

# FR-MRInet: A Deep Convolutional Encoder-Decoder for Brain Tumor Segmentation with Relu-RGB and Sliding-window

Farshid Rayhan

Department of Computer Science and Engineering  
United International University  
frayhan133057@bscse.uui.ac.bd

## ABSTRACT

Automatic detection of tumorous tissue in MRI scans plays an important role in computer-aided diagnosis. We present a novel deep fully convolutional encoding architecture for semantic segmentation of brain MRI scans termed, FR-MRInet. This trainable encoder works with a corresponding decoder of a fully connected network. The 32 layer deep encoding architecture is inspired by VGG16 and InceptionV3. The novelty of FR-MRInet is its architectural design that efficiently reduces input to a lower resolution feature map(s). The encoder uses strides instead of pooling in various layers to reduce feature maps without losing spatial information. We used a non-overlapping sliding window with a novel activation function called, Relu-RGB to train the model so that the model directly produces the final output instead of a mask. We compared our model with well known imagenets such as Alexnet and VGGnet, other recent models proposed by researchers testing for pixel wise accuracy, intersection over union (IoU) and mean square loss value. We conducted our experiment on BRATS dataset for benchmarking and one of the latest dataset which was proposed in 2016 consisting of, 3064 T1-weighted contrast-enhanced images from 233 patients. We also show that FR-MRInet provides an impressive performance on live images detecting tumors as well. To further investigate the matter, we have consulted with an MD about the usefulness and the future of these kind of projects. Our code is open sourced and freely available at [github.com/farshidrayhanuui/FR-MRInet](https://github.com/farshidrayhanuui/FR-MRInet).

## Keywords

Deep Convolutional Neural Networks , Semantic Pixel-Wise Segmentation , Encoder , MRI scan , Brain Tumor

## 1. INTRODUCTION

Nowadays, the usage of digital images for medical diagnosis are increasing rapidly. MRI scan is a kind of brain test results of which can lead to identifying presence of tumor in brains. Even though there are other kinds of tests for this same purpose, MRI is most popular among them for its zero exposure to ionizing and high soft-tissue contrast. Regardless of the convenience of MRI scans,

the brain tumor identification and classification still remains a challenging task. The conventional method requires a radiologist to review and analyze the MRI scan and provide proper interpretation of the test results. One of the major down side of this system is the involvement of human analyst. The whole process of tumor detection and classification depends solely in the skills and expertise of the radiologist. Moreover, this is also a very impractical solution where large numbers of data are involved. Therefore, computer assisted diagnosis are highly desirable for addressing this problem. The task primarily consists of two sub problems: (1) Identification of abnormal tissue, i.e, whether the brain contains any tumor cells or not. (2) Classification of the tumor type.

Automatically categorizing the tumor type is a relatively more challenging task comparing to the binary classification of normal and abnormal tissue and convolutional networks are found to be very successful in biological tasks [13, 53]. Thus currently studies aims to develop an approach that can classify and discriminate different pathological tumor type. Researchers have proposed many automatic and semi-automatic techniques for detection and identification of brain tumors [17, 25, 72]. Typically the process follows are certain pattern, first the tumor is detected and segmented then the segmented part is classified. The classification task involves two task, feature extraction and classification. Since classification accuracy is highly dependent of informative features, researches have proposed various ways such as, intensity and texture based features [49, 24], Gabor filters [25], wavelet transform [27], GLCM [80] to extract features for better classification accuracy. In [25], the authors proposed a 3D vortex based segmentation and classification method using Gabor features and adaBoost classifier. In [58], an automatic classification models based on least squares SVM was presented to identify normal and abnormal tissues. Texture based features with fuzzy weighting and SVM were employed in [24] for multi-class classification. The authors of [27] used discrete GLCM [80] and wavelength for detection and classification. Patch based BoW representations have also been a notable representation scheme where the pixels are replaced with image patches and vector quantization are replaced with scalar quantization. Generally features extracted by BoW, GLCM are computed in a global scale which inevitably ignores spatial informations. There

have been several approaches proposed to address this issue as well [26, 59, 20, 33].

In order to research on a topic a rich collection of dataset is required. For the task brain tumor detection and classification from MRIs many researcher have open sourced their datasets. Among them CT scan, X-Ray, MRI scan of various parts of human body are available. *Open-Access Medical Image Repositories* [3] is one for the most largest on-line source for data collection. This repository not only contains its own datasets but also datasets from other recognized organizations like *Cancer Imaging Archive* [1], *Oasis Brains* [2]. It also contains datasets proposed by individual researchers as well [47, 77]. Regardless of the quality of datasets, researchers have to choose datasets very cautiously. Many of the datasets are outdated and others fall outside the scope of intended research goal. Due to the rapid improvement of scan quality it is often advised to use datasets not older than 5 years in order to keep the research up to date. Recently in [11], a dataset was created and open sourced for the purpose of further research in to the specific topic of brain tumor detection and classification. The dataset contains 3064 images (image resolution  $512 * 512 * 1$ ) containing three typical brain tumors: (a) meningioma; (b) glioma; and (c) pituitary tumor. It was acquired from Guangzhou, Nanfang Hospital, Tianjing Medical University and General Hospital of China. This dataset recently have been receiving attention of other researchers as well [78, 4]. In our research, we use the mentioned dataset with a slight modification.

The contributions of this paper are the following,

- We demonstrate a novel Encoding architecture for MRI scans with conclusive experimental results in support of its effectiveness.
- Manipulate the original dataset for better effectiveness.
- We use a non overlapping sliding window to work on the actual  $512 \times 512 \times 3$  sized images instead of resizing them to a lower resolution.
- We demonstrate comprehensive comparison between the effectiveness of using De-convolution and Fully Connected Networks as decoder using several top imagenets including our proposed encoding network.
- We propose a novel activation function that allows to generate output images accurately without the need of pixel-wise classification.
- We analyze FR-MRInet's performance on not only the whole dataset but also on several subset of the dataset and draw conclusions on the effect of each sub sets.
- We also compare our method with recent deep learning model which were designed for pixel wise segmentation.
- We also convert some of the recent deep architectures that works on brain MRI scans to pixel wise image segmentation network which were not originally designed for tumor segmentation then compare and analyze theirs performance as well.
- Furthermore, we employ Neighborhood Cleaning Rule as a smoothening method on the output images which boosts up the performance significantly.
- Finally, we have also consulted with an MD for professional opinion on several aspects of the project.

The article is structured as follows, in section 2 we discuss about important related works done so far. In section 3 we go further in details about image segmentation. In the following section, we discuss the materials used in our experiment and its importance. The methods used in this experiment, including our proposed encoding

architecture, are discussed in section 5. Section 6 describes the process of constructing training data and various configurations of the experimental setup. We provide our results in section 7 with detailed discussion. Finally, we provide some possible future works in section 8 and conclude the paper.

## 2. RELATED WORKS

Object detection in images has been an area of research interest for over a decade. Haar features were first introduced by *P Viola* in 2001 for rapid object detection using boosted cascade [67]. Following that, researchers have been proposing better and faster object detection methods ever since. Haar like features with boosted classifiers [68, 57, 52, 51] received a lot of attention [36, 35] for its accurate and real time detection but due to its limited capability to scale it started to lose its applicability for more complex tasks.

Before convolutional networks, many methods have been proposed by researchers for image classification like, Class-specific hough forests [14], Bayesian modeling [60], ensemble SVMs [39], Multiple kernels [66]. These works can be divided in to two categories, (1) Feature extraction method and (2) Classification method. In feature extraction methods authors have proposed several ways to extract informative features that was aimed to improve classification accuracy. In the classification methods the goal was to propose better classifiers to improve classification quality. Despite of their impressive performances, each of those methods had one primary limitation, that is the methods often failed to generalize. Which means even though their performance on datasets were impressive, the methods often performed poorly in real world applications.

In the recent years, deep convolutional-nets started to receive a lot of attention for its ability to extract features and classify images with great accuracy. Even though convolutional nets are computationally quite costly, its impressive capability to generalize often out weights that limitation. Object detection task has two sub-tasks (1) Objected segmentation, the task of locating the object in the image, (2) object classification, the task of classifying the located object. Segmented object is also known as **Region Of Interest (ROI)** where the object to be located is regarded as the interested region. The task of finding ROI from an image is comparatively more complex than classifying it. For image classification, the task is to classify the image only. Researchers have proposed many convolutional network for image classification, among them some of the notable image classifiers are Alex-net [31], VGG-net [62] and google's Inception-net [64]. While these nets are able to classify images with great accuracy they do not indicate the location of the object in a image. For the task of finding ROI in a image, several convolutional architectures have been proposed known as **Region Proposal Networks** or RPN. These networks are especially aimed to locate objects in images and trained in a end to end scheme. These methods became well accepted in the computer vision community for its ability to scale without compromising quality and speed [65, 23, 19].

In 2014, a model named R-CNN was introduced which merged the task of locating objects with image classification for a fully automated object detection model that can not only identify the location of an object but also classify [16]. Later improved version of the mentioned model was introduced named Fast R-CNN [15], Faster R-CNN [54] and Mask R-CNN [22]. Faster R-CNN was the first model to use RPNs for location proposing which significantly improved the speed of detection. The methods were trained in the following way, given an image as input and the output was an array containing the *X, Y* co-ordinates of the object with height and width followed by a confidence indicator ranging from 0 to 1. The

Network	Decoder	Dataset	Mean Square Loss (train)	Accuracy (train)	Accuracy (validation)
Alexnet	FCD	Top view	12.3	69.2	69.5
Alexnet	Deconvolution	Top view	16.5	70.0	60.2
VGGnet	FCD	Top view	8.3	72.6	63.1
VGGnet	Deconvolution	Top view	9.7	72.3	62.9
FR-MRI <sub>net</sub>	FCD	Top view	<b>3.9</b>	80.8	<b>83.0</b>
FR-MRI <sub>net</sub>	Deconvolution	Top view	4.5	<b>81.5</b>	81.8

Table 1. : Table containing accuracy on train and validation set of top view using Alexnet, VGGnet and FR-MRI<sub>net</sub> with both FCD and Deconvolutional decoder

Network	Decoder	Dataset	Mean Square Loss (train)	Accuracy (train)	Accuracy (validation)
Alexnet	FCD	Side view	12.3	70.2	71.6
Alexnet	Deconvolution	Side view	16.5	69.1	71.3
VGGnet	FCD	Side view	8.3	72.1	72.2
VGGnet	Deconvolution	Side view	9.7	72.1	72.4
FR-MRI <sub>net</sub>	FCD	Side view	<b>3.9</b>	<b>75.7</b>	75.9
FR-MRI <sub>net</sub>	Deconvolution	Side view	4.5	74.0	73.8

Table 2. : Table containing accuracy on train and validation set of top view using Alexnet, VGGnet and FR-MRI<sub>net</sub> with both FCD and Deconvolutional decoder

X, Y co-ordinates with high and width forms a rectangular shaped box around the indented object. These rectangular shaped boxes are known as anchor boxes. The shape and size of the box remains another challenging area of research interest [61, 22].

