Brain Lesion Segmentation using Convolutional Neural Networks

Divyesh Vanjare (ddv245), Priyanka Ashok Sapkal (pas571), Shiva Sanketh Ramagiri Mathad (srm714)

***Abstract*—Multiple Sclerosis (MS) is the most common neurological disability, that often leads to severe physical and cognitive problems. Lesion detection is important not only for studying stages of MS but also for taking the best treatment decisions. Magnetic Resonance Images (MRI) are mainly used for detecting such lesions. However, manual segmentation of lesions using MRI scans is very tedious and time consuming. Also, the results of such manual segmentation depend largely on the machines used for MRI scans and the knowledge of experts. Hence, there is a need for a reliable and automatic lesion segmentation technique. However, the format of the MRI scans as well as the irregularity in the shape and size of brain lesions makes it a challenging task for automatic segmentation.**

In this report, we present three different convolutional neural networks (CNNs) that fit best for brain lesion segmentation. The first model is 3D U-Net, which takes 3D MRI input and consist of an auto-encoder and auto-decoder. The second model is V-Net, which is similar to U-Net and is specially designed for volumetric medical image segmentation. The third model is 3D High Resolution Representation Learning Network (3D HR-Net) that consists of high resolution and low resolution layers connected in parallel to maintain the input resolution for better accuracy. We analyse and implement these models and present our results for each model tested on MSSEG challenge data.

***Index Terms*—multiple sclerosis, brain lesion segmentation, convolutional neural network, 3D U-Net, V-Net, 3D HR-Net.**

I. Introduction

Multiple sclerosis (MS) is an autoimmune-mediated disorder that affects the central nervous system (CNS) and often leads to severe physical, cognitive and neurological problems in young adults [1]. The National MS Society estimates nearly 1 million people in the US living with MS [7]. Fig 2. shows a slice of human brain with two lesions. Detection of such lesions is necessary for studying the stage of MS and predicting the best treatment for MS patients. Such lesions can be visualised using MRI imaging technique. However, MRI images are 3D in nature. Manual segmentation of lesions using MRI images takes up valuable time and is dependent on the expert's knowledge. Moreover, the MRI images obtained by different scanning machines can be different. Therefore, it is important to obtain reliable automatic lesion segmentation. This is a challenging task as the size, shape and location of the lesion is highly variable. Now-a-days, many machine learning techniques have emerged that can solve this problem with less amount of time and produce accurate results. However, many of these techniques rely on models that take 2D image as input. For this purpose, the 3D MRI volumes are sliced along each axis to produce 2D images, that are in turn fed to the model.

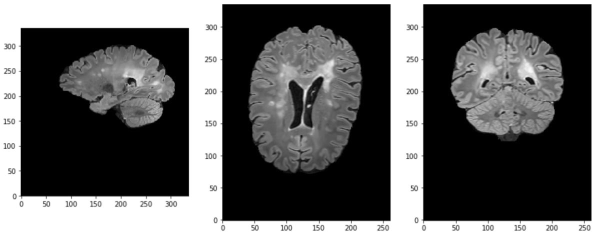

Fig. 1. Slices of brain image along sagittal, transverse and coronal plane respectively.

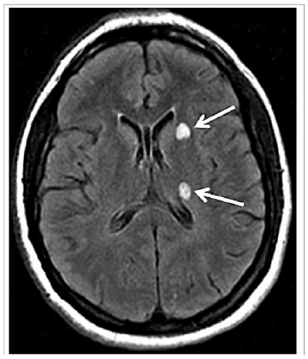

Fig. 2. Slice of brain with lesion.

II. Literature Survey

The MICCAI Challenge Proceedings on Multiple Sclerosis Lesions Segmentation mentions various approaches to segment lesions. One of the approaches is training a deep convolutional neural network on image slices along different axes [9]. This could find good locality of the lesions and good dice scores of 0.64 but not provide accurate crisp boundaries.

