Brain Tumor Segmentation using CNN based Residual Neural Network

Christopher Sam Roy
Computer Science and Engineering
University at Buffalo
croy2@buffalo.edu

Abstract

Nearly 80,000 brain tumors are diagnosed in the USA each year. Although approximately 32% of them are considered malignant, Glioblastoma Multiforme (GBM) constitutes 45% of all malignant brain tumors [1]. It is also considered the deadliest type of brain tumor. The proposed model is trained to segment/identify the glioblastomas with the help of MRI images. Predicting the formation of a GBM tumor is a difficult task as it can show up in any part of the brain and have any shape and size. This property inclines the task of segmentation to a more traditional Convolution based Neural Network model. Also, manually segmenting the tumor is an arduous task as the number of images produced by MRI for a single patient is large. GBM being so malignant, it becomes even more essential to identify it early so that proper treatments could be started as soon as possible. Images obtained by MRI varies in intensities which requires a pre-processing step that normalizes the intensity across all images to help the model recognize the underlying pattern during segmentation rather than focusing too much on variable intensities. MRI images used for training the segmentation models are also highly imbalance as the tumorous region will always be only a small part as compared to the healthy tissues. This problem was addressed by training the models on 33x33 patches that are randomly cropped from training data and are balanced by data augmentation. The CNN model proposed consist of a 3x3 size kernel, it helps the model to give more importance to the tumorous region as compared to the surrounding areas that are further away from the tumor. The model is inspired from ResNet model that provides a residual path that skips some layers and helps the model in better understanding the context. The residual path also helps the model to generalize well as deeper models are more prone to overfitting. The model classifies each pixel to be in one of the 5 classes- necrosis (dead tissues), edema (swelling), enhancing tumor, non-enhancing tumor, normal tissues. The accurate segmentation of these regions is very important in determining the course of treatment. The segmentation output by the model was evaluated as a measure of DICE similarity coefficient with respect to the ground truth (manually segmented) image. The model was trained and evaluated using the Brain Tumor Segmentation Challenge 2013 dataset (BRATS 2013) and the results were comparable to the models reported in their leaderboard.

1. Introduction

Brain Tumors especially Glioblastomas are very aggressive and malignant at the same time, which results in an aggressively reproducing tumor that becomes lethal quickly. They are nourished by an ample and abnormal tumor vessel blood supply. The tumor is predominantly made up of abnormal astrocytic cells, but also contain a mix of different cell types (including blood vessels) and areas of dead cells (necrosis). Glioblastomas are infiltrative and invade into nearby regions of the brain [2]. They are often surrounded by swollen tissues or a fluid (edema). The median survival time with glioblastoma is 15 to 16 months in people who get surgery, chemotherapy, and radiation treatment [3]. For diagnosing and planning the treatment of an individual MRI scans are generally used. The scans are then carefully studied by experts to segment the tumorous region manually. This process takes quite some time as the MRI imaging forms many scans (in the form of slices) of a single patient. Even when a patient is undergoing treatment the MRI scans hold an essential step as it helps the doctor to understand how the tumor is responding to the adopted treatment method. Glioblastomas being such an aggressively spreading tumor, the doctors cannot afford any delay in the segmented result of MRI images. Hence it is important to automate this process of segmentation with utmost precision. The segmented MRI slice from the Neural Network model with the largest area from axis could be then intricately analyzed to fine tune the result. This could save significant man-power and precious time of both the patients and the doctors.