Recently, pixel-wise object classification or image segmentation have been receiving a lot of attention. In 2015, an image segmentation method named *SEGnet* [7] mainstreamed the concept of pixel classification. Image segmentation is a training process where instead of trying to find X, Y height and width of the object's location, each pixel's probability of belonging to that object is fed to the network. One of the benefit of this process is that it allows segmentation independent of any anchor box shapes [42, 40]. In the last few years, variants of deep CNN models have been proposed for image segmentation. In [42], 6 separate CNN layers were used from there different neighborhood to predict labels of each pixel. A network for 2 class patch-wise prediction was proposed in [41] where the final fully connected network outputs  $16 \times 16$  patches of pairwise labels. A detailed comparison between Fully Connected Networks (FCNs) and per pixel CNNs have been done in [69]. Anatomically Constrained Neural Networks or ACNNs have been applied to cardiac image enhancement and segmentation [48]. In [28], a similar comparison was reported with use of one upsampling layer. In both cases [48, 28], the results stayed below state of the art because the encoder-decoder networks lacked skipped connections to fully support the upsampling steps and also the results were in favor of FCNs. Atrous Convolution along with CNNs and FCNs was also found very useful for semantic image segmentation [10].

In [11], an automatic classification method for classifying tissue type was proposed which was able to propose ROIs accurately 82.31% for intensity histogram, 84.75% for GLCM and 88.19% for BoW model. They used their feature extraction method with SVM, K-Nearest Neighbor (KNN) and sparse representation-based classification (SRC) [71] to check and verify the features ability extract informative features. In the article [80], the authors have

also open sourced their dataset constructed with 233 real life patient's brain MRI scans. In 2016, the same datasets was used in [73] where the authors proposed spatial pyramid matching kernels to address the problem. In the following year, a method was proposed titled *Non Sub-sampled Contourlet Transform Based Feature Extraction Technique* for the purpose of for differentiating Glioma Grades Using MRI Images of the mentioned dataset [79]. In [78], the authors proposed a rectangular window based image cropping method to generalize brain neoplasm classification systems using the dataset of [11]. In [4], the authors employed **Capsule Networks** (also referred as CapsNets [56]) on the dataset proposed in [11] and showed a comparison between CNNs and CapsNets on the task of brain tumor classification problem, where the results were in favor of CapsNets. Our experiment is most similar with the one proposed in [50] where convolutional neural networks were used to segment brain tumors from MRI scans. But the difference is that, our model takes only the MRI scan as input where the method proposed in [50] requires the MRI scan as well as the masked file of that scan. This also makes our model more real life applicable. In [32] an automatic brain tumor stage classification was proposed by using probabilistic neural network where segmentation process was done by using K-means clustering. A method using convolutional nets to find region of interest, ie. the tumor, was proposed in [5] where the tumor was cropped by a rectangular shaped design and classified afterwards. Proposed method in [21] is also quite similar to our method, the difference being that their model classifies tumor types (edema/enhanced/tumor/necrosis/non-enhanced/tumor) from T1, T2, T1-enhanced, Flair plates and our model does semantic segmentation of the tumor. In the article [6] a brain tumor segmentation model was proposed which was claimed state of the art by the authors. The model was tasked with both patchwise segmentation and pixelwise segmentation.

Network	Decoder	Dataset	Mean Square Loss (train)	Accuracy (train)	Accuracy (validation)
Alexnet	FCD	Back view	11.4	69.7	70.1
Alexnet	Deconvolution	Back view	9.1	69.1	70.0
VGGnet	FCD	Back view	7.1	72.3	72.7
VGGnet	Deconvolution	Back view	6.1	71.0	72.1
FR-MRInet	FCD	Back view	<b>2.1</b>	<b>80.0</b>	<b>84.5</b>
FR-MRInet	Deconvolution	Back view	3.0	79.8	81.4

Table 3. : Table containing pixel wise accuracy on train and validation set of back view using Alexnet, VGGnet and FR-MRInet with both FCD and Deconvolutional decoder

### 3. IMAGE SEGMENTATION

Image segmentation is task where each pixel of an image is classified with the probability of its belonging to a object. This method is regarded highly in the computer vision society because of (1) its simplistic training procedure and (2) ability to detect multiple object. One of the many of training scheme goes as follows, an image is fed (preferably RGB) as input and the output is a grayscale image (for subject and object detection) with the background colored black and the subject in white. A dataset proposed in [34] is regarded as one of the benchmark for image segmentation task. Table 4 shows some of the train and test images.

Traditionally, an auto encoder is used where the encoder is tasked to encode the image and the decoder decodes that to the desired output image. Various imageNets such as Alexnet, VGGnet are some of the most used network as encoder where the final fully connected layer for classification is replaced with a decoder [10, 7].

In order to make the network end to end trainable the segmented image is considered as an numeric array where each of the element is considered as class value. Our experiment is highly influenced by Segnet [7] which is a deep convolutional encoder-decoder for image segmentation which uses VGG16 as the encoder. In [46], a similar method was proposed where the VGG16 was used as encoder and the authors proposed a deconvolution network and used that as the decoder. Another convolutional neural network for image segmentation was proposed in [38] where Alexnet and VGGnet were the encoder and fully connected layers were used to decode the image.

#### 3.1 Binary Segmentation

Binary image segmentation is a sub category of general image segmentation. In this case, there is only one type of object considered as subject in the image and the rest of the image is considered as background [18, 75]. This scheme is very common in medical images where the subject is the tumor and the rest is background. This also makes training process much simpler as the output is a 1D array of 0s and 1s. For our project, the masked image is a binary image where the tumor covered area is denoted by 1s and others by 0s.

### 4. MATERIALS

The benchmark datasets used in this experiment are from Brain Tumor Image Segmentation Challenge (BRATS) 2013 and 2015 [30, 43]. We used the BRATS 2013 dataset which contains 30 patient datasets and 50 synthetic datasets from 2013/2014 and BRATS 2015 dataset. Both of them contains hand labeled ground truths and four labels necrosis, edema, non-enhancing tumor, enhancing tumor and everything else. But in our experiments we mainly focus

on the segmentation task rather than the classification task. We use 2 common metrics known as pixel wise accuracy and IoU (Intersection over Union) for measuring and comparing the performance of our model with others.

The dataset which we exhaustively focus on was created by the authors of [11, 12] in 2015/6. The dataset consists of 30 T1-weighted contrast-enhanced image. 233 patients were randomly chosen from hospitals in China to collect the samples. The mentioned 3064 MRI scans consists of 708 slices of meningioma, 1426 slices of glioma and 930 slices of pituitary tumor. This dataset is organized in MATLAB's .mat data format and openly available here<sup>1</sup>.

#### 4.1 Modification

One of the shortcoming of the mentioned dataset is that all the images in the dataset contains any of the 3 tumor. But we want our model to be able to conclude when there is a no tumor present, which makes it more real life applicable. Since the actual dataset do not contain any image without any tumor we have manually inserted 10 tumor-less images. This has allowed us to train our model to be able not only locate a tumor but also give a probability of its possibility of being a real tumor.

While designing a model for the purpose of tumor detection it very important to provide a confidence value with output results. Otherwise, it will become very difficult if not impossible to make use of this model in real life.

### 5. METHODS

In this section, we are going to discuss the models we have experimented with as encoders and decoders. We are also going to discuss our proposed architecture and its effectiveness. We will also discuss some of the possible reasons of why our encoder performs better than other state of the art imagenets. The intuition to compare FCNs as decoder instead of either convolution or deconvolution was based on the evidence showed in [41, 28] where it was found that FCNs perform better as decoder than CNNs.

#### 5.1 Encoder Variants

In this subsection, we will discuss the types of encoders we have used in this experiment including the proposed convolutional encoder, FR-MRInet.

**5.1.1 Alexnet.** Alexnet was proposed in [31] as a image classifier for Oxford Flower Dataset [44, 44, 45]. The network had 14 layers including 5 convolutional layers, 3 max pool layers, 3 local response normalization layer and 2 FCNs with 4096 neurons each.

<sup>1</sup>[https://figshare.com/articles/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/brain_tumor_dataset/1512427)

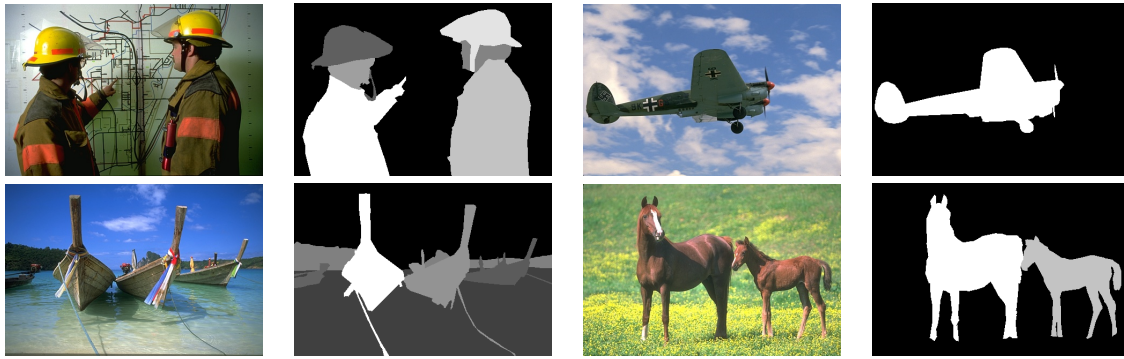


Table 4. : From each images Objects are segmented and are colored lighter in the segmented image while the rest of the image is colored black to represent the background.

Network	Pixel2Pixel Accuracy (train)	Pixel2Pixel Accuracy (validation)	IoU(train)	IoU(validation)
Alexnet + FCD	67.4%	67.7%	0.75	0.75
Alexnet + Deconvolution	67.4%	67.1 %	0.76	0.75
VGGnet + FCD	75.2%	76.2 %	0.81	0.82
VGGnet + Deconvolution	75.8%	76.0 %	0.81	0.82
[21] + FCD	74.8%	75.1 %	0.77	0.77
[21] + Deconvolution	74.9%	75.1 %	0.75	0.78
DeepNat + FCD	52.1%	50.2 %	0.68	0.70
DeepNat + Deconvolution	51.8%	49.8 %	0.71	0.71
[6] + FCD	82.7%	86.2 %	0.86	0.87
[6] + Deconvolution	83.2%	86.1 %	0.85	0.87
FR-MRInet+ FCD	<b>90.6%</b>	<b>91.2 %</b>	<b>0.9</b>	<b>0.9</b>
FR-MRInet+ Deconvolution	90.1%	90.8 %	<b>0.9</b>	<b>0.9</b>

Table 5. : Table contains Pixel2Pixel accuracy and IoU on train and validation set on BRATS 2013 dataset using different encoder variants with both FCD and Deconvolutional decoder.

