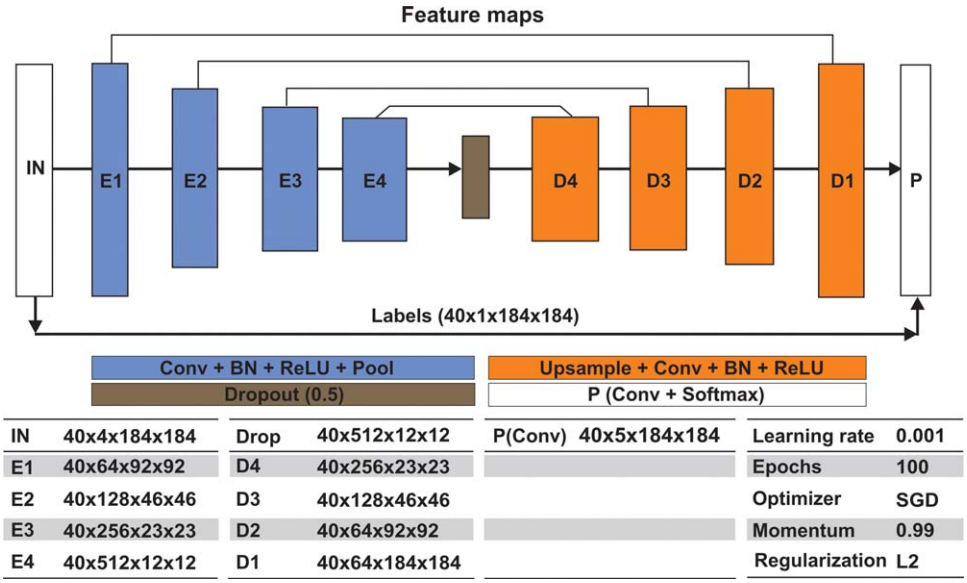


FIGURE 2 SkipNet - feature maps from encoder blocks are concatenated with inputs of corresponding decoder blocks [Color figure can be viewed at wileyonlinelibrary.com]

a tensor with shape $H \times W \times C$. r^2 to $rH \times rW \times C$. This makes the computation faster and smarter. The process of down sampling is just the reverse of it (Shi et al., 2016).

The squeeze and excitation block (SE block) attempts to ensure the network sensitivity for informative features which are exploited by subsequent transformations and to suppress the less important features. This is done by explicitly modeling channel interdependencies to calibrate filter responses. It is done in two phases known as squeeze and excitation on input obtained by the previous transformation. This recalibrated input is then fed to next transformation block. A Convolutional layer produces multiple feature maps that may contain interdependencies however the weight filters operating on each feature map work on local receptive fields which limits them to incorporate contextual information from a region outside the local receptive field. Global average pooling is used to generate channel wise statistics by converting 3D tensor to 1D tensor.

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i,j) \tag{9}$$

Here, U is the output taken from previous transformation block (a 2D feature map) and z_c is statistics calculated from it. The excitation function is given by the following equation:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \tag{10}$$

this function tries to capture channel wise dependencies. Here, δ refers to ReLU function and σ for nonlinearity. The final output of the block is generated by rescaling the transformation output with activations.

