INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1857)
Available online at: https://www.ijariit.com

# Segmentation approach for brain tumor detection

Shreyash Gupta
shreyashgupta.cs18@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Shreyas S.
shreyass.cs18@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Lalitha V. P.
lalithavp@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

## ABSTRACT

***The task of detecting tumours is an important one as it can lead to many lives being saved as nearly 20k lives are lost every year due to brain tumour. The task is challenging as in few cases brain tumours are not detected in early stages. With the help of artificial intelligence this gap can be bridged. The task of detecting regions of tumour can be classified into the popular computer vision task of image segmentation. Given the actual MRI scans and the masks representing the region of tumour, the machine learning models can learn the functions which map the regions to presence or non presence of tumour. The models are trained using a pair of MRI images and corresponding masks and the given an MRI image, the model should detect the region of tumour if present. This paper explores different state of the art architectures to perform the task of image segmentation.***

***Keywords***—*Computer Vision, Artificial Intelligence, Segmentation, Machine Learning.*

## 1. INTRODUCTION

Recent interest in augmented reality wearables, home-automation devices, and self-driving vehicles has created a strong need for semantic-segmentation or visual scene understanding algorithms. Semantic segmentation, or image segmentation, is the task of clustering parts of an image together which belong to the same object class. In this article the application of semantic segmentation in the medical field is explored.The task of segmenting tumour scans is difficult considering that the mask represents the region of the brain which has tumour cells. Since it is the medical field any slight changes in the region predicted can lead to wrong surgeries and can have harmful effects. However with the help of a segmentation model, the medical experts can narrow down the regions of presence of tumour as well as it helps in earlier detection of tumour. Tumour detection has been a challenge even for the medical professionals and with the help of a segmentation model, such challenges can be faced with certain ease.Artificial intelligence has been making huge strides in the field of medical domain by finding its application in various tasks such as cancer detection, syndrome prevention and helping disabled in their day to day activities. Similarly in the task of tumour detection, machine learning algorithms like segmentation have been found to produce optimal results and assist doctors in early detection of diseases and tumours. In fact machine learning algorithms have also been used to detect diseases in crops as well.The problem statement is to detect tumour regions in the MRI images of brain scan using deep learning based segmentation models. The models are trained used a pair of MRI images and corresponding masks and the given an MRI image, the model should detect region of tumour if present.Artificial Intelligence has found its applications in various sectors of the medical field with some interesting applications like detecting cancerous cells, assisting disabled people and many more. In this article one such application which is detecting tumour cells in the brain is studied. The problem statement is to detect the region where tumour is present in the brain given the MRI images of the brain. The above task is popular in the computer vision field as the segmentation task. The use of current state of the art models is presented in the next section and three such models are used for training and testing.

## 2. LITERATURE REVIEW

The current paper provides a comparison of three different neural network models for the task of segmentation of MRI Images. One of the models is inspired by the paper[8] which introduces the architecture called SegNet. The novelty of SegNet lies in the manner in which the decoder up samples the lower resolution feature maps. The model provides a reasonably good performance on the benchmark dataset and is space efficient as well. In paper[12], the authors have extended the concept of Faster RCNN to propose Mask RCNN by adding a branch for predicting segmentation masks in each region of interest in parallel with the existing branch for classification and bounding box regression. The paper[1] proposes a multiscale architecture called Bidirectional Pyramid Network which takes the shape of a pyramid in the sense that information flows from bottom to top and from left to right in a systematic manner. In paper[2] an attention based approach is presented which combines multi scale predictions and this approach is found to improve the accuracy score as well as detection time. In paper[4] the author aims to unify the instance and semantic segmentation at the architecture level by designing a single network for both the tasks. In paper[7], a new mobile architecture called MobileNetv2 is proposed. In the architecture instead of using regular convolution blocks, inverted residual blocks are used which are a combination of 2D convolution layers and ReLU units. The paper[9] introduces the architecture of LinkNet. This has an encoder decoder architecture but the output of the encoder block is not only passed on to the next block but also to the corresponding decoder block. By performing such connections, the spatial information can be recovered and used by the decoder during the upsampling operations. In paper[15] the authors present a novel neural network based architecture called UNet used for the task of medical image segmentation. The authors have modified the fully convolutional networks to introduce an upsampling layer and a downsampling layer. Skip connections are introduced such that the upsampling layer can make use of the features learnt by the downsampling layer. The paper[5] is an extension of the base UNet model. The hypothesis behind the architecture is that the model can more effectively capture fine-grained details of the foreground objects when high-resolution feature maps from the encoder network are gradually enriched prior to fusion with the corresponding semantically rich feature maps from the decoder network. The paper[3] is an extension of the UNet++ model. The proposed model takes advantage of full scale skip connections and deep supervision. The model was used to test against benchmark medical datasets. The paper[10] proposes a novel idea for generating the pixel wise masks. The idea is to let low-resolution images go through the full semantic perception network first for a coarse prediction map. Then a cascade feature fusion unit and cascade label guidance strategy are proposed to integrate medium and high resolution features, which refine the coarse semantic map gradually. The paper[11] proposes a novel architecture called the pyramid scene parsing network to embed difficult scenery context features in an FCN based pixel prediction framework. The authors have also presented an effective optimization strategy for ResNet based networks. Paper[13] presents an architecture which has been optimized for fast inference and high accuracy. It consists of bottleneck blocks which are made up of convolution layers. Among the recent architectures which are giving promising results are the Generative Adversarial Networks. GAN's have found their applications in segmentation as well with paper[6] which proposes CGAN's which are suitable for image to image translation tasks, in which conditions are imposed on the input image and the corresponding output image is generated.