For our experiment we removed all the FCNs of the original design and used that as an encoder for the task.

**5.1.2 VGGnet.** The VGG net was proposed in 2014 in the article [62]. It was initially targeted for classification of flower dataset [44, 44, 45]. Unlike Alexnet, the VGGnet is quite simple in design. It takes RGB image with the size of  $224 \times 224$  and passes them through a series of maxpool operation with stride 2 and convolution operation with filter size 3. It follows the following structure, 2 convolution layer 64 filters and a maxpool operation. Then the same 3 layers again but this time the number of filters are increased to 128. Following that, there are 3 convolution layers with 256 filters and a maxpool operation then the same set of layers twice with 512 filters each time. The original network had 2 FCNs with 4096 neurons before the output layer but for our task we replaced the FCNs for the purpose of using it as an encoder.

**5.1.3 FR-MRInet.** Figure 2 shows our proposed architecture of the encoding network. It is a 33 layer deep model consisting of only convolutional layer and merge layer that takes input images with resolution of  $256 \times 256 \times 3$ . Our design was motivated from "inception" imagenet[65, 64] which merged tensors after convolution operation with different filter sizes. We have used the similar methodology in several layers of FR-MRInet. From the input layer, we take the input and employ convolutional operation with filter

sizes 2, 3, and 5. Similar methodology was applied throughout the network where before each convolution act with any filter size, we used a conv operation with filter size of 1 to reduce the computational complexity. It was found in [37] that using a convolution operation with filter 1 reduces computational cost significantly with little to no affect on the performance. Table 11 shows a detailed description of the network. For example, lets consider layer 2 to 4.3. In layer 2 and 3, two consecutive conv operation happens with 128 filters each filter with size of 3. Following that, layer 4.1.a, 4.1.b and 4.1.c does convolution operation with filter size of 1 and 4.2.a, 4.2.b and 4.2.c does the same with filter size of 2, 3 and 5 respectively. Finally, in layer 4.3, they are merged together. The rest of network follows similar pattern with occasional change in the stride value to 2.

## 5.2 Decoder Variants

In this subsection, we will discuss two most common type of decoder which we have used in our experiment.

**5.2.1 Fully Connected Decoder.** Fully connected decoder(FCD) is a group of fully connected networks(FCN) used for decoding the encoded network before the final layer. Many types of FCDs have been proposed the researchers and they argue that in a system with encoder-decoder the decoder is more challenging area of interest [28]. In our experiment, we use 3 layers of FCN with 512, 6144 and

Network	Pixel2Pixel Accuracy (train)	Pixel2Pixel Accuracy (validation)	IoU(train)	IoU(validation)
Alexnet + FCD	68.4%	68.4%	0.73	0.74
Alexnet + Deconvolution	68.7%	68.4 %	0.71	0.71
VGGnet + FCD	73.9%	73.7 %	0.81	0.82
VGGnet + Deconvolution	73.0%	73.2 %	0.86	0.81
[21] + FCD	71.1%	71.1 %	0.73	0.73
[21] + Deconvolution	71.2%	71.1 %	0.79	0.78
DeepNat + FCD	58.8%	57.5 %	0.61	0.6
DeepNat + Deconvolution	58.8%	47.9 %	0.78	0.68
[6] + FCD	89.1%	89.6 %	0.89	0.88
[6] + Deconvolution	89.4%	90.1 %	0.81	0.84
FR-MRInet+ FCD	94.5%	<b>95.9%</b>	<b>0.91</b>	<b>0.93</b>
FR-MRInet+ Deconvolution	<b>94.8%</b>	93.7 %	0.89	0.9

Table 6. : Table contains Pixel2Pixel accuracy and IoU on train and validation set on BRATS 2015 dataset using different encoder variants with both FCD and Deconvolutional decoder.

Network	Pixel2Pixel Accuracy (validation)	Pixel2Pixel Accuracy (test)
Alexnet + FCD	70.3 %	72.6 %
Alexnet + Deconvolution	70.2 %	71.4 %
VGGnet + FCD	75.2 %	73.1 %
VGGnet + Deconvolution	74.1 %	71.3 %
[21] + FCD	71.5 %	70.7 %
[21] + Deconvolution	69.1 %	70.1 %
DeepNat + FCD	65.2 %	63.1 %
DeepNat + Deconvolution	67.5 %	61.8 %
[6] + FCD	78.2 %	80.1 %
[6] + Deconvolution	78.1 %	80.3 %
FR-MRInet+ FCD	<b>84.6 %</b>	<b>85.6 %</b>
FR-MRInet+ Deconvolution	83.9 %	83.8 %

Table 7. : Table containing pixel wise accuracy on validation and test set of total dataset using Alexnet, VGGnet and FR-MRInet with both FCD and Deconvolutional decoder

4096 neurons respectively (see figure 4). The most common FCDs consists of 2 FCN with 4096 neurons each. Our choice of selected number of neurons were found by fine-tuning for this problem.

**5.2.2 Deconvolutional Decoder.** Deconvolution is transpose operation of convolution operation which arguably can deconvolutionize a convolution operation [9]. Researchers have proposed various types of deconvolution layers and made arguments that since an encoder uses convolution to encode the image, a deconvolution operation can effectively decode the information [76]. We used 2 de-convolutional layer and between them were 2 upsampling layer(using nearest neighbor) with stride of 2.

## 6. IMPLEMENTATION DETAILS AND CONFIGURATION

In our experiment, we used a Tensorflow wrapper API called TFLearn using python 3.5 and Tensorflow version 1.5. The benchmark dataset used in this experiment was proposed in [11, 12] in 2015 and 2016 respectively which is the latest dataset for Brain

MRI tumors detection. Although we have made some changes to make it more appropriate to design a real life applicable method.

We quantify the performance of FR-MRInet on the dataset using our Tensorflow implementation. Instead of predicting the black and white mask we aimed to predict the whole input image using semantic segmentation. The dataset contains 3 types of tumor (*meningioma, glioma, pituitary*) with 3 types of views such as top, side and back. Table 10 shows some input, masked, output images along with predictions by the proposed network.

### 6.1 Training Data Construction

For the training data construction, we use *MATLAB*'s image processing tool [63] to generate the ground truth(GT) images. We take the grayscale MRI scan with the respective mask file of the image and use the *imfuse* function to create the GTs. The function takes 2 images and overlays them. We create the ground truth images with the help of gray scale mask images so that we can avoid using them as inputs which significantly limits the applicability of the model. Also it is much more comforting to receive an output highlighting

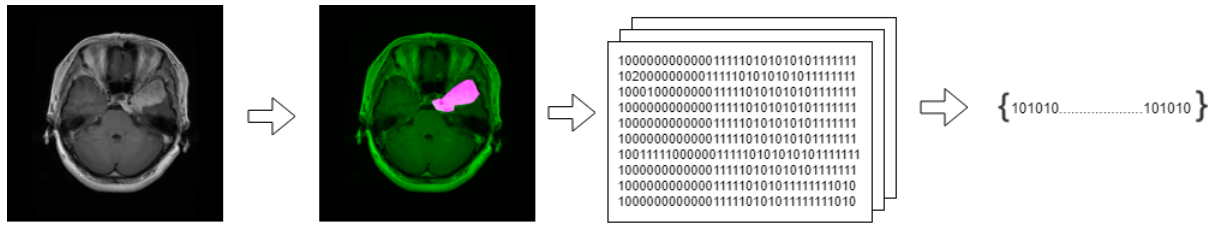


Fig. 1: Preprocessing of creating an end to end trainable image segmentation network where the image on the left is fed and input and the 1D array at the right is the expected output.

Network	Decoder	Dataset	Mean Square Loss (train)	Accuracy (train)	Accuracy (validation)
Alexnet	FCD	Total	12.3	70.2	70.3
Alexnet	Deconvolution	Total	16.5	10.2	70.2
VGGnet	FCD	Total	8.3	73.1	75.2
VGGnet	Deconvolution	Total	9.7	72.8	74.1
[21]	FCD	Total	8.4	83.1	71.5
[21]	Deconvolution	Total	10.2	84.1	69.1
DeepNat [70]	FCD	Total	9.7	83.7	65.2
DeepNat [70]	Deconvolution	Total	8.4	<b>88.1</b>	67.5
[6]	FCD	Total	15.3	78.3	78.2
[6]	Deconvolution	Total	16.3	78.4	78.1
FR-MRInet	FCD	Total	<b>3.9</b>	80.7	<b>84.6</b>
FR-MRInet	Deconvolution	Total	4.5	80.3	83.9

Table 8. : Table containing pixel wise accuracy on train and validation set of total dataset using Alexnet, VGGnet and FR-MRInet with both FCD and Deconvolutional decoder

the tumor instead of getting a gray scale mask. For each of the MRI scan containing a tumor is trained with confidence value 1 and the scans without any tumor is trained with 0. This allows us to choose threshold level manually to check if there is a tumor or not. In our experiment we chose 0.5 as the threshold value where any output containing confidence level below or equal to 0.5 is considered as non-tumor. Each training image is  $512 \times 512 \times 3$  resolution and the output is  $512 \times 512 \times 3$ . Our gpu could not process the whole  $512 \times 512 \times 3$  sized images so we improvised with sliding window. We divided each image into 4 equal non-overlapping sub-images sized  $256 \times 256 \times 3$  and fed them to the model sequentially. We took the outputs sequentially as well and merge all 4 equal non-overlapping sub-images to generate the final output image. The process is visually described in figure: 3. Instead of dividing merging the images in each iteration, we created a whole new dataset using this methodology where each input is a one quarter of the original image stored sequentially. The same method was applied to the outputs as well. But since accuracy on one quarter of a input doesn't make any sense, through out the rest of the paper, all the results are calculated on the whole output image. We used 5% data for testing and of the 95% training data 0.2% were used for validation.