Results of using Hybrid Artificial Neural Networks on 3D images [8] are relatively accurate for a short training and testing time. It mentions using pre-processing and post-processing on a single layer neural network to segment lesions. Their proposed method incorporates training on finding spatial-based and intensity-based features.

We propose using best of both approaches mentioned above, training deeper neural networks on 3D images for potentially getting even better dice scores and accurate lesion boundaries.

III. Proposed Methods

Diagnostic and interventional imagery usually consists of 3D images. Converting the images in 2D may lead to losing out on important spatial information necessary to locate the lesion accurately. Therefore, processing the images in 3D is necessary not only to reduce overhead of pre-processing but also to retain their spatial context.

A. 3D U-Net

The network is similar to the 2D U-Net architecture, which has a series of down convolutions and up convolutions and produces an output of same shape as the input. The network we are using takes 3D images as input and processes them with 3D versions of convolution, max pooling, up-sampling, and batch normalization functions. Due to memory constraints, we resized the original image and its mask to 48x100x100 before feeding into the network.

Similar to 2D U-Net, our network has a max pooling path and an up-sampling path. In the max pooling path, each of the three layers has a convolution with kernel size 5 followed by a batch norm, a rectified linear unit (ReLu), and a 3D max pooling with kernel size 2 with stride of 2 in each dimension. In the synthesis path, each layer consists of an up-sampling unit with kernel size 2 and stride of 2 in each dimension, followed by one convolution with kernel size 5, a batch norm and a ReLu unit. Skip connections from layers of same resolution in the max pooling path provide the high-resolution features to the up-sampling path. In the output layer, a convolution of kernel size 1 followed by a sigmoid activation function is used. This reduces the number of channels to 1 depicting lesion probability labels on the voxels.

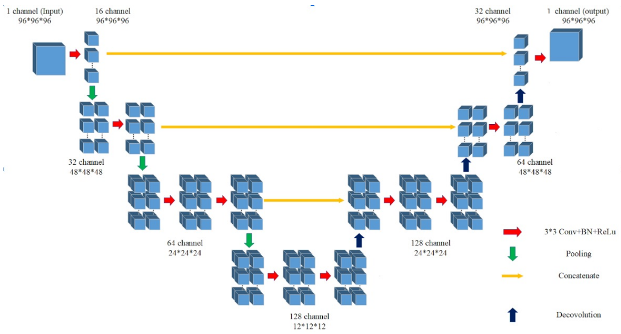

Fig. 3. 3D U-Net Architecture

B. V-Net

Due to memory and GPU constraints, instead of passing the entire volume to the network, we converted the 3D input images into blocks of size 32. We refer to this as Block Processing. Here, we first normalize the image and its corresponding mask, then convert the 3D image into smaller cubes of size 32. We use this data as the input to our V-Net model. Converting the image into blocks also allowed us to send only those blocks to the network that contained lesions. We checked the presence of lesion in a block of image by taking the sum of the corresponding mask of the image block. Only those blocks with a sum greater than 0 were passed to the network.