## 3. DATASET

The dataset contains brain MRI images together with manual FLAIR abnormality segmentation masks. The images were obtained from The Cancer Imaging Archive (TCIA). The dataset contains 3929 images and the corresponding masks which indicate the position of tumour in the brain. The mask is a binary image with the pixels representing the tumour in the brain. The images consist of both tumorous and non tumorous samples. Of the 3929 images there are 1373 images which are of scans having no tumour and the rest 2556 are of images which contain brain tumour.

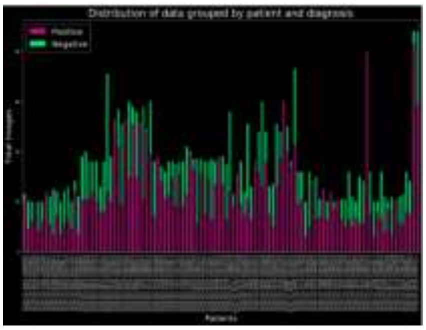


**Figure 1. Distribution of scans**

Both MRI scan images and their corresponding mask are in tiff format and. Brain MRI scan images have 3 channels ie RGB whereas the masks have a single channel. Figure 1 shows MRI scans corresponding to a patient. Figure 2 shows MRI brain scan, its corresponding mask and its canny.

Image augmentation is a technique of altering the existing data to create some more data for the model training process. In other words, it is the process of artificially expanding the available dataset for training a deep learning model. Due to limited dataset and to make the trained model more robust, image augmentation is also performed.It helps in preventing overfitting during the training stage Image Augmentation is performed using the Image Data Generator class of the Tensorflow library. The various augmentations performed are described in the following section. Width shift range and Height shift range is applied to the images in which the images are shifted in the horizontal and the vertical scale.

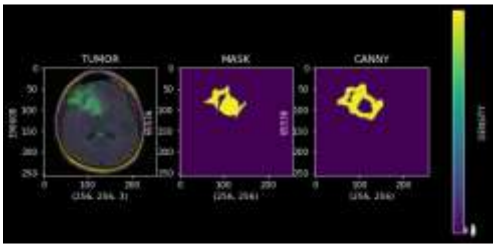


**Figure 2. Mask and Canny for a scan**

Shear transformation is applied to the images controlled through the shear range. Few images are randomly zoomed in and out. Few images are flipped horizontally and vertically. Rotation range is applied to the images in which images are rotated randomly. Below Images shows augmented brain MRI scans and their corresponding masks. Figure 3 shows augmented MRI scans and Figure 4 shows corresponding masks.