If the brain tumor segmentation problem is analyzed technically it can be observed that there is no single positional pattern or specific size/structure that could help in segmentation as the tumor could appear anywhere in the whole brain with any shape and size. Also segmenting

healthy tissues from tumorous ones are a result of slight variations in the intensities of the tissues in MRI scans. MRI scans in itself include variable intensities across slices of a single patient and also across different MRI scanners. This makes the segmentation task more complicated and hence ensures the need of a normalizing step and a bias correction step that makes all the images to be uniform. As mentioned before, the segmentation task relies on varying intensities among the brain tissues, hence a variety of imaging modalities can be used for mapping tumor-induced tissue changes, including T2 and FLAIR MRI (highlighting differences in tissue water relaxational properties), post-Gadolinium T1 MRI (showing pathological intratumoral take-up of contrast agents), perfusion and diffusion MRI (local water diffusion and blood flow), and MRSI (relative concentrations of selected metabolites), among others. Each of these modalities provides different types of biological information, and therefore poses somewhat different information processing tasks [4].

For the object detection and segmentation task, deep learning models and especially CNN's have been widely popular because of its hierarchical structure and powerful feature extraction capabilities from an image making it a very robust algorithm [5]. As the CNN models operate using kernel filters they are bound to take the context into consideration, which proves helpful in biological segmentation tasks. The tumor segmentation task however, is very sensitive to the context because, if the kernel size is bigger the model may tend to learn positionally more from the training data whereas the position of occurrence of a tumor is less predictable in general which leads to overfitting. In contrast, if the kernel size is smaller, the model may fail to learn important regions surrounding the core of the tumor, leading to underfitting. Some models also try to use the 3D nature of the MRI data by using 3D kernels, however it increases the computation load [6]. The proposed model uses the remarkable work done by Kaiming He et al, in Deep Residual Learning for Image Recognition [7], where they proposed a way to build deeper networks without a significant increase in the complexity of the models. The idea was to not only make the kernel filters extract features from the output of its previous layers but also from the output of a layer that is much above the current layer. It creates a residual path in the network making it easier to optimize and generalize better. In the preprocessing step, the intensity normalization method proposed by Nyúl et al [8], have been implemented to offset the variability in the intensities of MRI images taken from varying scanners. To generalize the model from not learning much of the structural or the spatial features of the tumor, an image augmentation method was also used in addition to the training data.

2. Task Statement

The Brain Tumor segmentation task using a neural network-based model can be broadly divided into two types of models as mentioned by Havaei et al [9]- Generative model and Discriminative model. Generative models demand domain knowledge as the idea is to map the tumorous regions on an atlas (template) of MRI images of several healthy brains. This allows the model to learn the texture and intensity of the tissues at a particular region of the brain. The model aims to segment the tumorous region from healthy region by computing a probability of occurrence based on the intensity and texture localized to a particular region. A well-defined threshold is used to determine if the probability values computed is enough to flag the tissues as tumorous. The models have also been proposed that learns the differences in contours in the brain tissues and correlates it to the left and right brain symmetry. Discriminative models do not demand much prior domain knowledge as it focuses on extracting the underlying features between several patches of input training data. This type of models aims to understand the underlying distribution (histogram) of pixels, textures, contours, edges, gradients, positional/regional shapes and sizes. In MICCAI Brain Tumor Segmentation challenges conducted each year, discriminative model based on ensemble learning such as Random Forest has been mostly successful.

The challenge with training a discriminative model is that it works best when the training data is balanced. The class balanced training data enables the model to learn the underlying features that distinguishes the classes from each other. Although in general the training data is not class balanced and especially in case of Brain Tumor segmentation, the region covered by tumor is smaller than the healthy part of the brain. This imbalance might provide a bias to the model hence it is important to do preprocessing or augmentation on the data before training the model with it. The importance of preprocessing in case of brain tumor segmentation becomes evident when the most accurate discriminative models in the BRATS challenge are analyzed. Most common preprocessing was to extract 3D or 2D patches from the MRI images before feeding it to the model. In the proposed model, the input is a 2D patch of size 33x33 that is randomly selected from all the slices available in the training dataset.