## 6.2 Configuration

We implement the proposed network, FR-MRInet based on TFLearn API of Tensorflow framework. The *Adam* optimizer is employed for optimization, where initial learning rate is set to 0.0001. We used 3 learning rates for fine tuning the weights and biases of the network. The 1st 200 epochs were trained with *lr* 0.0001 and the

weights were stored. Then we employed transfer learning to load the weights and trained 200 epochs with learning rate 0.00001 and stored the weights again. Finally, we reloaded the the weights and trained with 0.000001 learning rate. We initialized the weights of the proposed convolution network using zero-mean Gaussian. We used 0.8 keep probability values in each dropout layer of fully connected decoding network. The network converges after approximate 600 epochs. We used the *Root-Mean-Square* error as the loss function, RELU as activation and a mini batch of size 10 was used to train and validate the network. The training takes approximately 5 days on a single Nvidia GTX 980ti over-clocked with 6GB memory.

## 7. RESULTS AND DISCUSSION

In this section, we discuss, analyze and provide possible reasons for certain phenomenons and outcomes. To compare the quantitative performance of FR-MRInet with different decoders, we used performance measures like IoU, local accuracy for measuring the percentage of pixels correctly classified in the validation set and global accuracy for measuring in the test set. We also show the loss value in both train and validation set as well. There are three types of images available in the dataset. We first only considered the scans of top views of the brain. Table 1 shows the Mean Square loss in the training phase for each imagenet with different decoder variants. It also shows the accuracy on validation and test set of each networks.

In table 5 and 6, we provide the performance of our model and other models in term of pixel to pixel accuracy and Intersection



Network	Pixel2Pixel Accuracy (test)	IoU
Alexnet + FCD	73.1 %	84%
Alexnet + Deconvolution	71.8 %	86%
VGGnet + FCD	73.3 %	87%
VGGnet + Deconvolution	71.4 %	87%
[21] + FCD	71.0 %	78%
[21] + Deconvolution	70.7 %	78%
DeepNat + FCD	63.2 %	69%
DeepNat + Deconvolution	61.7 %	68%
[6] + FCD	81.7 %	91%
[6] + Deconvolution	81.1 %	89 %
FR-MRInet+ FCD	<b>91.4 %</b>	<b>94 %</b>
FR-MRInet+ Deconvolution	88.6 %	93 %

Table 9. : Table contains accuracy on test set using different encoder variants with both FCD and Deconvolutional decoder after smoothening the output images with NCL.

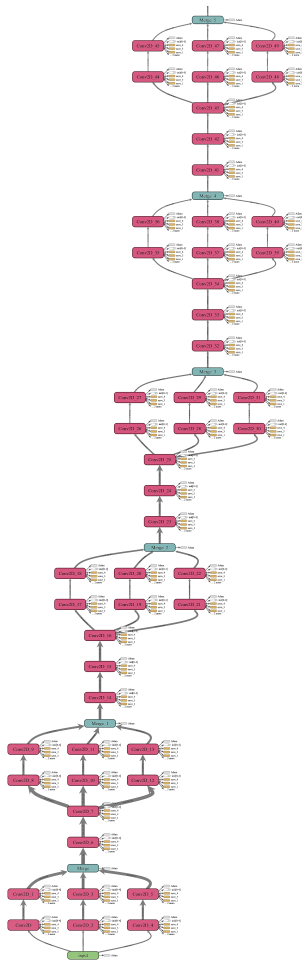


Fig. 2: Proposed convolutional Encoder

over Union (IoU). The first thing noticeable is that except for DeepNat every models performance on validation set slightly increased or remained same as the train which means the models were able to generalize well. We also observe that each individual model shows quite similar performance on both BRATS 2013 and 2015 dataset while sometimes the metrics show that model performed slightly better on 2015 dataset. This is because the 2015 dataset includes all the data from 2013 along with some new ones and also it contains more images than 2013 dataset. So due to the increased number of data the models were able to perform slightly better on the dataset of 2015. On both datasets, FR-MRInet were able to achieve the highest accuracy and IoU on both train set and test set.

In table 1, Alexnet obtained an accuracy of 69.2 in train set and 69.5 in validation set with a loss of 12.3 using the FCD. But the performance deteriorated with the use of de-convolutional decoder. Similar pattern was found using VGGnet and FR-MRInet. The VGGnet obtained a accuracy of 72.6 in train set and 73.1 in validation set employing FCD with a loss value of 8.3. The accuracy reduced to 72.3 (train) and 72.9 (validation) and the loss value increased to 9.7 when the de-convolutional decoder was used. The proposed network, FR-MRInet, achieved the the highest accuracy in both validation and test set with lowest loss value using FCD. The second best score was also achieved by FR-MRInet using the de-convolutional decoder. This not only resemblance that FR-MRInet is an efficient encoder but also shows that FCD performs generally better as decoder than de-convolutional decoder, aligning with similar conclusion drawn in [48, 28].

Table 3 and 2 illustrates the Mean Square loss value with accuracy on train and validation set. Following the pattern of table 1, Alexnet performs with lower loss values and higher accuracy with fully connected decoder compared with deconvolutional decoder. It obtained 12.3 loss value and accuracy of 70.2 on train and 71.6 on validation set in table 2 and 11.4 loss value and accuracy of 69.7 on train and 70.1 on validation set in table 3. VGGnet performed a bit higher accuracy scores on both sets with lower loss value. FR-MRInet was able to obtain the highest accuracy value on both sets with the lowest Mean Square loss value.

Finally, we trained the model on all three types of images and the results are displayed in table 8. Alexnet, VGGnet and FR-MRInet showed their better scores in all aspects comparing with table 1, 2 and 3. This is probably because when we used the whole dataset for training, the larger number of data enabled the network to learn more efficiently. Following the pattern of table 1, 2 and 3, Alexnet's performance was lower than VGGnet and the proposed network



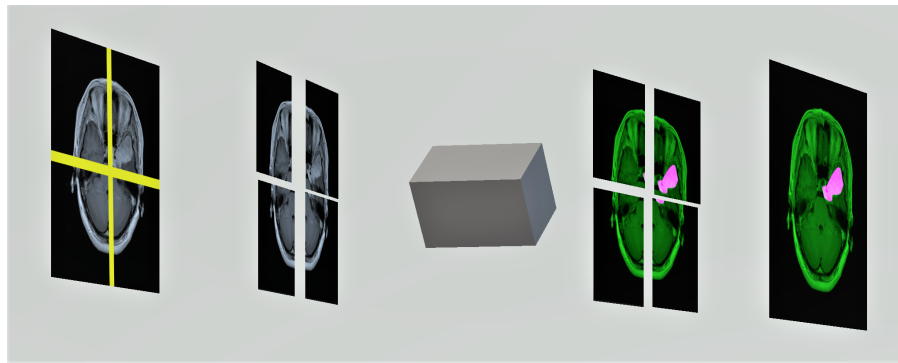


Fig. 3: Each of the input image is divided into 4 equal non-overlapping sub-images (highlighted by the yellow lines). Those 4 images are then fed to the network. The outputs are corresponding images of those 4 equal non-overlapping sub-images. They are finally merged together to generate the final output.

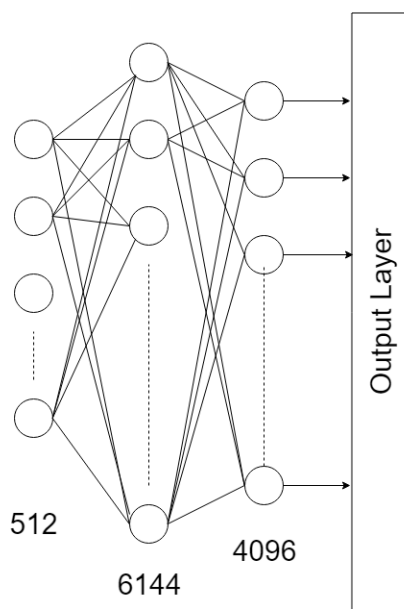


Fig. 4: Fully connected decoder used with each imagenet used in this experiment

was able to learn with the lowest loss value with highest accuracy on both sets. An important fact to note is that while FR-MRInet takes images with resolution of  $256 \times 256 \times 3$ , Alexnet and VGGnet are given input size of the original design which is  $227 \times 227 \times 3$  and  $224 \times 224 \times 3$ . It is quite fascinating to observe that FR-MRInet was able to show superior performance while working on a part of an image at each time. This is likely because of the usage of pooling layer in both Alexnet and VGGnet. Whenever any type of pooling operation occurs it reduces the feature map for computational benefit but also loses spatial information in the process. FR-MRInet avoids this by using a stride of 2 in various layers (see table 11) of the encoding architecture. Also Alexnet has fixed filter sizes of 11, 5 and 3 and VGGnet uses filter size 3 in all layers. The proposed network uses the inception-like module with filter sizes of 2, 3 and

5. The original design proposed in [65] used filter sizes of 2, 3 and 5 with a pooling layer. In our design, we removed the pooling layer so that the network is forced to choose information from the 3 chosen filter sizes that doesn't lose spatial information in the process. This also allows the network to choose the appropriate filter size automatically and assign weights and biases accordingly.

Many recent articles showed that they were able to classify the tumor type as well after detection [78, 4, 79]. In our research, we focused more on the visual aspect of the problem than the biological aspect. This is why we have separately observed results on scans with top, side and back view before merging them and drawing final conclusion instead of trying to classify the tumor type. Many researchers have proposed methods showing that their method can classify various tumor types from 2D MRI scans accurately. While their accuracy in classification was quite high, the methods often failed to generalize therefore performing poorly in real life [17]. Classification with 3D scans showed more promise but it suffers from the same problem [29]. Thus instead of trying to classify the tumor type [74], segmenting tumor from MRI scans was found more challenging and applicable in real world [50]. Also in real life, tumor classifications are made with the help of MRI scans plus the symptoms of the patients which are absent in datasets.