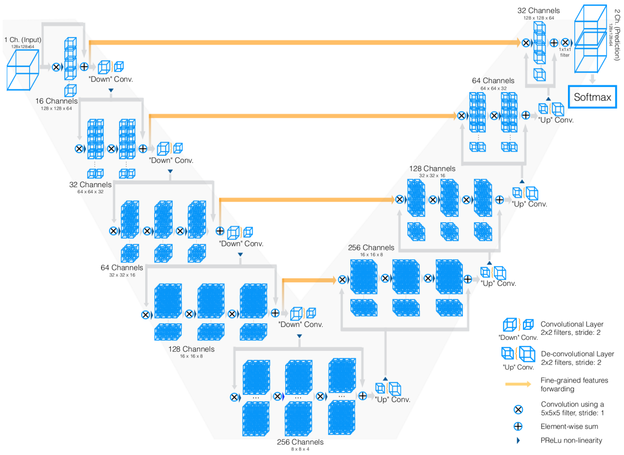

Fig. 4. V-Net Architecture

The V-Net model [2] consist of a compression and decompression path, similar to a U-Net. Here, we performed 3D convolutions to extract features as well as to down-sample and up-sample the image as it passes the network. We used kernels of size 5 to perform feature extraction and kernel of size 2 with stride 2, to perform down-sampling and up-sampling. As the image passes through the network, the size of the image decreases by a factor of 2 and the number of channels increases by a factor of 2. In addition to this, there are additional convolutional layers at each layer. Similar to U-Net, the early extracted features from the left part of the V-Net are fed to the right part of the V-Net to reduce the vanishing gradient problem. V-Net also enables residual learning. At each layer, the input is passed to the sequence of convolutional layers as well as added to the output of the last convolutional layer. We used 3D batch normalization and have applied Parametric Rectified linear unit (PRelu) non-linearities throughout the network. Finally, the last convolutional layer i.e. the output layer uses a kernel of size 1 and a sigmoid activation function to produces an output mask of size 32.

C. 3D HR-Net

Deep High-Resolution Representation Learning [3] also called the HR-Net mainly focuses on learning high-resolution representations. Our previous approaches recover high-resolution representations from low-resolution representation, instead HR-Net maintains high-resolution throughout the process. The HR-Net architecture is a 2D architecture intended for Human Pose Estimation. We developed a 3D version of the HR-Net architecture by inflating the 2D HR-Net architecture and correspondingly tuning the channels for our requirement as shown in Fig 5. HR-Net architecture consists of three layers, top most layer maintains the original resolution, middle layer is down-sampled using two-strided convolution to lower the resolution once and the third layer is further down-sampled. A regular 3D convolution unit uses a kernel size of 3. HR-Net has three key features-

- Connect high-to-low resolution convolutions in parallel.
- Maintain high-resolution representations throughout the process.

- Fuse low resolution and high resolution features repeatedly to strengthen and obtain better feature representations.

These three key features are essential for better data representation.

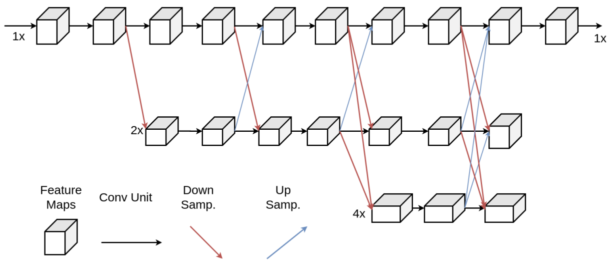

Fig. 5. 3D HR-Net Architecture

IV. Experimental Setup

A. Dataset

To perform analysis on all the three models, we used the MSSEG challenge data. It consists of 15 sample volumes collected from 3 different scanners. The dataset consisted of different MRI sequences such as 3D FLAIR, 3D T1, 3D T2 etc. along with 7 manual segmentation masks and one consensus mask for each patient. For our analysis, we made use of the 3D FLAIR sequence and its corresponding consensus mask. Out of the 15 samples, we used 3 samples per machine i.e. a total of 9 samples for training part, one sample per machine i.e. a total of 3 samples for each of the validation and testing part.

B. Training

1) 3D U-Net: The 3D U-Net has 386,353 parameters and was trained with Adam optimizer and optimized on dice loss for 500 epochs. We used 12 samples for training with leave-one-out cross validation to avoid over-fitting. We used adaptive learning with initial rate of 0.1 and step size of 0.1

2) V-Net: The V-Net model has 45 million trainable parameters. The original V-Net model [2] was trained for 48 hours on a dataset of 30 samples. However, due to resource restrictions, we could run our model for approximately 4-5 hours with 200 epochs and batch size of 16 on training dataset of 9 samples using Nvidia P4 GPU. We used Adam optimizer and adaptive learning rate with an initial rate of 0.1 and a step size of 0.1.