A CNN based model is often considered among the first choices to address a segmentation/detection challenge. This is because of the generalizability of the CNN feature extraction algorithm which gives more importance on the contours, varying gradients in the patches rather than on the position. Hence naturally in case of brain tumor segmentation, where the desired region is not specific to one part of the image CNN models are preferred. In the proposed model, a 3x3 size kernel filter is used which

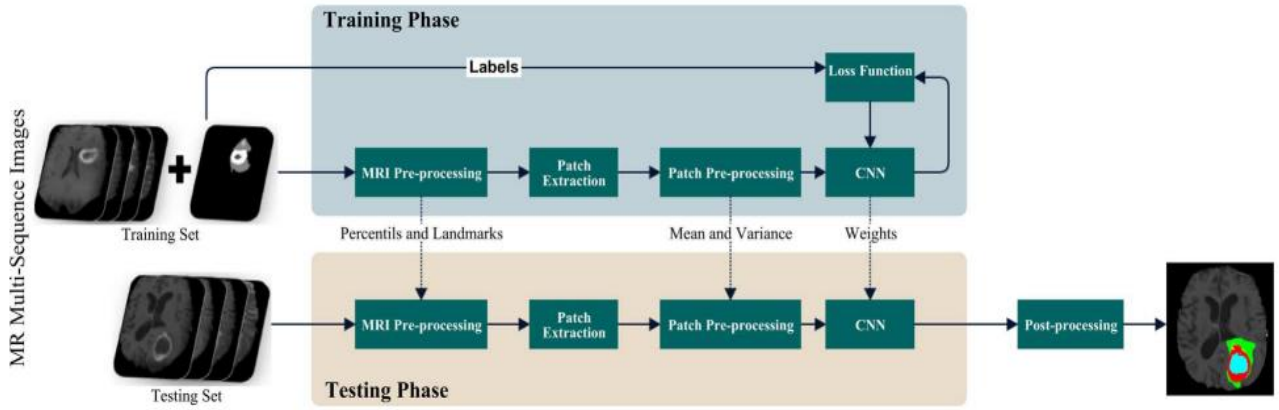

Figure 1: The pre-processing, training and testing phase [11]

allows the model to be deeper and extract underlying features more efficiently. Creating a deeper model allows the addition of more non-linear activation functions as well, which increases the capability of the model to learn non-linear feature correlations. To make the model deeper and preserve the complexity of the model from increasing too much, a ResNet inspired residual path was included. It helped the model in preserving the input structure not only to initial layers but also to a much deeper layer.

As mentioned earlier, the input to the model are randomly selected patches and the objective of the model is to predict the class (5 class classification) of the center pixel. The training images being unbalanced with respect to classes it was essential to select the patches carefully such that we balance the number of training samples per class. This was achieved by first selecting the center pixel and then extracting a patch with that pixel at the center. Similarly, patches with same central pixel value corresponded to one class and the number of patches selected was even among all five classes that are necrosis (dead tissues), edema (swelling), enhancing tumor, non-enhancing tumor, normal tissues. For better generalization of the model, the patches selected were also augmented in a certain way (rotation augmentation) to make sure the model does not give more importance to the structure and size of the tumor. When the model segments all the slices corresponding to a single brain, the segmented region with largest core area from its axis was considered the best segment because most of the slices of a brain wasn't representing the complete brain structure as 3D MRI images consists of several 2D slices.

The training dataset consists of images from 4 different modalities (ways of MRI scanning) that illuminates different tissues with varying intensities. This information can help the model learn to segment between tumorous and health tissues by providing more data points to extract features from. The proposed model takes this into consideration when extracting patches. The input size of the model is 4x33x33, which represents 4 modalities of 33x33 patches. The popularity of CNN based models in the domain of bio-medical imaging is highly growing because of several algorithmic advantages, one such interesting research was done on boundary reconstruction of neural circuity from 3D electron microscopy data by Huang and Jain [10]. The proposed model aims at similar segmentation task but in the field of brain tumors.

3. Method

Figure 1 describes the flow of training the model and then testing the learnings on a test set. The complete structure can be divided broadly into three steps- Pre-processing and patch extraction, CNN model training, Testing.