To further investigate in this matter, we asked an MD the following questions, (1) **Is it possible to classify brain tumors just using MRI scans?** (2) **How relevant are the symptoms for tumor classification with T1-weighted MRIs?** (3) **What is your professional risk assessment on diagnosing tumor types using only MRI scans?** (4) **How much do you think a project like this can help radiologists?** We present here some of the key lines of our consultation respectively and the whole answer to each question is given in the appendix. The answers to the questions are (1) "...MRIs are usually considered insufficient for definite diagnosis by itself, however, in some cases it can be used to confirm the diagnosis..." (2) "Very relevant. MRIs are most often evaluated with context, and the scan itself is almost certainly never initiated without context. It is the symptoms that lead to an MRI scan..." (3) ".....The MRI film itself is usually rather straightforward to evaluate, but cannot and should not be used as a definite diagnosis.....The exact type of brain tumor is generally not diagnosed with certainty without a biopsy to confirm the type.....". (4) "A ratio analyzer that detects and evaluates disproportionate areas of the brain. To a human eye, small disproportions can be missed, but a ratio analyzer

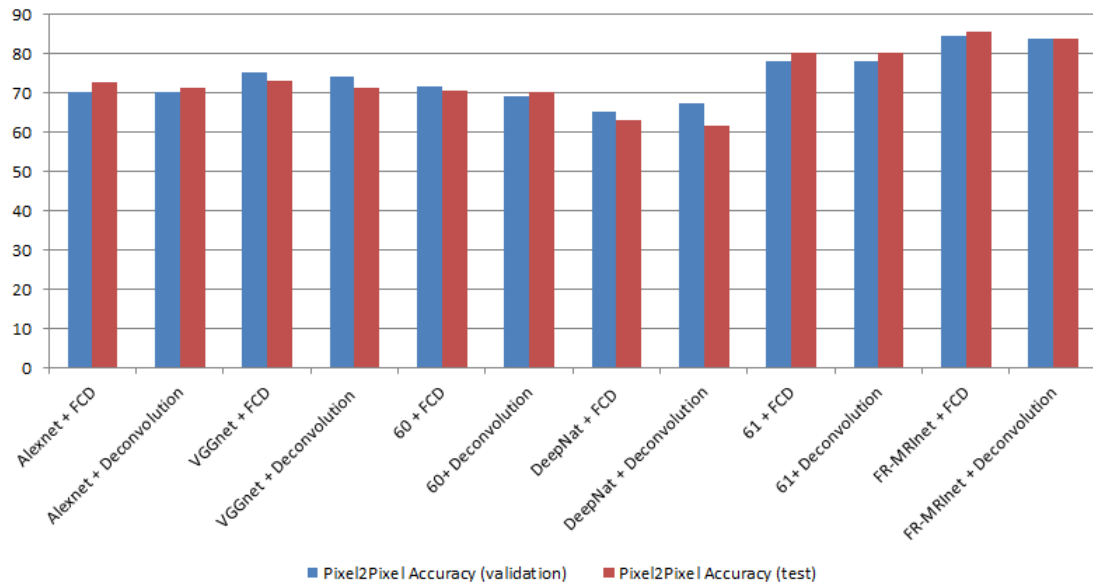


Fig. 5: Performance comparison on pixel wise accuracy metric of each of the model on validation and test set. The blue colored columns resembles performance on validation set and red the color illustrates performance on test set. The X axis represents accuracy ranging from 0 to 100%

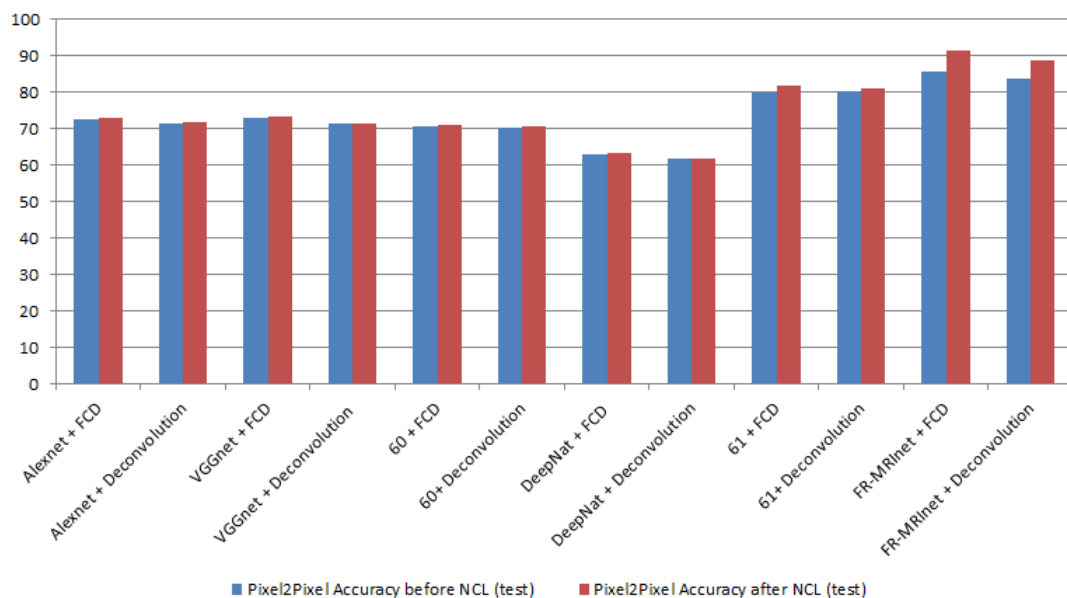


Fig. 6: Performance comparison on pixel wise accuracy metric of each of the model on test set before and after applying NCL. The blue colored columns resembles performance before using NCL and red the color illustrates performance afterwards. The X axis represents accuracy ranging from 0 to 100%

would point out irregularities.....if this project can be expanded in the future to also evaluate the malignancy of the tumor based on its appearance in the MRI, this can be of greater help to students...". From the comparison shown in table 8, we see that FR-MRInet was able to achieve the highest accuracy with FCD. The model from

[21] was able to achieve impressive accuracy in the train set but was unable to reproduce that performance in the test set. This suggests that the model has over-fitted probably due to the fact that the model was not originally designed for classification purpose. The model only has three convolutional layer (encoding section) with channel

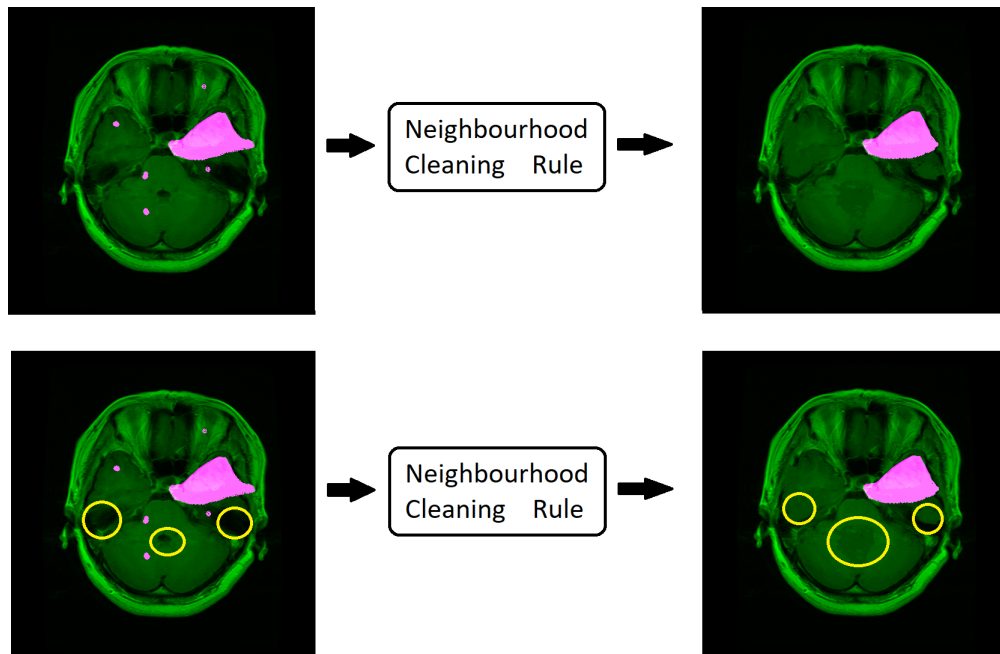


Fig. 7: After each of the model predicted their final output, each of the output contained noises on the final image. We used the Neighborhood Cleaning Rule (NCL) to identify anomalies on the image and replace them with its neighbor pixels. With yellow circles we denote some of the unintentional changes made by NCL.

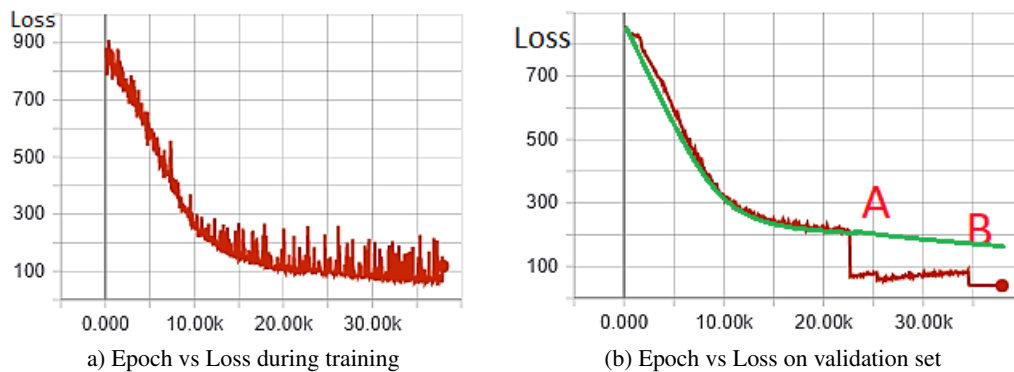


Fig. 8: Epoch vs Loss curves of training and validation set. Point A and B on curve (b) denotes where the learning rate was reduced.

sizes of 64, 160 and 224 which wasn't nearly enough to learn for it to generalize. The same conclusion can be drawn for DeepNat [70] as well, as DeepNat only uses 3 convolutional operation with some max-pooling and batch normalization in between. Both of these model's number of parameter pales in comparison with that of Alexnet, VGGnet and FR-MRInet which is probably why they performed poorly on the validation set. However, the model from [6] was not only able to generalize but also provide an impressive accuracy value as well. This was because of the deep nature of the architecture and the usage of softmax in the last layer for pixel wise classification. Unlike [6], our model do not use try to do pixel wise classification rather it tries to generate the tumor highlighted image straight from the input image. Although it may sound counter in-

tuitive, we used this method for output generation because in our case the output image is a simple RGB image where there are only 3 colors, green for the brain, pink for the tumor and black for background. Pixel wise classification are more efficient when there are more classes/objects are involved [7].