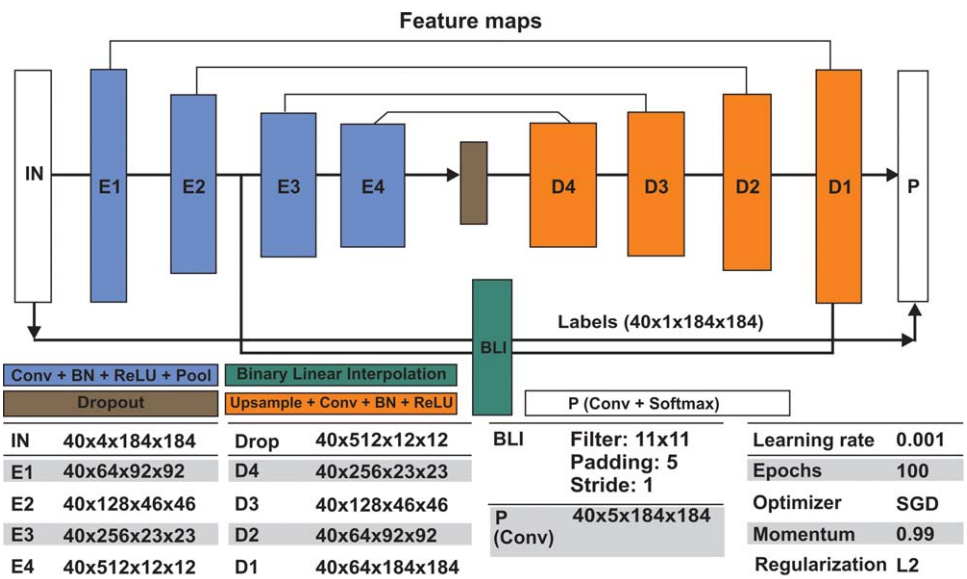
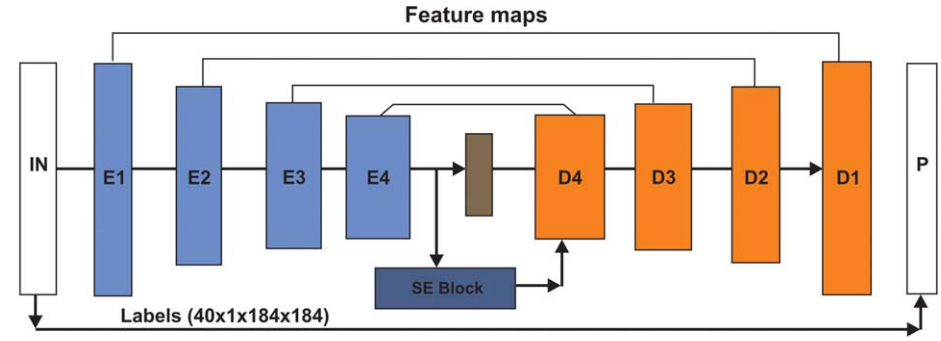


FIGURE 3 IntNet CNN architecture [Color figure can be viewed at wileyonlinelibrary.com]



IN	40x4x184x184	Drop	40x512x12x12	P(Conv)	40x5x184x184	Learning rate	0.001
E1	40x64x92x92	D4	40x256x23x23			Epochs	100
E2	40x128x46x46	D3	40x128x46x46			Optimizer	SGD
E3	40x256x23x23	D2	40x64x92x92			Momentum	0.99
E4	40x512x12x12	D1	40x64x184x184			Regularization	L2

FIGURE 4 (Top) CNN with SE block, (Bottom) SE block [Color figure can be viewed at wileyonlinelibrary.com]

$$\tilde{X}_c = F_{\text{scale}}(u_c, s_c) = s_c \cdot u_c \tag{11}$$

The scalar S_c is multiplied with previous transformation U_c. In proposed network, one block of SE after fourth pooling is employed as shown in the Figure 5. Proposed training algorithm is summarized as follow:

Algorithm 1: Training algorithm for proposed networks

Input: Training data x, maximum epochs E

Initialization: initialize the hyper parameters like network weights: msra, learning rate $\alpha = 0.01$, learning rate policy: step, momentum: 0.9, weight decay: 0.0005, optimizer: SGD, regularization type: L2

Output: learned weight model

Read 3D medical images

Rescale intensity to range 0–255

Converting 3D images to 2D slices

Preprocessing 2D images

while e < E:

e ← e +1

forward the training network to compute the loss L

backward the network to compute the gradients

update the network parameters and hyper parameters

if network loss is stabilized

pause training and reduce learning rate manually

resume training from where it is stopped

end while

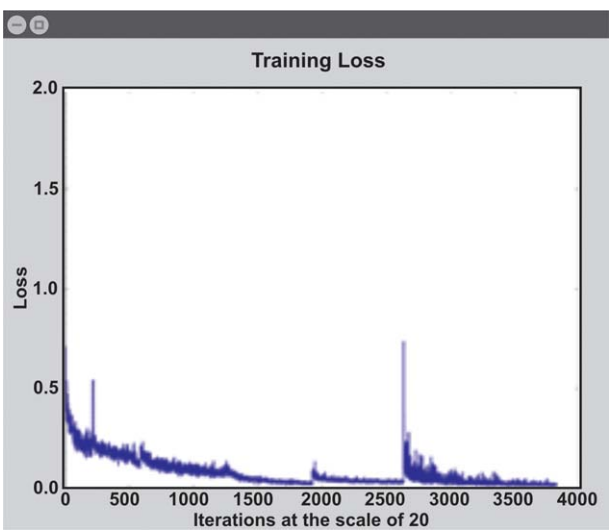


FIGURE 5 Training loss: The spikes show the change of learning rate [Color figure can be viewed at wileyonlinelibrary.com]