3.1. Pre-processing and patch extraction

The MRI images are bound to contain irregular intensities in different tissues among the slices of a patient. The intensity of tissues also varies in the scans taken at a different time from the same scanner. This violates the uniformity in the scans of the same patient, if the follow up scans are needed to be compared with the previous scans. To rectify this problem an intensity normalizing step is necessary on all the MRI scans. A normalizing method across modalities as proposed by Nyul et al [12] was implemented as a pre-processing step. However, applying only intensity normalization was not enough to rectify the contrast in same tissues across various slices of the MRI scans. Hence a bias correction method called as N4ITK [13] was implemented on top of the intensity normalization step.

Patch extraction is necessary because of unbalanced nature of the MRI slices in the training data. If the complete slice is provided as an input to the model, it will be biased towards classifying the pixels to a healthy (non-tumorous) class as majority of the pixels in the slice will belong to this class. Model will not be able to learn the underlying features

that are necessary to classify a pixel to a tumorous class and this requires knowledge of its surrounding regions. Extracting patches corresponding to a center pixel value helps the model learn to classify each pixel based on its surroundings. The BRATS dataset includes MRI scans from 4 different modalities that use varying radio frequency pulse sequences to illuminate different tissues differently. This information can be used by the model to differentiate between the classes as each pulse sequence show the tissues in different intensities thereby increasing the contrast between the classes. The modalities present in the BRATS dataset are- Fluid Attenuated Inversion Recovery (FLAIR), T1, T1-contrasted, and T2. Figure 2 shows how the 4 modes of imaging show contrasts in the same brain slice.

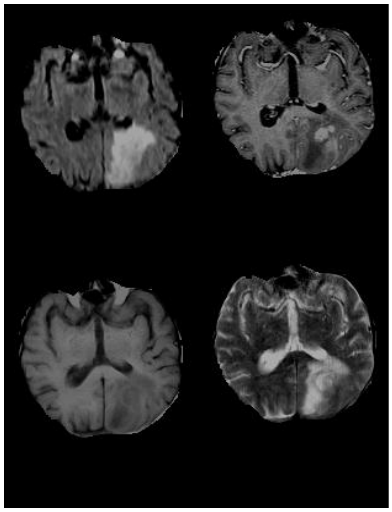

Figure 2: Single slice of MRI scan in FLAIR (top left), T1c, T1, T2 (bottom right) modalities

The proposed model uses these differences in modalities to classify the class a pixel will belong in. Therefore, patches are extracted across all 4 modes concurrently. Each patch size becomes 4x33x33 (here patch size is 33x33 across 4 modalities). Hence the convolution layer treats the modalities as a different channel and learns the useful differences among them to improve the classification.

Patch extraction algorithm could be implemented in different ways. I decided to randomly select pixel values belonging to a class and then selected the patch with that pixel at its center. This allowed the input data to be balanced across all classes, since the number of patches selected per class is same. This method faces a problem when selecting patches for the healthy tissue class. The ground truth displays all the healthy tissues and background of the image slices (black) to belong to same class. It only shows the tumorous regions and hence randomly selecting patches belonging to healthy tissues mostly contained patches of background (zero intensity). This led to model wrongfully classifying most of the tissues as tumorous (because of the absence of patches containing healthy tissues). To mitigate this issue, a threshold was decided so that the patches that consist of pixel intensities above the threshold are kept while others are discarded. The threshold decide was 75%, such that the selected patches will contain more than 75% of non-zero pixel intensities. Figure 3 shows the patches selected across all four modalities belonging to different classes (central pixel). To avoid the model from overfitting the patches were randomly augmented so that the model does not give more importance to the structure or orientation of the tumorous regions. Only rotation augmentation was used on the patches.