Finally, in table 7, pixel to pixel accuracy of each variant of encoder with both decoder types are illustrated. We have also tested our model against other DNN (Deep-neural Networks). Although they weren't meant for image segmentation, we discarded the last fully connected layer of the original network and replaced it with the 2 variant decoder used in this experiment. We used architecture from [21, 70] and [6] as well as some commonly used architecture like the AlexNet and VGGnet. While every variant was able to achieve

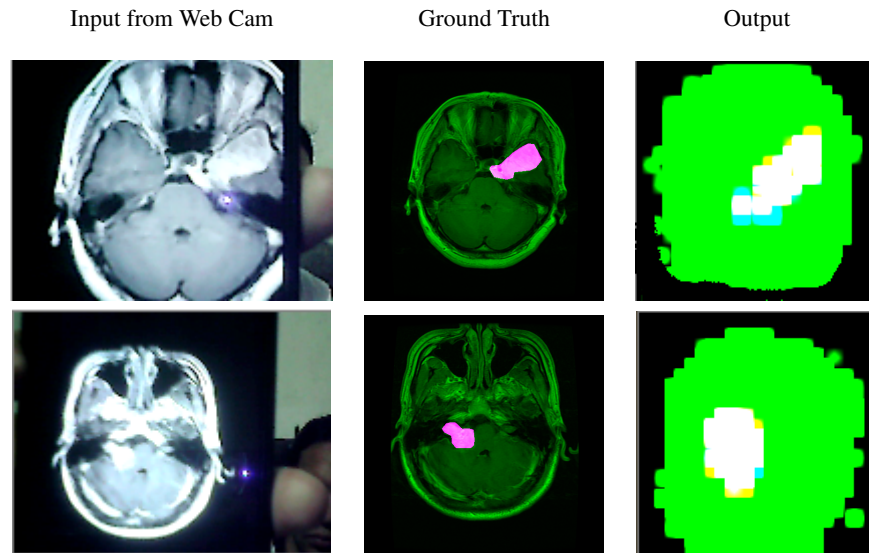


Fig. 9: Examples of Input images, masked image, Ground truth and predictions of FR-MRInet

over 60% detection accuracy, the proposed network with FCD was able to achieve the highest, 85.6% in the test set (ref table 8).

Figure 5 shows some outputs considering only the best performances of each model variant of table 7. From here we can observe that FR-MRInet clearly produces more satisfactory results comparing with other models. FR-MRInet, Alexnet and [6] were able to produce better score in the test set which means the models was able to generalize while DeepNat failed miserably. The highest accuracy was achieved by FR-MRInet with FCD in the test set which is just slightly higher than that of FR-MRInet with Deconvolutional decoder. The method from [6] performed very well which is due to the fact that it was originally designed for segmentation purpose, just not for brain MRI. It was able to show similar performance with both types of decoder which means the type of decoder has very little effect on the total performance. From this chart, we also see that generally encoder variants with FCD seems to perform slightly better when compared to deconvolutional decoder which aligns with the conclusion from [69, 48, 10] and [28].

Note that after each the model produced masks, each contained some anomalies. By anomaly, we mean some random small areas which are predicted as tumorous by mistake. This significantly reduces capability of each of the model. To overcome that we used a cleaning procedure called Neighborhood Cleaning Rule (NCL) [55]. The method was originally proposed to assist imbalanced data learning approaches [8] but we used its ability to detect anomaly in data-space to clean the output images. Figure 7 shows a visual example on how NCL smoothen the output for better efficiency. These anomalies occur as each of the model had positive loss value and none of them were able to converge to zero loss. Instead of further training we used this less computationally expensive method to address the problem. But as a by product, the smoothening process sometimes distort some of the pixel which were not anomalies (located by yellow circles in fig 7). Since we are more interested on the tumorous area, distortion of some not tumorous areas posses little harm. We used  $neighbour, k = 9$  (tuned as hyper parameter) as parameter for NCL. Table 9 shows the change in performance on the test set of each encoder with both types of decoder variants.

The highest performance gain was achieved by FR-MRInet with FCD which increased from 85.6% to 91.4%. The second highest was a 4.8% accuracy boost by FR-MRInet using deconvolutional decoder. All the rest of the model gains 1 to 2 % accuracy gain with the exception of DeepNat. With deconvolutional decoder DeepNat lost 0.1% accuracy when NCL was applied. The results are visualized in figure 6.

In table 10, we display some outputs generated from the testset by different methods. We see that VGGnet and [6] are quite successful at detecting the tumorous zone. Alexnet was also able to detect proper zones but it also contained too much noises which remained even after cleaning it using NCL. DeepNat was unable to learn any pattern which is why it always predicts the center area as region of interest (ROI). [6] and [21], both performed reasonably as they were able to detect the correct zone but often the detected area was much more wide spread than the actual ROI area and finally, FR-MRInet was able to generate the most satisfactory output among all of them.

In figure 8(a) and 8(b), the loss vs epoch is illustrated. We have used three learning rate ( $lr$ ) using transfer learning. The initial learning rate was 0.0001 and after 200 epochs we stopped and stored the weights. Then  $lr$  was reduced to 0.00001 for training the stored weights and after 200 epoch it was again reduced to 0.000001. We used this scheme as an alternative to the decayed learning rate value in order to reduce the hyper-parameter, *decay rate* from the training process. Although in figure 8(a), the loss value reduces gradually, in figure 8(b) the changes are more noticeable. There are two stages in figure 8(b) noted as *A* and *B* which shows are sudden change in the loss value due to lowering the  $lr$ . The green line in the figure 8(b) shows the predicted trajectory which points to a much higher loss function then the actual loss value. Therefore by using 3 different learning rates we were able to converge the model with less number of epochs. In our research, we have tested the model further lowering the  $lr$  but it was found not helpful. Thus we came to conclude that the mentioned learning rates already converged the model to its best and further training will only cause over fitting and worse performance.

20135cm

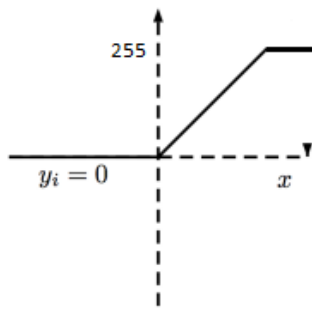


Fig. 10: Relu-RGB activation function

In several sections of the paper, we have mentioned that although the other models first generates the mask which is then overlaid on the original image to get the desired output (see table 10), our model generates the final output directly. We do that by having 3 channel as output in the final layer where each pixel value is between 0 to 255. To accomplish this task we propose a novel activation, called RGB-relu, which is a slight modification of the original Relu activation. We cut off the upper limit by 255 thus if  $X$  is greater than 255 the output will be 255 (see fig 10). Although limiting the upper value brings up the possibility of vanishing gradient problem, the model was able to overcome that as RGB-relu was only used once in the whole network.

Lastly, we have tested our model with an web camera to see if it can identify the tumor live, even though it wasn't designed with that intend. We placed a camera that feeds  $512 \times 512 \times 3$  images to the model and held a MRI image in front of the camera with a phone. In figure 9, we show some examples of input images with ground truth and the prediction. Although the outputs are not very precise, they are quite visually accurate in locating the tumor.

## 8. CONCLUSION AND FUTURE WORKS

In this paper, we have proposed an encoder and demonstrated its performance against state of the art imagenets with several decoder variants which addresses the problem of locating tumor from a MRI scan of a brain. This aims to help respected radiologists in their field and reduce some dependency of skills and expertise from the radiologist. The novelty of our encoder is its architectural design which performed well with fully connected networks as decoder. We have provided conclusive evidence by illustrating several experimental results which shows that FR-MRInet is a very effective encoder. We have also provided detailed analysis of the results as well as possible reasons for the proposed network's success. We have also used a non-over lapping sliding window technique to work on the whole  $512 \times 512 \times 3$  image instead of resizing them. We have also compared the effectiveness of fully connected decoder and deconvolutional decoder where the results in favor of FCD. We have also discussed effectiveness of this project with reputed professional in respective field.

We address the problem of detecting 3 types of brain tumor from T1 weighted MRI scans. We use pixel by pixel image classification to locate the tumor in the MRIs. Since MRI technology evolves rapidly, thus despite of having a lot of open sources brain MRI datasets, we chose the dataset from [11] which is one of the latest dataset designed for this particular problem. Using our encod-

ing model with Fully Connected Decoder, we were able to achieve 31.4 % pixel by pixel accuracy on finding the location. One of the shortcoming of this method is that the output image is generated at the size of  $512 \times 512 \times 3$  using a sliding window which was a choice made due to computational limitation. For that limitation, we could not compare the performance of the network with and without the sliding window. Many researchers have proposed methods that can classify the tumor types as well but we argue that without symptoms of the patient, the tumor classification is highly unrealistic for real life application. For that reason, we have consulted with a professional about the importance of the symptoms for identifying tumor types and how a project like ours can be useful in practical applications. Our dataset and codes are open sourced for future research and freely available here: <https://github.com/farshidrayhanui/FR-MRInet>.

## 9. ACKNOWLEDGMENT

Firstly, we would like thank very much Dr. Jonatan Nowakowski, MD for his involvement in the project. We would also like to acknowledge Mr. Hao-Tzu Wang ( haotzuw@gmail.com ) for conducting the interview and making it possible for us to collaborate with Dr Jonatan Nowakowski. We would finally like to thank Dr Rahul Savani (University of Liverpool), Dr Hakan Bilen (University of Edinburgh) and Professor Richard Hartley (Australian National University) for their advices and guidance towards preparing this manuscript.

## 10. REFERENCES

- [1] Cancer imaging archive.
- [2] Oasis-brains.
- [3] Open-access medical image repositories.
- [4] AFSHAR, P., MOHAMMADI, A., AND PLATANIOTIS, K. N. Brain tumor type classification via capsule networks. *arXiv preprint arXiv:1802.10200* (2018).
- [5] AHMED, K. B., HALL, L. O., GOLDFOF, D. B., LIU, R., AND GATENBY, R. A. Fine-tuning convolutional deep features for mri based brain tumor classification. In *Medical Imaging 2017: Computer-Aided Diagnosis* (2017), vol. 10134, International Society for Optics and Photonics, p. 101342E.
- [6] AKKUS, Z., GALIMZIANOVA, A., HOOGI, A., RUBIN, D. L., AND ERICKSON, B. J. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging* 30, 4 (2017), 449–459.
- [7] BADRINARAYANAN, V., KENDALL, A., AND CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 12 (2017), 2481–2495.
- [8] BEKKAR, M., AND ALITOUCHE, T. A. Imbalanced data learning approaches review. *International Journal of Data Mining & Knowledge Management Process* 3, 4 (2013), 15.
- [9] CAMPISI, P., AND EGIAZARIAN, K. *Blind image deconvolution: theory and applications*. CRC press, 2017.
- [10] CHEN, L.-C., PAPANDREOU, G., KOKKINOS, I., MURPHY, K., AND YUILLE, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 4 (2018), 834–848.