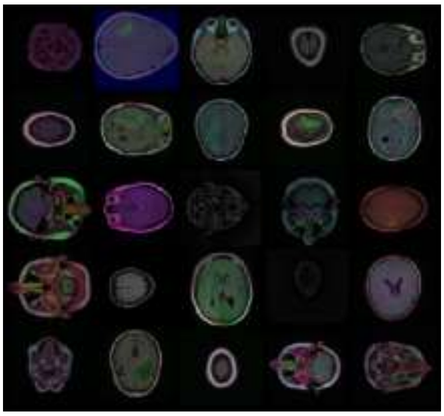

**Figure 3. Augmented Scan Images**

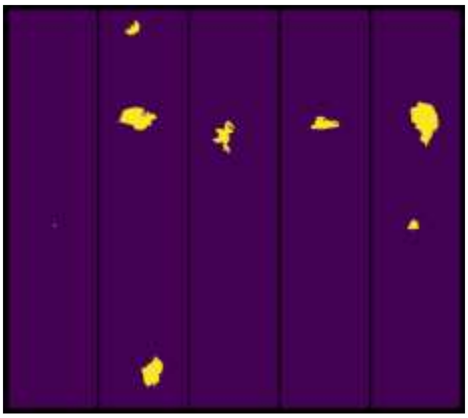

**Figure 4. Corresponding masks for scans in Fig3.**

 

## 4. MODEL ARCHITECTURE

A. UNet

The network is made up of two parts, an upsampling layer and a downsampling layer. Internally both the layers constitute convolutional blocks. Each block consists of 3x3 unpadded convolutions, followed by a ReLU activation function and a 2x2 max pooling operation. The difference between the downsampling and the upsampling layers is that at every downsampling layer the number of feature channels is doubled whereas in the upsampling layer the number of feature channels is reduced in half. At the final layer a 1x1 convolution is used to map each 64 component feature vector to the number of output classes which in the case of segmentation of MRI scans is 2. The entire network has 23 convolution layers in total.

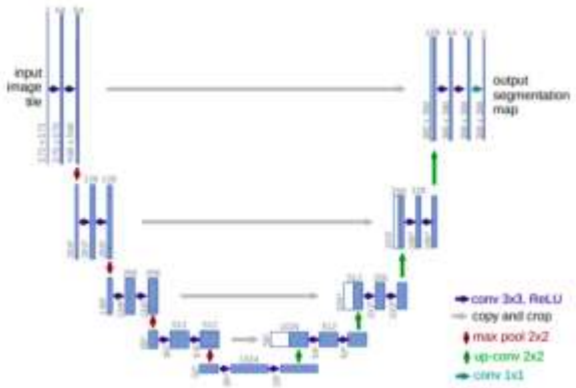


**Figure 5. UNet Architecture**

B. SegNet

SegNet has an encoder decoder architecture. The feature vectors are encoded by the encoding part and after the long scale vectors are resolved by the decoding network, a pixel wise classification layer is present which gives pixel wise output which highlights the presence or absence of brain tumour. The encoder consists of 13 convolution layers which correspond to the first 13 layers of the VGG16 network. This allows the use of pretrained weights of larger datasets. In a typical VGG16 network the convolution layers are followed by the fully connected layers.

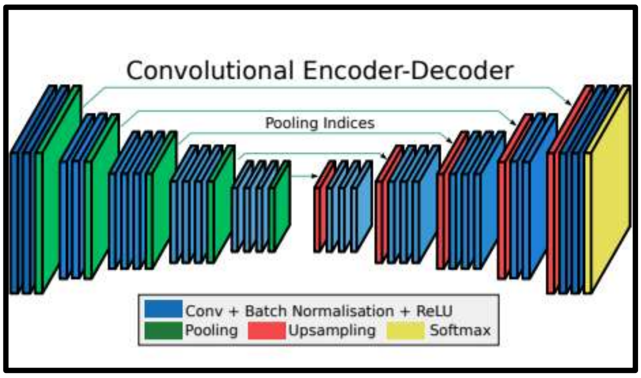


**Figure 6. SegNet Architecture**

In the SegNet architecture, the fully connected layers are discarded to retain the high resolution feature vectors obtained from the encoder. This reduced the number of parameters in the encoder significantly. Each encoder in the encoder network performs a convolution step, followed by batch normalization, ReLU activation and a max pooling operation. SegNet provides flexibility in the sense that the encoder network can remain fixed and multiple decoder networks can be tried and experimented to observe which backbone network can produce the highest results.