4 | EXPERIMENTAL RESULTS AND DISCUSSION

The proposed models are implemented in Caffe on Titan X (Pascal) GPU with 12 GB memory size. The training of each network took around 48 hr and average testing time takes 0.5 ms for one 2D image of size 240 × 240 and for one 3D image prediction time becomes 75 ms (includes 155 2D images read time).

We trained all three proposed models and did multiple experiments with each model using different hyper parameter settings. First, training on LGG data only, then training on HGG data only and finally, training on combined (HGG and LGG) data. We did fivefold cross-validation and report mean scores. Figure 5 shows the training loss in an experiment.

The experimental setup is given in Table 2:

TABLE 2 Experimental setup detail

Experiment	Train Set Size	Test Set Size
HGG	200	20
LGG	40	14
HGG_LGG	200 + 40 = 240	20 + 14 = 34

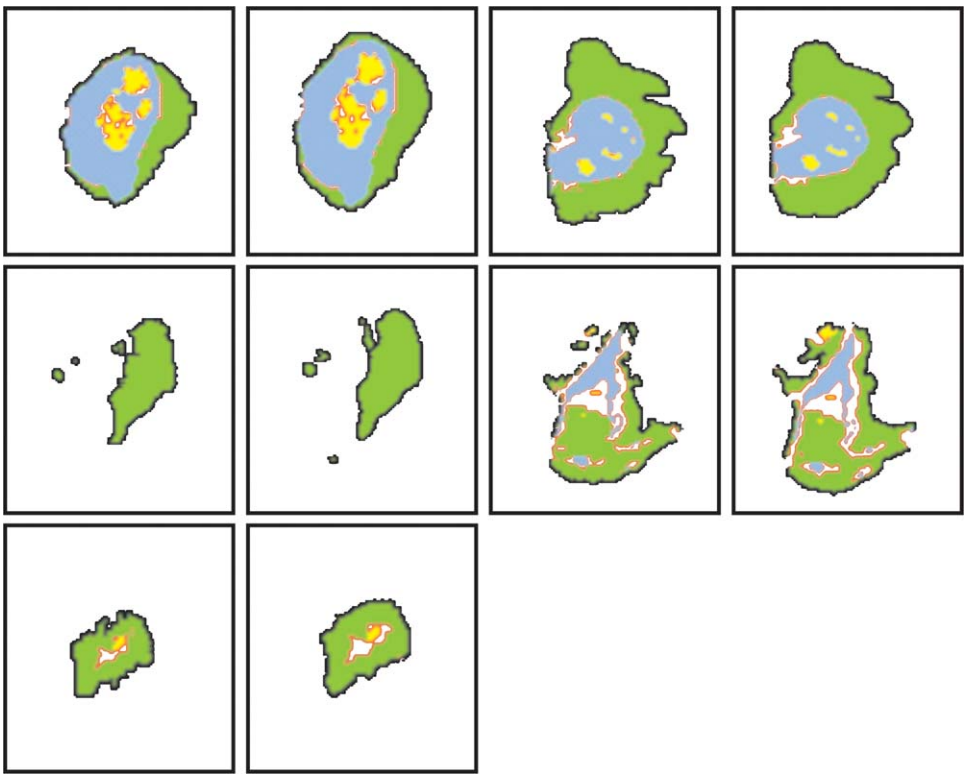


FIGURE 6 Ground truth and Segmentation: The first image is ground truth and the second image is the corresponding prediction [Color figure can be viewed at wileyonlinelibrary.com]

The purpose of this section is to exhibit details of the successful experiments. For a deep Convolutional network, implementation details are very important because little change in some hyperparameter could change results significantly. This work uses open source library Caffe (Jia et al., 2014). SGD Solver with step-based learning policy is used where learning rate is dropped whenever loss reduction becomes