- [11] CHENG, J., HUANG, W., CAO, S., YANG, R., YANG, W., YUN, Z., WANG, Z., AND FENG, Q. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one* 10, 10 (2015), e0140381.
- [12] CHENG, J., YANG, W., HUANG, M., HUANG, W., JIANG, J., ZHOU, Y., YANG, R., ZHAO, J., FENG, Y., FENG, Q., ET AL. Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PloS one* 11, 6 (2016), e0157112.
- [13] DU, Y., WANG, J., WANG, X., CHEN, J., AND CHANG, H. Predicting drug-target interaction via wide and deep learning. In *Proceedings of the 2018 6th International Conference on Bioinformatics and Computational Biology* (2018), ACM, pp. 128–132.
- [14] GALL, J., AND LEMPITSKY, V. Class-specific hough forests for object detection. In *Decision forests for computer vision and medical image analysis*. Springer, 2013, pp. 143–157.
- [15] GIRSHICK, R. Fast r-cnn. *arXiv preprint arXiv:1504.08083* (2015).
- [16] GIRSHICK, R., DONAHUE, J., DARRELL, T., AND MALIK, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 580–587.
- [17] GORDILLO, N., MONTSENY, E., AND SOBREVILLA, P. State of the art survey on mri brain tumor segmentation. *Magnetic resonance imaging* 31, 8 (2013), 1426–1438.
- [18] GORELICK, L., VEKSLER, O., BOYKOV, Y., AND NIEUWENHUIS, C. Convexity shape prior for binary segmentation. *IEEE transactions on pattern analysis and machine intelligence* 39, 2 (2017), 258–271.
- [19] GUPTA, S., GIRSHICK, R., ARBELÁEZ, P., AND MALIK, J. Learning rich features from rgb-d images for object detection and segmentation. In *European Conference on Computer Vision* (2014), Springer, pp. 345–360.
- [20] HARADA, T., USHIKU, Y., YAMASHITA, Y., AND KUNIIYOSHI, Y. Discriminative spatial pyramid. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), IEEE, pp. 1617–1624.
- [21] HAVAEI, M., DAVY, A., WARDE-FARLEY, D., BIARD, A., COURVILLE, A., BENGIO, Y., PAL, C., JODOIN, P.-M., AND LAROCHELLE, H. Brain tumor segmentation with deep neural networks. *Medical image analysis* 35 (2017), 18–31.
- [22] HE, K., GKIOXARI, G., DOLLÁR, P., AND GIRSHICK, R. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on* (2017), IEEE, pp. 2980–2988.
- [23] HOFFMAN, J., GUADARRAMA, S., TZENG, E. S., HU, R., DONAHUE, J., GIRSHICK, R., DARRELL, T., AND SAENKO, K. Lsda: Large scale detection through adaptation. In *Advances in Neural Information Processing Systems* (2014), pp. 3536–3544.
- [24] JAVED, U., RIAZ, M. M., GHAFOOR, A., AND CHEEMA, T. A. Mri brain classification using texture features, fuzzy weighting and support vector machine. *Progress In Electromagnetics Research* 53 (2013), 73–88.
- [25] JIANG, J., WU, Y., HUANG, M., YANG, W., CHEN, W., AND FENG, Q. 3d brain tumor segmentation in multimodal mr images based on learning population-and patient-specific feature sets. *Computerized Medical Imaging and Graphics* 37, 7-8 (2013), 512–521.
- [26] JIANG, Y., YUAN, J., AND YU, G. Randomized spatial partition for scene recognition. In *Computer Vision–ECCV 2012*. Springer, 2012, pp. 730–743.
- [27] JOHN, P., ET AL. Brain tumor classification using wavelet and texture based neural network. *International Journal of Scientific & Engineering Research* 3, 10 (2012), 1–7.
- [28] KAMPFFMEYER, M., SALBERG, A.-B., AND JENSSEN, R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on* (2016), IEEE, pp. 680–688.
- [29] KHOTANLOU, H., COLLIOT, O., ATIF, J., AND BLOCH, I. 3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy sets and systems* 160, 10 (2009), 1457–1473.
- [30] KISTLER, M., BONARETTI, S., PFAHRER, M., NIKLAUS, R., AND BÜCHLER, P. The virtual skeleton database: An open access repository for biomedical research and collaboration. *J Med Internet Res* 15, 11 (Nov 2013), e245.
- [31] KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.
- [32] LAVANYADEVI, R., MACHAKOWSALYA, M., NIVETHITHA, J., AND KUMAR, A. N. Brain tumor classification and segmentation in mri images using pnn. In *Electrical, Instrumentation and Communication Engineering (ICEICE), 2017 IEEE International Conference on* (2017), IEEE, pp. 1–6.
- [33] LAZEBNIK, S., SCHMID, C., AND PONCE, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer vision and pattern recognition, 2006 IEEE computer society conference on* (2006), vol. 2, IEEE, pp. 2169–2178.
- [34] LI, H., CAI, J., NGUYEN, T. N. A., AND ZHENG, J. A benchmark for semantic image segmentation. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on* (2013), IEEE, pp. 1–6.
- [35] LIENHART, R., KURANOV, A., AND PISAREVSKY, V. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *Joint Pattern Recognition Symposium* (2003), Springer, pp. 297–304.
- [36] LIENHART, R., AND MAYDT, J. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on* (2002), vol. 1, IEEE, pp. I–I.
- [37] LIN, M., CHEN, Q., AND YAN, S. Network in network. *arXiv preprint arXiv:1312.4400* (2013).
- [38] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.
- [39] MALISIEWICZ, T., GUPTA, A., AND EFROS, A. A. Ensemble of exemplar-svms for object detection and beyond. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (2011), IEEE, pp. 89–96.
- [40] MANINIS, K.-K., PONT-TUSET, J., ARBELÁEZ, P., AND VAN GOOL, L. Convolutional oriented boundaries: From image segmentation to high-level tasks. *IEEE transactions on*

*pattern analysis and machine intelligence* 40, 4 (2018), 819–833.

- [41] MARCU, A., AND LEORDEANU, M. Dual local-global contextual pathways for recognition in aerial imagery. *arXiv preprint arXiv:1605.05462* (2016).
- [42] MARMANIS, D., SCHINDLER, K., WEGNER, J. D., GALLIANI, S., DATCU, M., AND STILLA, U. Classification with an edge: improving semantic image segmentation with boundary detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 135 (2018), 158–172.
- [43] MENZE, B., JAKAB, A., BAUER, S., KALPATHY-CRAMER, J., FARAHANI, K., KIRBY, J., BURREN, Y., PORZ, N., SLOTBOOM, J., WIEST, R., LANCZI, L., GERSTNER, E., WEBER, M.-A., ARBEL, T., AVANTS, B., AYACHE, N., BUENDIA, P., COLLINS, L., CORDIER, N., CORSO, J., CRIMINISI, A., DAS, T., DELINGETTE, H., DEMIRALP, C., DURST, C., DOJAT, M., DOYLE, S., FESTA, J., FORBES, F., GEREMIA, E., GLOCKER, B., GOLLAND, P., GUO, X., HAMAMCI, A., IFTEKHARUDDIN, K., JENA, R., JOHN, N., KONUKOGLU, E., LASHKARI, D., ANTONIO MARIZ, J., MEIER, R., PEREIRA, S., PRECUP, D., PRICE, S. J., RIKLIN-RAVIV, T., REZA, S., RYAN, M., SCHWARTZ, L., SHIN, H.-C., SHOTTON, J., SILVA, C., SOUSA, N., SUBBANNA, N., SZEKELY, G., TAYLOR, T., THOMAS, O., TUSTISON, N., UNAL, G., VASSEUR, F., WINTERMARK, M., HYE YE, D., ZHAO, L., ZHAO, B., ZIKIC, D., PRASTAWA, M., REYES, M., AND VAN LEEMPUT, K. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* (2014), 33.
- [44] NILSBACK, M.-E., AND ZISSERMAN, A. Delving into the whorl of flower segmentation. In *Proceedings of the British Machine Vision Conference* (2007), vol. 1, pp. 570–579.
- [45] NILSBACK, M.-E., AND ZISSERMAN, A. Automated flower classification over a large number of classes. In *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing* (Dec 2008).
- [46] NOH, H., HONG, S., AND HAN, B. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 1520–1528.
- [47] OISHI, K., FARIA, A., JIANG, H., LI, X., AKHTER, K., ZHANG, J., HSU, J. T., MILLER, M. I., VAN ZIJL, P. C., ALBERT, M., ET AL. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and alzheimer's disease participants. *Neuroimage* 46, 2 (2009), 486–499.
- [48] OKTAY, O., FERRANTE, E., KAMNITSAS, K., HEINRICH, M., BAI, W., CABALLERO, J., COOK, S. A., DE MARVAO, A., DAWES, T., OREGAN, D. P., ET AL. Anatomically constrained neural networks (acnns): Application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging* 37, 2 (2018), 384–395.
- [49] PATIL, S., AND UDUPI, V. A computer aided diagnostic system for classification of brain tumors using texture features and probabilistic neural network. *Int J Comput Sci Eng Inf Technol Res* 3 (2013), 61–66.
- [50] PEREIRA, S., PINTO, A., ALVES, V., AND SILVA, C. A. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE transactions on medical imaging* 35, 5 (2016), 1240–1251.
- [51] RAYHAN, F., AHMED, S., MAHBUB, A., JANI, M., SHATABDA, S., FARID, D. M., ET AL. Cusboost: Cluster-based under-sampling with boosting for imbalanced classification. *2nd International Conference on Computational Systems and Information Technology for Sustainable Solution* (2017).
- [52] RAYHAN, F., AHMED, S., MAHBUB, A., JANI, M., SHATABDA, S., FARID, D. M., RAHMAN, C. M., ET AL. Meboost: Mixing estimators with boosting for imbalanced data classification. *11th international Conference on Software, Knowledge, Information Management and Applications (SKIMA)* (2017).
- [53] RAYHAN, F., AHMED, S., MOUSAVIAN, Z., FARID, D. M., AND SHATABDA, S. Frnet-dti: Convolutional neural networks for drug-target interaction. *arXiv preprint arXiv:1806.07174* (2018).
- [54] REN, S., HE, K., GIRSHICK, R., AND SUN, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (2015), pp. 91–99.
- [55] RIQUELME, J., RUIZ, R., RODRÍGUEZ, D., AND MORENO, J. Finding defective modules from highly unbalanced datasets. *Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos 2*, 1 (2008), 67–74.
- [56] SABOUR, S., FROSST, N., AND HINTON, G. E. Dynamic routing between capsules. In *Neural Information Processing Systems* (2017), pp. 3859–3869.
- [57] SEIFFERT, C., KHOSHGOFTAAR, T. M., VAN HULSE, J., AND NAPOLITANO, A. Rusboost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 40, 1 (2010), 185–197.
- [58] SELVARAJ, H., SELVI, S. T., SELVATHI, D., AND GEWALI, L. Brain mri slices classification using least squares support vector machine. *International Journal of Intelligent Computing in Medical Sciences & Image Processing* 1, 1 (2007), 21–33.
- [59] SHARMA, G., AND JURIE, F. Learning discriminative spatial representation for image classification. In *BMVC 2011-British Machine Vision Conference* (2011), BMVA Press, pp. 1–11.
- [60] SHEIKH, Y., AND SHAH, M. Bayesian modeling of dynamic scenes for object detection. *IEEE transactions on pattern analysis and machine intelligence* 27, 11 (2005), 1778–1792.
- [61] SIMONYAN, K., VEDALDI, A., AND ZISSERMAN, A. Learning local feature descriptors using convex optimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 8 (2014), 1573–1585.
- [62] SIMONYAN, K., AND ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [63] SIOGKAS, G. *Visual Media Processing Using Matlab Beginner's Guide*. Packt Publishing Ltd, 2013.
- [64] SZEGEDY, C., LIU, W., JIA, Y., Sermanet, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCHE, V., RABINOVICH, A., ET AL. Going deeper with convolutions. *Cvpr*.
- [65] SZEGEDY, C., REED, S., ERHAN, D., ANGUELOV, D., AND IOFFE, S. Scalable, high-quality object detection. *arXiv preprint arXiv:1412.1441* (2014).