C. ResNeXt50

In this architecture, we use UNet with a pretrained ResNeXt50 backbone. ResNeXt50 is a simple, highly modularized network architecture for image classification. The network is constructed by repeating a building block

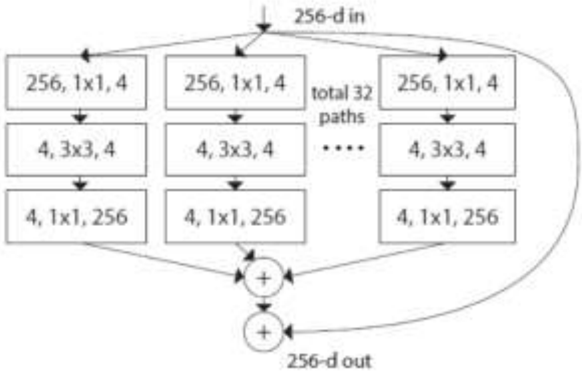


**Figure 7. ResNeXt50 Architecture**

that aggregates a set of transformations with the same topology. The simple design results in a homogeneous, multi-branch architecture that has only a few hyper-parameters to set. This strategy exposes a new dimension, which we call "cardinality" (the size of the set of transformations), as an essential factor in addition to the dimensions of depth and width. On the ImageNet-1K dataset, we empirically show that even under the restricted condition of maintaining complexity, increasing cardinality is able to improve classification accuracy. Moreover, increasing cardinality is more effective than going deeper or wider when we increase the capacity.

## 5. PARAMETERS, OPTIMIZERS, LOSS FUNCTIONSAND ACCURACY CALCULATIONS

Adam Optimizer is an optimization technique used to optimize the weights of the model during the gradient descent algorithm. It is a combination of the gradient descent with momentum algorithm and the RMSP algorithm. Here mt and mt-1 are the weighted averages of the gradients at times t and t-1, beta (b) is the moving average parameter, dL is the derivative of the loss function with respect to the weights and vt is the sum of squares of previous gradients. The most commonly used

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

**Figure 8. Loss function**

loss function for image segmentation tasks is pixel wise cross entropy loss. This loss function compares individual pixels to the one hot encoded target vector. In the current scenario since it is a binary classification, the one hot vector is replaced by binary numbers representing the mask. Another popular loss function is the Dice coefficient, which is a measure of the overlap between the two samples. Dice coefficient can range between 0 and 1 with 1 indicating perfect overlap. This coefficient is used to compare the predicted mask with the actual mask. Here 'A' represents pixel value from the predicted mask and B from the actual mask. The numerator represents the common elements between sets A and B. In order to simplify the calculation of intersection, it is approximated to the element-wise product between the predicted mask and the actual mask. IoU score is the most popular measure of calculating the accuracy of segmentation models. IoU score stands for the ratio of area of intersection over area of union. A score of 1 indicates that both the intersection and union areas overlap and that the predicted and the actual results are the same. The score can also be calculated over the multiple regions of the mask and the average is taken over the entire score represented by the mean IoU score.

$$DICE(A,B) = \frac{2|A \cap B|}{|A|+|B|}$$

**Figure 9. Dice Coefficient**

$$IOU(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

**Figure 10. Intersection Over Union**

The third measure of accuracy index is the Jaccard index. It is defined as the size of the intersection over the size of the union. As input the index is given the actual scores and the predicted scores, currently which are pixel values given by the model. The index 0 indicates the absence of positive or true values in case of binary classification and score of 1 indicates perfect match.

| Parameter | Training Scores | Validation scores | Testing scores |
|---|---|---|---|
| Accuracy Score | 0.9961 | 0.9954 | 0.988 |
| Loss | -0.5571 | -0.5367 | -0.555 |
| Dice Coefficient | 0.5570 | 0.5175 | 0.5513 |
| IoU Score | 0.3907 | 0.3580 | 0.3836 |

**Figure 10. SegNet Results**

| Parameter | Final Iteration | Average |
|---|---|---|
| Accuracy | 0.9740 | 0.9703 |
| Loss | -0.2901 | -0.2421 |
| Dice Coefficient | 0.2901 | 0.2518 |
| IoU Score | 0.1735 | 0.1678 |

**Figure 11. UNet Results**

| Parameter | Training Scores | Validation scores | Testing scores |
|---|---|---|---|
| Accuracy Score | 0.9961 | 0.9928 | 0.9897 |
| Loss | 0.1473 | 0.1643 | 0.1813 |
| Dice Coefficient | 0.8658 | 0.8821 | 0.8690 |
| IoU Score | 0.86 | 0.89 | 0.90 |

**Figure 12. ResNeXt50 Results**

## 6. RESULTS AND CONCLUSION

All 3 models were trained for 10 epochs. Figures 10, 11, 12 show the final results.Segmentation is a very popular task in machine learning with continuous research in the domain. The fundamental task remains the same and newer approaches to the task in the form of model architectures are being studied. In the current paper the use of three such modern state of the art models are used for the task and the performances of these individual models are compared. The popular model is the UNet which has also been presented in the model. Along with UNet two other models, SegNet and ResNet are used and predictions made by the models are also presented. The brain images dataset is a memory exhaustive dataset and consumes memory during training. This places a constraint on the model size and hence on the iterations of the training. However with the more epochs and optimized bottlenecks the accuracy of the predictions can be improved. The SegNet model is found to produce more false positives and false negatives when compared to other models. Since the field of segmentation is continuously exploding with newer models, the use of much more modern state of the art models such as Pyramid Parsing Networks and Feature Pyramids Networks can be made in order to get a diverse range in predictions. The use of such architectures can lead to better accuracy scores and smaller weights file resulting in smaller models which can be deployed in resource
intensive platforms.