TABLE 3 Tumor segmentation accuracy: A comparison

		Dice Score			Sensitivity			Specificity		
	Dataset	Complete	Core	Active	Complete	Core	Active	Complete	Core	Active
Proposed (SENet)	HGG	0.91	0.86	0.83	0.90	0.84	0.86	0.88	0.9	1.0
	LGG	0.86	0.82	0.81	0.84	0.80	0.81	0.84	0.84	0.78
	Combined	0.88	0.81	0.81	0.86	0.80	0.80	0.83	0.84	0.84
Proposed (IntNet)	HGG	0.88	0.90	0.80	0.85	0.90	0.73	0.76	0.81	0.73
	LGG	0.84	0.83	0.69	0.76	0.74	0.64	0.67	0.67	0.66
	Combined	0.90	0.87	0.80	0.86	0.80	0.81	0.73	0.73	0.66
Proposed (SkipNet)	HGG	0.85	0.88	0.82	0.83	0.87	0.80	0.77	0.77	0.78
	LGG	0.83	0.83	0.68	0.77	0.74	0.63	0.70	0.67	0.66
	Combined	0.87	0.86	0.79	0.83	0.79	0.80	0.73	0.71	0.65
Lun and Hsu (2016)	Combined	0.87	0.67	0.85	-	-	-	-	-	-
Dong et al. (2017)	HGG	0.88	0.87	0.81	-	-	-	-	-	-
	LGG	0.84	0.85	0.00	-	-	-	-	-	-
	Combined	0.85	0.76	0.74	-	-	-	-	-	-
Shi et al. (2016)	Combined	0.90	0.81	0.57	-	-	-	-	-	-
Xue et al. (2017)	Combined	0.85	0.70	0.66	0.92	0.80	0.69	0.80	0.65	0.62

smooth. Momentum for SGD solver and weight decay are set to 0.9 and 0.0005, respectively. During experiments, we observed that providing class-wise weights to balance class samples, in loss layer, can improve the performance significantly. These weights can be calculated in different ways. We tested average, median and maximum frequency based class balancing as described in (Eigen, Puhrsch, & Fergus, 2014).

The size of training data and mini-batch size also play a significant role in network performance. Due to limited memory on available GPU card, mini batch of size 40 is used. The obtained experimental predictions are exhibited in Figure 6.

We have validated three sub-tumoral regions as required by BRATS 2015 challenge: Complete, Core and Enhanced. The network training is done on categorical cross entropy loss function. The generated predictions are then processed by the script to calculate DSC, Sensitivity, and Specificity.

Dice similarity is the measure of similarity between the actual label and predicted label. Dice score is calculated for three sub-tumoral regions using the formula:

$$\mathrm{DSC} = 2\mathrm{TP}/(\mathrm{FP} + 2\ \mathrm{TP} + \mathrm{FN})$$

Sensitivity is used to measure the true positive rate i.e., for actual true value, how often trained network predicts true label and it is given by:

$$\mathrm{Sensitivity} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

and final measure used is specificity that tells the true negative rate.

$$\mathrm{Specificity} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP})$$

Moreover, Table 3 presents the average region-wise accuracy of segmentation:

We are aware of the fact that Isensee et al. (2017) performed well at the BRATS 2015 challenge, however, the results obtained by proposed method are good enough and better than many published works. In this work, we observed that VGG like architecture works good for tumor segmentation problem. We took this as proposed baseline and improved it. Proposed second model performed better than the first model that shows the application of BLI layer that could be beneficial in improving semantic segmentation results. The third model, SENet, performed even better than proposed second model that shows the weighting of feature maps in intermediate layers can produce better results.

5 | CONCLUSION AND FUTURE WORK

The current research has presented a comparable network architecture for brain tumor segmentation using multi-modal images. We presented three different models with increasing level of performance. The results exhibit that use of intermediate Convolutional maps and interpolation techniques are workable and produce promising results. Most of the networks presented in the literature are deep networks that need a lot of training to converge. However, proposed network structure is small, fast, and less memory demanding. As future directions, we will explore use of SE blocks at different layers. It is also interesting to explore whether one type of weighting strategy is better and which one or their combinations could perform better.

ACKNOWLEDGMENT

This work was supported by the Machine Learning Research Group; Prince Sultan University Riyadh; Saudi Arabia [RG-CCIS-2017-06-02]; The authors are grateful for this support.