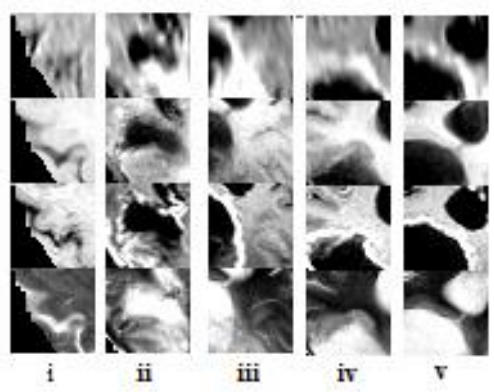

Figure 3: Patches of size 4x33x33 from all 5 classes- (i) healthy tissues, (ii) necrosis (iii) edema (iv) non-enhancing tumor (v) enhancing tumor

3.2. CNN model Architecture

The proposed model has CNN based ResNet inspired architecture. The CNN based models have long been gaining prominence for the task of object detection and segmentation. The idea behind a CNN model is to have convolutional layers that have a set of kernels to perform the convolution operation over the input to extract the features. The extracted features influence the weights by which the kernels are connected between each layer. These weights are updated during the training phase by calculating the cost function during each epoch and backpropagating the gradients to update the weights. CNN models are less susceptible to overfit the training data because the same kernels are shared among all the training images and hence not allowing the complexity of the model to increase too much. The output from the convolution layers are usually applied to a non-linear activation function which infuses much needed non-linearity in the model. The size of the kernels is determined by analyzing the importance of context information in making the classification. Since same kernels are passed through the entire input image the model output does not get influenced by the position of the patch in the image.

The ResNet architecture was introduced to build deeper networks without compromising the overall complexity of the model. In sequential CNN architecture the features extracted by the deeper layers were increasingly abstract which lead to overfitting the training data. This problem is addressed by providing skip connections, that allows alternate way of data to flow between the layers in the network. When implementing a model with skip connections the input and output sizes of the layers needed to be matched, this is insured by using an extra convolutional layer in the skip connection with a stride of more than one. The activation functions are applied differently in a ResNet model, wherever there is a combination of sequential and skip connection the activation layer is applied after the combined inputs. This can also be observed in Figure 4, where ReLU activation is applied to the combination of sequential and skipped connection.

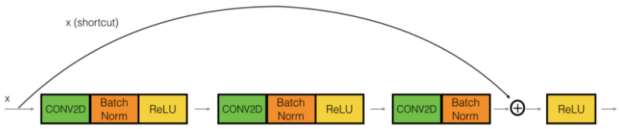

Figure 4: Skip connections in a ResNet architecture [14]

The proposed model consists of two Max-pooling layers applied after combining the outputs of skip and sequential connection. The max-pool layers are often used to join features and possibly eliminate redundant features. It is also used to increase the number of channels by reducing the size of the input. However sometimes using max-pool layers could lead to the model eliminating important features and to limit this the pooling size was kept 3x3 while maintaining the stride of 2x2, this resulted in an overlapping pooling operation.

The model included 3 fully connected layers at the end and used softmax activation function on the output of last layer. The activation function used after all the other layers was 'Leaky ReLU' which provided a small negative slope and helped in avoiding the problem of vanishing/exploding gradient during the training.

To avoid overfitting the model several regularization steps were implemented. The input patches extracted were augmented to increase the number of images the model could train on, while maintaining the balance (class-wise) in the training data. The augmentation was kept limited to angular shift or rotation because the patches were selected such that model learns to classify the center pixel and hence any other augmentation could result in the model learning irregular correspondences between features and central pixel values. Another regularization method was adding dropout to the fully-connected layers so that these layers learn more domain specific features rather than image specific.

The loss function used to train the model was Categorical cross-entropy shown in Figure 5, it allows the model to calculate the probability of the image patch being classified to 5 classes.

$$H = -\sum_{j \in voxels} \sum_{k \in classes} c_{j,k} \log(\hat{c}_{j,k})$$

Figure 5: ĉ represents the probabilistic predictions and c is the target [11]

Figure 7 shows the complete architecture of the proposed model with the input/output dimensions of each layer. It also shows the flow of data in the model including the skip/ residual connections in the model. The layer "Add" is used to combine the sequential and skip connections together.