3) 3D HR-Net: 3D HR-Net model has 84 million trainable parameters. We use block processing and each sample in our training data is broken down into 16x16x16 blocks due to memory and computational power constraints. And this is fed as input to our 3D HR-Net model. We use 200 Epochs, BCE Loss function, Adam Optimizer and adaptive learning with initial rate of 0.1 and step size of 0.1. The training took approximately 11 hours on Nvidia Tesla P40 GPU.

V. Results

A. 3D U-Net

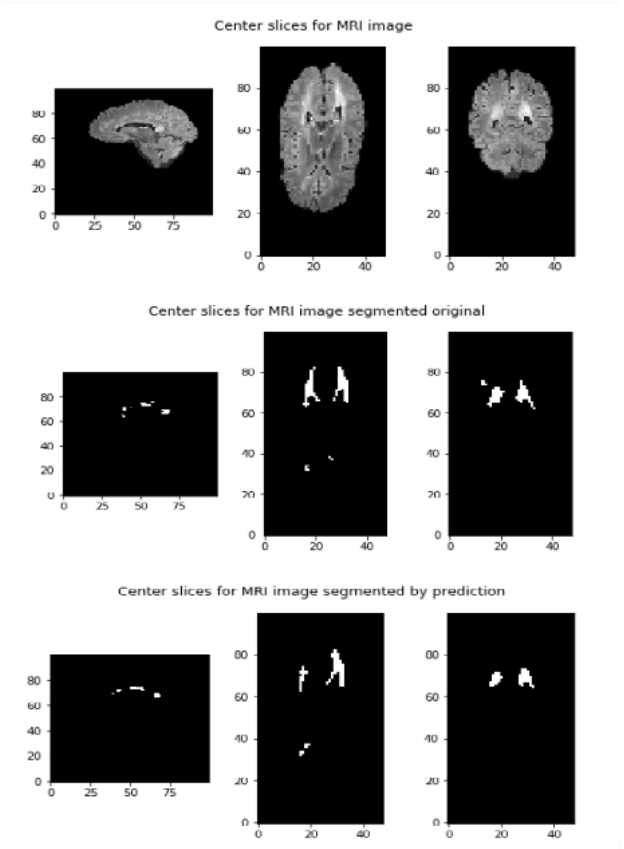
Center slices for MRI image
Center slices for MRI image segmented original
Center slices for MRI image segmented by prediction
0
20
40
60
80
25
50
75

Fig. 6. Slices of input image block, original mask and predicted mask for 3D U-Net

Fig 6. depicts a slice of 3D FLAIR image passed as an input to the 3D U-Net model at the top. The actual mask segmented by medical expert in the middle and the mask predicted by the 3D U-Net model at the bottom.

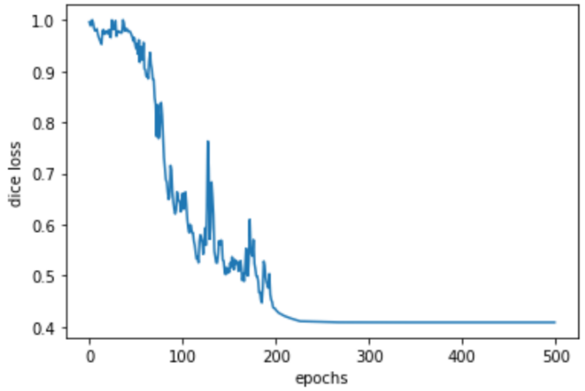

Fig. 7. Graph depicting training dice loss vs epochs for 3D U-Net

B. V-Net

Figure 8(a). is an example of image block passed as an input to the V-Net model. Figure 8(b). is of the true mask of the image block. Figure 8(c). shows the predicted mask obtained using V-Net.

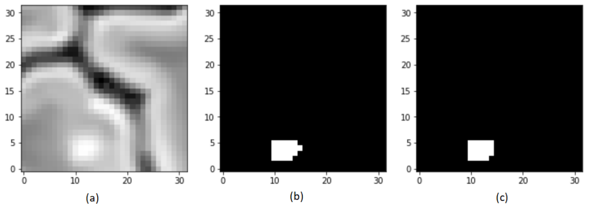

Fig. 8. A slice of MRI image block, its true mask and mask predicted by V-Net model respectively.