ORCID

Amjad Rehman http://orcid.org/0000-0002-3817-2655

REFERENCES

Abbas, N., Saba, T., Mohamad, D., Rehman, A., Almazyad, A. S., & Al-Ghamdi, J. S. (2016). Machine aided malaria parasitemia detection in Giemsa-stained thin blood smears. *Neural Computing and Applications*, 1–16. https://doi.org/10.1007/s00521-016-2474-6

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561.

Brownlee, J. (2014). Discover feature engineering, How to engineer features and how to get good at it. Machine Learning Process.

Chaichulee, S., Villarroel, M., Jorge, J., Arteta, C., Green, G., McCormick, K., & Tarassenko, L. (2017). Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring.

Clark, T., Wong, A., & Haider, M. A. (2017). Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted MR images. In Machine Learning for Medical Image Computing, 14th Int. Conf. on Image Analysis and Recognition (ICIAR) (pp. 1–8).

Daykin, M., Beveridge, E., Dilys, V., Lisowska, A., Muir, K., Sellathurailan, M., & Poole, L. (2017). Evaluation of an Automatic ASPECT Scoring System for Acute Stroke in Non-Contrast CT. In: Valdés Hernández, M., & González-Castro, V. (Eds.), *Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science*, vol 723. Springer, Cham.

Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks. Medical Image Understanding and Analysis (MIUA) 2017.

Eigen, D., Puhrsch, C., & Fergus, R. (2014). Depth map prediction from a single image using a multi-scale deep network. In Advances in neural information processing systems (pp. 2366–2374).

Gan, Y., Xia, Z., Xiong, J., Li, G., & Zhao, Q. (2017). Tooth and alveolar bone segmentation from dental computed tomography images. *IEEE Journal of Biomedical and Health Informatics*.

Havaei, M., Davy, A., Farley, W. D., Biard, A., Courville, A., Bengio, Y., ... Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, *35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004.

Husham, A., Hazim Alkawaz, M., Saba, T., Rehman, A., & Saleh Alghamdi, J. (2016). Automated nuclei segmentation of malignant using level sets. *Microscopy Research and Technique*, *79*(10), 993–997. doi. 10.1002/jemt.22733

Iftikhar, S., Fatima, K., Rehman, A., Almazyad, A. S., & Saba, T. (2017). An evolution based hybrid approach for heart diseases classification and associated risk factors identification. *Biomedical Research*, *28*(8), 3451–3455.

Iqbal, S., Khan, M. U. G., Saba, T., & Rehman, A. (2017). Computer-assisted brain tumor type discrimination using magnetic resonance imaging features. *Biomedical Engineering Letters*, 1–24.

Isensee, F., Kickingereder, F., Bonekamp, D., Bendszus, M., WickHeinz-Peter, W., Klaus, H. P., & Maier-Hein, K. (2017). Brain Tumor Segmentation Using Large Receptive Field Deep Convolutional Neural Networks. In: Maier-Hein, geb., Fritzsche K., Deserno, geb., Lehmann T., Handels, H., & Tolxdorff, T., (Eds.), *Bildverarbeitung für die Medizin 2017. Informatik aktuell* (pp. 86–91). Berlin, Heidelberg: Springer Vieweg, https://doi.org/10.1007/978-3-662-54345-0_24

Jamal, A., Alkawaz, M., H., Rehman, A., & Saba, T. (2017). Retinal imaging analysis based on vessel detection. *Microscopy Research and Technique*, *80*, 799–811. https://doi.org/10.1002/jemt.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., … Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678. ACM, 2014.

Jie, H., Shen, L., & Sun, G. (2017). Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507.

Lun, T. K., & Hsu, W. (2016). Brain tumor segmentation using deep convolutional neural network. In Proceedings MICCAI-BRATS Workshop.

Mandal, D., Chatterjee, A., & Maitra, M. (2017). Particle swarm optimization based fast Chan-Vese algorithm for medical image segmentation. In *Metaheuristics for medicine and biology* (pp. 49–74). Berlin Heidelberg: Springer.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., … Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024.

Mughal, B., Muhammad, N., Sharif, M., Saba, T., & Rehman, A. (2017). Extraction of breast border and removal of pectoral muscle in wavelet, *Biomedical Research*, *28*(11), 5041–5043.

Muhsin, Z. F., Rehman, A., Altameem, A., Saba, A., & Uddin, M. (2014). Improved quadtree image segmentation approach to region information. *The Imaging Science Journal*, *62*(1), 56–62. doi. https://doi.org/10.1179/1743131X13Y.0000000063.