While training the model, the Stochastic Gradient Descent optimization method was used with a momentum of 0.9. The addition of momentum to the optimization algorithm help the model to converge faster as it directs the gradient vectors in the right direction during the backpropagation step. Nesterov accelerated gradient method was also used to update the momentum so that the gradients are adequate to escape the local minima.

3.3. Testing and Evaluation criteria

The model was validated on BRATS 2013 dataset, which consists MRI images of 20 patients with either of anaplastic astrocytomas or glioblastoma multiforme tumors. There is a total of 155 slices in each of the 4 modalities of scanning for a single patient. Hence each patient contributes about 620 images in the dataset. It also consists of manually segmented tumors in all the slices which are considered as ground truth during the training. The ground truth for test set is not available publicly and hence a small part of training images was kept unseen from the model to evaluate its output segmentation with the ground truth. The ground truth contains tumor segmentation into 5 classes- necrosis, edema, enhancing tumor, non-enhancing tumor, normal tissues and is evaluated with the segmentation of the model using Dice Similarity Coefficient metric. This metric measures the spatial overlap between the segmented output and ground truth image. The complete segmentation of MRI image is not considered for evaluation rather only the core of the tumor (necrosis, enhancing tumor and non-enhancing tumor), enhancing tumor and complete tumor (except background and healthy tissues) are considered. The formula used to calculate Dice similarity coefficient for the three regions mentioned above is explained in figure 6.

$$DSC = \frac{2TP}{2TP + FP + FN}$$

Figure 6: Calculating Dice Similarity coefficient using TP- True Positives, FP- False Positives, FN- False Negatives [15]