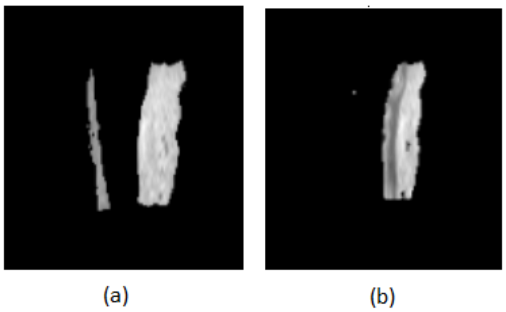

Fig. 9. True segmentation and segmentation predicted by the V-Net model for the MRI image respectively.

Figure 9(a). shows the true segmentation and Figure 9(b). shows the predicted segmentation. Note that the segmentation is not 100% accurate since the model was trained for a limited number of epochs and time due to limited number of GPUs and memory. The model accuracy can be increased with more resources and training samples.

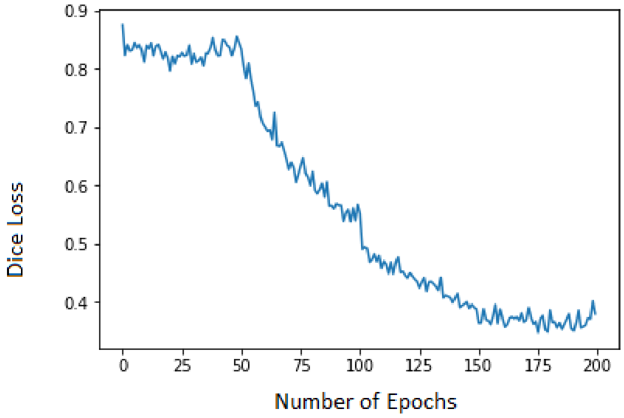

Fig. 10. Graph depicting number of epochs vs dice loss for V-Net

Figure 10. shows the graph of dice loss versus the number of epochs used during training the V-Net. The dice loss is approximately 0.9 at the beginning of the training part and reduces to approximately 0.3 by the end of 200 epochs. The dice loss can further decrease by running the model for more number of epochs and by adjusting the learning rate.

C. 3D HR-Net

In Figure 11., at the left side is the slice of the image block passed as an input to the 3D HR-Net model. The middle image is of the true mask of the image block. The last image shows the predicted mask obtained using 3d-HR Net. In Figure 12., we can see the graph of BCE Loss vs Epoch. The loss is very high initially, i.e. approximately 5 and reduces to approximately 0.1 by the end of 200 epochs.

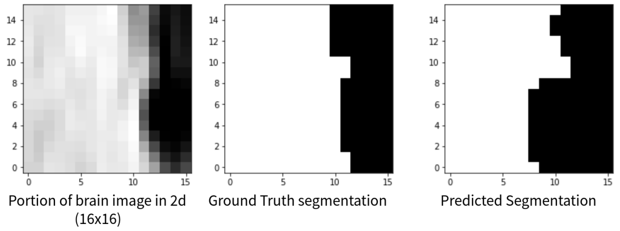
14
12
10
8
6
4
2
0
0
5
10
15
Portion of brain image in 2d (16x16)
Ground Truth segmentation
Predicted Segmentation

Fig. 11. A slice of image block, its true mask and its predicted mask respectively.