Norouzi, A., Rahim, M. S. M., Altameem, A., Saba, T., Rada, A. E., Rehman, A., & Uddin, M. (2014). Medical image segmentation methods, algorithms, and applications. *IETE Technical Review*, *31*(3), 199–213. https://doi.org/10.1080/02564602.2014.906861.

Pereira, S., Oliveira, A., Alves, V., & Silva, C. A. (2017). On hierarchical brain tumor segmentation in MRI using fully convolutional neural networks: A preliminary study. In 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG) (pp. 1–4).

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, *35*(5), 1240–1251.

Quine, B. M., Valery, T., Henok, M., & Richard, H. (2007). Determining star-image location: A new sub-pixel interpolation technique to process image centroids. *Computer Physics Communications*, *177*(9), 700–706.

Rad, A. E., Rahim, M. S. M., Rehman, A., Altameem, A., & Saba, T. (2013). Evaluation of current dental radiographs segmentation approaches in computer-aided applications. *IETE Technical Review*, *30*(3), 210–222.

Rahim, M. S. M., Norouzi, A., Rehman, A., & Saba, T. (2017a). 3D bones segmentation based on CT images visualization. *Biomedical Research*, *28*(8), 3641–3644.

Rahim, M. S. M., Rehman, A., Kurniawan, F., & Saba, T. (2017b). Ear biometrics for human classification based on region features mining. *Biomedical Research*, *28*(10), 4660–4664.

Rao, V., Sarabi, M. S., & Jaiswal, A. (2015). Brain tumor segmentation with deep learning. MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS), pp.56–59.

Saba, T. (2017). Halal food identification with neural assisted enhanced RFID antenna. *Biomedical Research*, *28*(18), 7760–7762.

Saba, T., Al-Zahrani, S., & Rehman, A. (2012). Expert system for offline clinical guidelines and treatment. *Life Sci Journal*, *9*(4), 2639–2658.

Saba, T., Rehman, A., & Sulong, G. (2011). Cursive script segmentation with neural confidence. *International Journal of Innovative Computing and Information Control (IJICIC)*, *7*(7), 1–10.

Saba, T., Rehman, A., & Elarbi-Boudihir, M. (2014). Methods and strategies on off-line cursive touched characters segmentation: A directional review. *Artificial Intelligence Review*, *42*(4), 1047–1066. https://doi.org/10.1007/s10462-011-9271-5

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., … Ang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Cardoso, M., et al. (eds) *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2017, ML-CDS 2017. Lecture Notes in Computer Science* (Vol. 10553, pp. 240–248). Springer, Cham.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In AAAI (pp. 4278–4284).

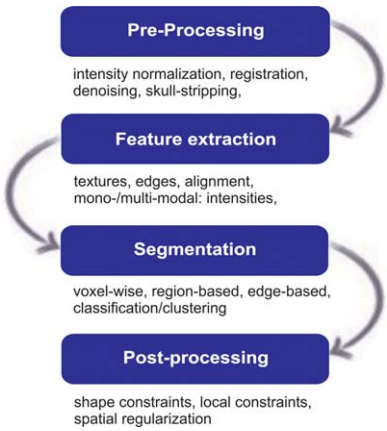
Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M. … Vercauteren, T. (2017). Interactive medical image segmentation using deep learning with image-specific fine-tuning. arXiv preprint arXiv: 1710.04043.

Wang, H., Wang, Y., Zhang, Q., Xiang, S., & Pan, C. (2017). Gated convolutional neural network for semantic segmentation in high-resolution images. *Remote Sensing*, *9*(5), 446–460. https://doi.org/10.3390/rs9050446.

Xue, Y., Xu, T., Zhang, H., Long, R., & Huang, X. (2017). SegAN: Adversarial Network with Multi-scale $L_1 $Loss for Medical Image Segmentation. arXiv preprint arXiv:1706.01805.

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis*, *43*, 98–111.

How to cite this article: Iqbal S, Ghani MU, Saba T, Rehman A. Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN). *Microsc Res Tech*. 2018;81:419–427. https://doi.org/10.1002/jemt.22994



SGML and CITI Use Only
DO NOT PRINT

The research presents a deep CNN to segment brain tumor in MRI. Proposed architecture consists of multiple CNN layers connected in sequential order using Convolutional feature maps at peer level. Experiments on BRATS 2015 exhibit promising results.