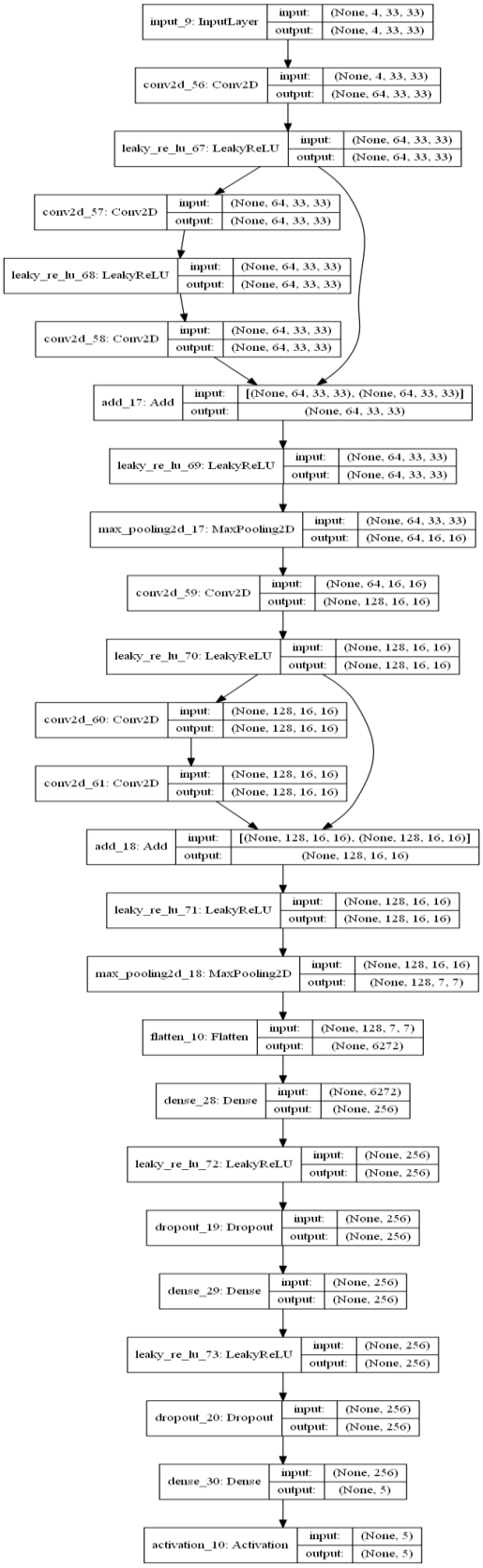

Figure 7: Architecture and Layer dimensions of the model

4. Experiments and Results

The brain tumor segmentation problem is basically a multi-class classification problem. In this case the number of possible classes are 5. The input is a patch of size 33x33 with 4 modalities stacked on top of each other making the model input be of dimension 4x33x33. The model is trained on patches sampled from the training dataset so that the model learns to segment the pixels without any class bias. A total of 20,000 patches (10,000 original+10,000 augmented) were sampled from the training dataset with their corresponding ground label.

Stage	Hyperparameter	Value
Initialization	weights	Xavier
Leaky ReLU	α	0.333
Dropout	probability	0.2
Training	Epochs	20
	Batch	128
	Learning rate	0.003
	Momentum	0.9

Table 1: Hyperparameters used for training the model

The performance of model was evaluated using different regularization techniques, adding/removing some layers, changing the pre-processing of the model input while keeping the hyperparameters show in Table 1 as constant. The changes in Dice coefficient was observed and for better comparison of the segmented output, the regions were colored differently as shown in Figure 8.

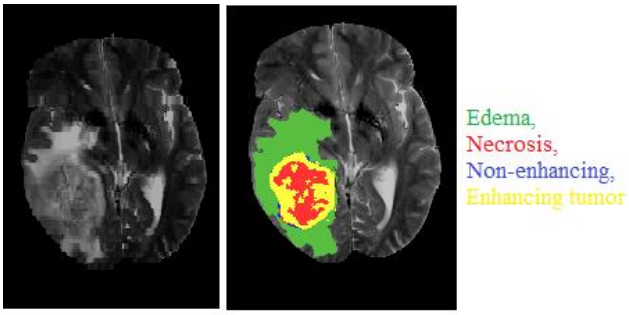

Figure 8: Ground Truth segmentation with respective color

The model classifies each pixel in the given input to one of the five classes. Changes in the architecture of the model results in changes in the classification and is shown in figure 9. A direct comparison can be made with respect to the ground truth label and affects of different setting in the model. It was observed that using either of Average pooling or Batch normalization resulted in smoothing of the segmented patches related to edema. It is observed by increased number of green patches in column ii, iii, iv of