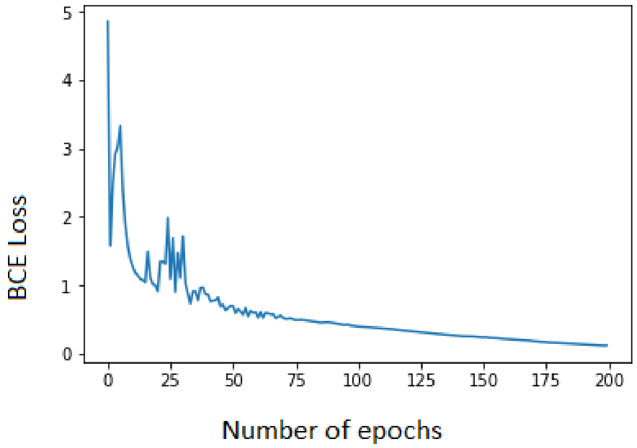

Fig. 12. Graph depicting BCE loss vs number of Epochs

1) Architecture Comparison: In the table below, we compare the Dice coefficient of the three developed architectures by comparing them on the basis of number of trainable parameters, type of input the model takes and total number of epochs trained for.

Architecture Comparison			
Features	**3D-UNet**	**V-Net**	**3D-HRNet**
Trainable Parameters	386353	45 million	84 million
Model Input	Input Resized to 48x100x100	Blocks of 32x32x32	Blocks of 16x16x16
Total trained Epochs	500	200	200
Dice coefficient	**0.64**	**0.41**	**0.32**

We compare Dice coefficient of our architectures with some of the pre-existing models

- Voxel-wise comparison of multi-channel Magnetic Resonance Images[4] - 0.4231
- Random Forest for Multiple Sclerosis Lesion Segmentation[5] - 0.63

We can observe from the above that our 3D U-Net outperforms both the pre-existing models. Our V-Net and 3D HR-Net models were not trained completely due to memory and computational power limitations, both were trained for 200 epochs. From the progression of loss vs epochs graphs of these 2 models, we can observe that, if trained for significantly more epochs, our V-Net and 3D HR Net model would outperform the pre-existing models.

VI. Discussion and Summary

We studied and developed different 3D models to perform Brain Lesion Segmentation and analysed their results. Future scope of the project includes training VNet and 3D HRNet on the entire dataset using suitable resources and compare the results with the existing 3D-UNet architecture.

Appendix

The source code for all the three models presented in this report can be found at the link given below.
https://github.com/shivasanketh-rm/Brain-Lesion-Segmentation-using-Convolutional-Neural-Networks
Project task division can be found in the presentation.

Acknowledgment

We would like to thank Zhipeng Fan for his guidance throughout the course of the project.

References

[1] Ghasemi, Nazem et al. “Multiple Sclerosis: Pathogenesis, Symptoms, Diagnoses and Cell-Based Therapy.” Cell journal vol. 19,1 (2017): 1-10. doi:10.22074/cellj.2016.4867 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5241505/

[2] Fausto Milletari 1 , Nassir Navab 1 , 2 , Seyed-Ahmad Ahmadi 3: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation

[3] ke, Sun Xiao, Bin Liu, Dong Wang, Jingdong. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. https://arxiv.org/abs/1902.09212

[4] Karpate, Y., Commowick, O., Barillot, C.: Robust Detection of Multiple Sclerosis Lesions from Intensity-Normalized Multi-Channel MRI (Feb 2015)

[5] Vera-Olmos, H. Melero and N. Malpica: Random Forest for Multiple Sclerosis Lesion Segmentation

[6] MSSEG Challenge Dataset: https://portal.fli-iam.irisa.fr/msseg-challenge/overview

[7] National Multiple Sclerosis Society Facts: https://www.nationalmssociety.org/About-the-Society/MS-Prevalence

[8] Amirreza Mahbod, Chunliang Wang, Orjan Smedby, Automatic Multiple Sclerosis Lesion Segmentation Using Hybrid Artificial Neural Networks, In: Proceedings of the 1st MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure — MICCAI-MSSEG, O. Commowick, F. Cervenansky, and R. Ameli (Eds), pp. 29, 2016.

[9] Richard McKinley, Tom Gundersen, Franca Wagner, Andrew Chan, Roland Wiest and Mauricio Reyes, Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation: application to white-matter lesion segmentation in multiple sclerosis, In: Proceedings of the 1st MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure — MICCAI-MSSEG, O. Commowick, F. Cervenansky, and R. Ameli (Eds), pp. 37, 2016.