## 7. REFERENCES


[1] Ye Huang, Wenjing Jia, Liu Liu, Yuxin Li, Dacheng Tao,"Channelized Axial Attention for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition on 21 Jan 2021.

[2] Dong Nie , Jia Xue , Xiaofeng Ren, "Bidirectional Pyramid Networks for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition in 2020.

[3] Andrew Tao, Karan Sapra, Bryan Catanzaro of Nvidia, "Hierarchical Multi Scale Attention for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition on 21 May 2020.

[4] Barret Zoph, Golnaz Ghiasi, Tsung Yu Lin, "Rethinking Pre-training and Self training" presented at the Conference on Computer Vision and Pattern Recognition on 11 Jun 2020.

[5] Huimin Huang , Lanfen Lin , Ruofeng Tong , Hongjie Hu , Qiaowei Zhang , Yutaro Iwamoto , Xianhua Han , Yen-Wei Chen, Jian Wu, "Unet 3+ : A full scale connected unet for medical image segmentation", presented in IEEE International Conference on Acoustics, Speech and Signal Processing, in 2020.

[6] Yuhui Yuan, Xiaokang Chen, Xilin Chen, Jingdong Wang, "Segmentation Transformer: Object- Contextual Representations for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition on 24 Sep 2019.

[7] Alexander Kirillov Ross Girshick Kaiming He Piotr Dollar, "Panoptic Feature Pyramid Network" presented at the 2019 IEEE Conference on Computer Vision and Pattern Recognition in Jan 2019.

[8] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "UNet++ : A Nested UNet architecture for medical image segmentation" presented at the IEEE Conference on Computer Vision and Pattern Recognition on 18 July 2018.

[9] Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros, "Image to Image Translation with Conditional Adversarial Networks" presented at the Conference on Computer Vision and Pattern Recognition on 26 Nov 2018.

[10] Mark Sandler Andrew Howard Menglong Zhu Andrey Zhmoginov Liang-Chieh Chen, " MobileNetV2: Inverted Residuals and Linear Bottlenecks", presented at the Conference on Computer Vision and Pattern Recognition on 13 Jan 2018.

[11] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, "SegNet: A Deep Convolutional Encoder- Decoder Architecture for Image Segmentation" presented at the IEEE Transactions on Pattern Analysis and Machine Intelligence Volume 39 Issue 21 on Dec 1, 2017.

[12] Simon Jégou, Michal Drozdzal, David Vazquez, Adriana Romero, Yoshua Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition on 31 Oct 2017.

[13] Abhishek Chaurasia, Eugenio Culurciello, "LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation" presented at the IEEE Visual


 

Communications and Image Processing Conference in 2017. Department of Computer Science and Engineering 2020-21 Page 29

[14] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, "ICNet for Real-Time Semantic Segmentation on High-Resolution Images" presented at the IEEE Conference on Computer Vision and Pattern Recognition on 27th April 2017.

[15] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia, "Pyramid Scene Parsing Network" presented at the IEEE Conference on Computer Vision and Pattern Recognition in 2017.

[16] Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick, " Mask RCNN" presented at the Conference on Computer Vision and Pattern Recognition on 20 March 2017.

[17] Fisher Yu, Vladlen Koltun, "Multi Scale Context Aggregation by dilated convolutions" presented at the Conference on Computer Vision and Pattern Recognition on 30 Apr 2016.

[18] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello, " ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition in 2016.

[19] Evan Shelhamer Jonathan Long and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation" presented at the Conference on Computer Vision and Pattern Recognition on 8 March 2016.

[20] Olaf Ronneberger, Philipp Fischer, Thomas Brox, " U-Net : Convolutional Networks for Biomedical Image Segmentation" presented at the Computer Vision and Pattern Recognition Conference, 18 May 2015

 