- [66] VEDALDI, A., GULSHAN, V., VARMA, M., AND ZISSERMAN, A. Multiple kernels for object detection. In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 606–613.
- [67] VIOLA, P., AND JONES, M. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (2001), vol. 1, IEEE, pp. I–I.
- [68] VIOLA, P., AND JONES, M. Fast and robust classification using asymmetric adaboost and a detector cascade. In *Advances in neural information processing systems* (2002), pp. 1311–1318.
- [69] VOLPI, M., AND TUIA, D. Dense semantic labeling of sub-decimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 55, 2 (2017), 881–893.
- [70] WACHINGER, C., REUTER, M., AND KLEIN, T. Deepnat: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage* (2017).
- [71] WRIGHT, J., YANG, A. Y., GANESH, A., SASTRY, S. S., AND MA, Y. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence* 31, 2 (2009), 210–227.
- [72] WU, Y., YANG, W., JIANG, J., LI, S., FENG, Q., AND CHEN, W. Semi-automatic segmentation of brain tumors using population and individual information. *Journal of digital imaging* 26, 4 (2013), 786–796.
- [73] YOUNG, J. Spatial pyramid match kernels for brain image classification. In *Pattern Recognition in Neuroimaging (PRNI), 2016 International Workshop on* (2016), IEEE, pp. 1–4.
- [74] ZACHARAKI, E. I., WANG, S., CHAWLA, S., SOO YOO, D., WOLF, R., MELHEM, E. R., AND DAVATZIKOS, C. Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme. *Magnetic resonance in medicine* 62, 6 (2009), 1609–1618.
- [75] ZAITOUN, N. M., AND AQEL, M. J. Survey on image segmentation techniques. *Procedia Computer Science* 65 (2015), 797–806.
- [76] ZEILER, M. D., KRISHNAN, D., TAYLOR, G. W., AND FERGUS, R. Deconvolutional networks. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (2010), IEEE, pp. 2528–2535.
- [77] ZHANG, Y., BRADY, M., AND SMITH, S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging* 20, 1 (2001), 45–57.
- [78] ZIA, R., AKHTAR, P., AND AZIZ, A. A new rectangular window based image cropping method for generalization of brain neoplasm classification systems. *International Journal of Imaging Systems and Technology* (2017).
- [79] ZIA, R., AKHTAR, P., AZIZ, A., SHAH, M. A., ET AL. Non sub-sampled contourlet transform based feature extraction technique for differentiating glioma grades using mri images. In *Australasian Joint Conference on Artificial Intelligence* (2017), Springer, pp. 289–300.
- [80] ZULPE, N., AND PAWAR, V. Glcm textural features for brain tumor classification. *IJCSI International Journal of Computer Science Issues* 9, 3 (2012), 354–359.

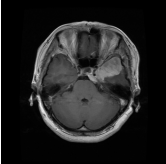
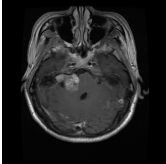
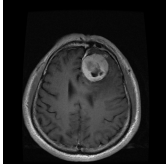
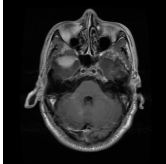
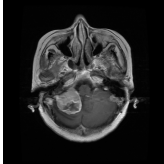
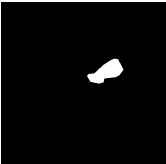
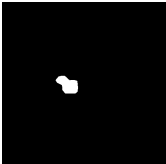
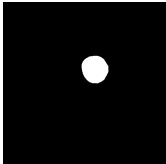
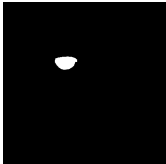

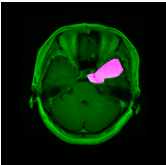
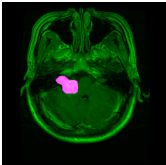
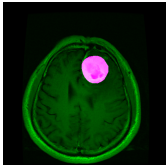
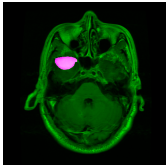
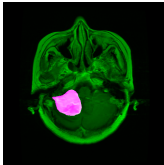
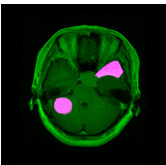
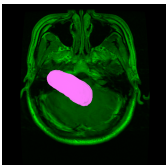
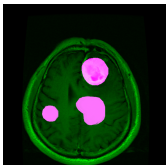
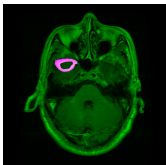
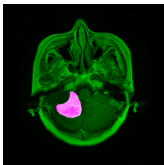
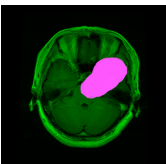
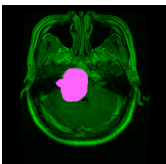
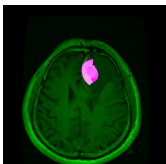
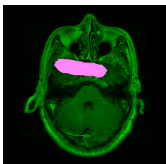
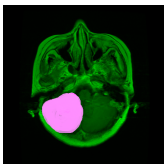
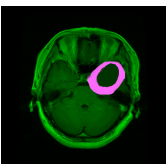
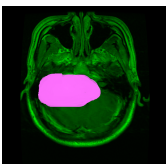
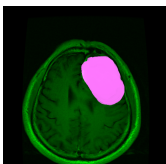
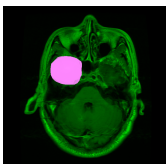
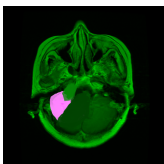
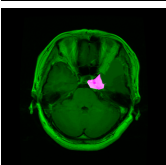
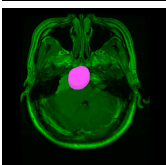
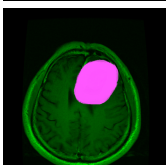
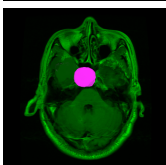
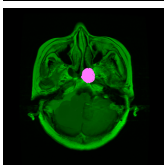
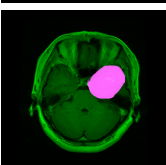
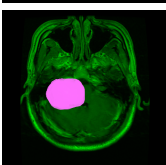
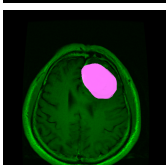
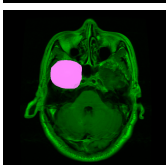
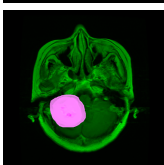
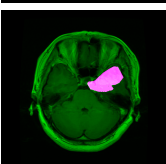
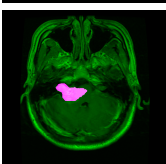
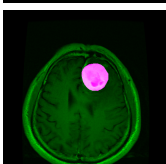
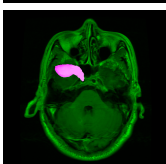
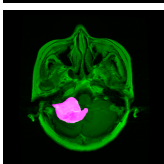
Input Image					
Mask					
Ground Truth					
AlexNet					
Vggnet					
[21]					
DeepNat [70]					
[6]					
FR-MRInet					

Table 10. : This table illustrates some example output of on the test set of different models. The ground truth image is created using the original image overlay-ed by the Masks. Except for FR-MRInet, each and every other model generates the mask and the final output is created by overlaying them. FR-MRInet directly produces the output image skipping the need of intermediate post processing step of the other methods.

Layer	Operation	Filter number	Filter size	Stride	Remarks
1.1.a	Convolution	16	1	1	
1.1.b	Convolution	16	1	1	
1.1.c	Convolution	16	1	1	
1.2.a	Convolution	64	2	1	
1.2.b	Convolution	64	3	1	
1.2.c	Convolution	64	5	1	
1.3	Merge	192			Merging 1.2.{a + b + c}
2	Convolution	128	3	1	
3	Convolution	128	3	1	
4.1.a	Convolution	16	1	1	
4.1.b	Convolution	16	1	1	
4.1.c	Convolution	16	1	1	
4.2.a	Convolution	64	2	2	
4.2.b	Convolution	64	3	2	
4.2.c	Convolution	64	5	2	
4.3	Merge	192			Merging 4.2.{a + b + c}
5	Convolution	128	3	1	
6	Convolution	128	3	1	
7	Convolution	128	3	2	
8.1.a	Convolution	32	1	1	
8.1.b	Convolution	32	1	1	
8.1.c	Convolution	32	1	1	
8.2.a	Convolution	96	2	1	
8.2.b	Convolution	96	3	1	
8.2.c	Convolution	96	5	1	
8.3	Merge	288			Merging 8.2.{a + b + c}
9	Convolution	256	3	1	
10	Convolution	256	3	1	
11	Convolution	256	3	2	
12.1.a	Convolution	32	1	1	
12.1.b	Convolution	32	1	1	
12.1.c	Convolution	32	1	1	
12.2.a	Convolution	128	2	1	
12.2.b	Convolution	128	3	1	
12.2.c	Convolution	128	5	1	
12.3	Merge	384			Merging 12.2.{a + b + c}
13	Convolution	256	3	1	
14	Convolution	256	3	1	
15	Convolution	256	3	2	
16.1.a	Convolution	32	1	1	
16.1.b	Convolution	32	1	1	
16.1.c	Convolution	32	1	1	
16.2.a	Convolution	128	2	1	
16.2.b	Convolution	128	3	1	
16.2.c	Convolution	128	5	1	
16.3	Merge	384			Merging 16.2.{a + b + c}
17	Convolution	512	3	1	
18	Convolution	512	3	1	
19	Convolution	512	3	1	
20.1.a	Convolution	32	1	1	
20.1.b	Convolution	32	1	1	
20.1.c	Convolution	32	1	1	
20.2.a	Convolution	128	2	1	
20.2.b	Convolution	128	3	1	
20.2.c	Convolution	128	5	1	
20.3	Merge	384			Merging 20.2.{a + b + c}
21	Convolution	3	1	1	3 channel for RGB image output

Table 11. : Detailed description of proposed encoding model, FR-MRInet.