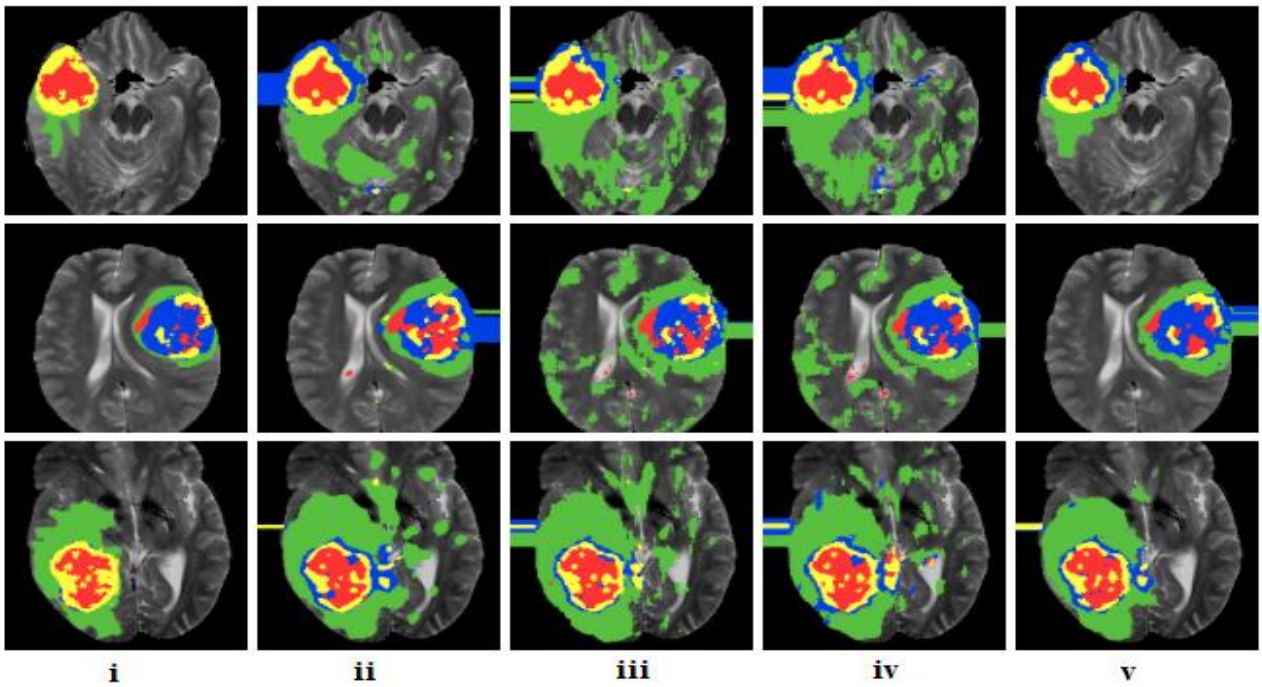

Figure 9: Comparison of classification output of the model on 3 different patients with changes in the architecture, each row corresponds to a single patient- Column (i) Color labeled Ground Truth, Column (ii) model with average pooling instead of Max-pool, Column(iii) model with batch normalization layers added, Column (iv) Model with residual path and Batch Normalization, Column(v) proposed model with residual path

figure 9. It was also observed that the model with added residual paths better classified the core regions of the tumor and is evident by the Dice Score in table 2. Some small green patches can be randomly seen in the output image away from the tumor, it is because the model classifies Cerebrospinal fluid as edema (swelling), since some modalities show them in similar intensities. This is largely evident in model with Batch Normalization layers. The proposed model gives evidently best classification results on the tumor core. It can also be observed in figure 9, the column(v) is closest to the ground truth for each patient. Since the model is trained using patches sampled from the original MRI slice, it struggles to classify the fine boundaries and outputs a smoothed patch as compared to the ground truth which has more details in its class boundaries.

Overall the proposed model with residual paths allowing skip connections was able to perform better than the sequential model and was also able to better segment the core region of the tumor.

Experiments	**Dice Similarity Coefficient**		
	Complete	**Enhancing**	**Core**
Patches without augmentation	0.92	0.71	0.81
Patches with augmentation	0.92	0.88	0.88
Original architecture without Residual path	0.92	0.73	0.77
Original architecture with Batch Normalization	0.89	0.83	0.76
Proposed model with residual path	0.92	0.90	0.91
Proposed model with residual path and batch normalization	0.93	0.85	0.87

Table 2: Changes in Dice score with different model settings

References

[1] https://www.nfcr.org/blog/blog7-facts-need-know-brain-tumors/
[2] https://www.abta.org/tumor_types/glioblastoma-gbm/
[3] Glioblastoma: Survival Rates, Treatments, and Causes (healthline.com)
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4833122/
[5] Convolutional Neural Networks — Simplified | by Prateek Karkare | AI Graduate | Medium
[6] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7426413
[7] https://arxiv.org/abs/1512.03385
[8] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=836373
[9] https://arxiv.org/pdf/1505.03540v3.pdf
[10] https://arxiv.org/pdf/1310.0354.pdf
[11] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7426413
[12] https://ieeexplore.ieee.org/document/836373
[13] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3071855/
[14] Understanding and Coding a ResNet in Keras | by Priya Dwivedi | Towards Data Science
[15] https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient