

Improved UNet Framework with attention for Semantic Segmentation of Tumor Regions in Brain MRI Images

Heena Kousar¹, Arimanda Chaitanya Sri², Saranu Charitha Sri³

^{1,2,3} Student of Dept. of CSE, Vignan's Foundation for Science, Technology & Research, Vadlamudi, India

Abstract - Brain Tumor Segmentation is a crucial task in medical image processing. Brain tumors must be detected early in order to improve treatment options and increase patient survival rates. A challenging and time-consuming task is detecting tumor from a large number of clinical MRI images for cancer diagnosis. Deep learning algorithms for automatic segmentation have recently gained traction due to the fact that these methods produce cutting-edge results and are better suited to this problem than other approaches. Deep learning approaches can also be used to efficiently and objectively process massive amounts of MRI-based image data. Several review papers on classic MRI-based brain tumor image segmentation algorithms are available. Because Semantic Segmentation assigns a class label to each pixel in a given image, it can be used to segment brain tumor images from the provided images. In the proposed methodology, we perform a batch training where each randomly created batch is passed to the variation of UNet, a popular Segmentation model. In this model, we have added batch normalizations following every convolution layer with the hope that a deeper network helps extracting the better features which turned out to be true. Here we prefer to use the metric as Intersection over Union (IoU) [1] rather than accuracy because it is less influenced by the inherent class imbalances in foreground/background segmentation tasks. With the proposed methodology, we achieve an averaged IoU of 84.3 and dice coefficient value is 91.4.

Key Words: Brain Tumor, Segmentation, Semantic Segmentation, U-Net, Intersection over Union (IoU), Dice coefficient

1. INTRODUCTION

Cancer is defined as uncontrollable and abnormal cell division and proliferation in the body. A brain tumor is an abnormal mass of unnatural cell growth and division in brain tissue. Brain tumors are one of the most fatal cancers¹, despite their rarity.

Brain tumors [2] are classified as either primary or metastatic based on where they originate. Primary cancer cells originate in brain tissue, whereas metastatic cancer cells become cancerous elsewhere in the body and spread to the brain. Gliomas are brain tumors that develop from glial cells. While these modalities are used in conjunction to provide the most detailed information, because of its high soft tissue contrast and widespread availability, MRI is routinely used to obtain information regarding brain malignancies. The conventional method MRI (magnetic resonance imaging) is a non-invasive in-vivo imaging technique that employs radio frequency pulses to excite tissues.



Figure 1.1: Sample for segmenting the Brain Tumor images

We use segmentation to efficiently locate and segment brain tumors in order to perform successful surgery. Brain tumors can be classified into two types. The first is manual segmentation, which is a subjective decision that does not produce the desired results because completely removing brain tumors without destroying healthy brain tissue is difficult. As a result, automatic segmentation for treatment planning and quantitative evaluation, the second method is required. It quickly and accurately diagnoses brain tumors.

Since both the location and size of the tumors are required to be identified, the problem comes under the task of segmentation and it particularly comes under semantic segmentation. Segmentation is something beyond the tasks like Image classification,

localization and object detection. In image classification, we just require the given image to be classified into one of the classes (binary or multi).

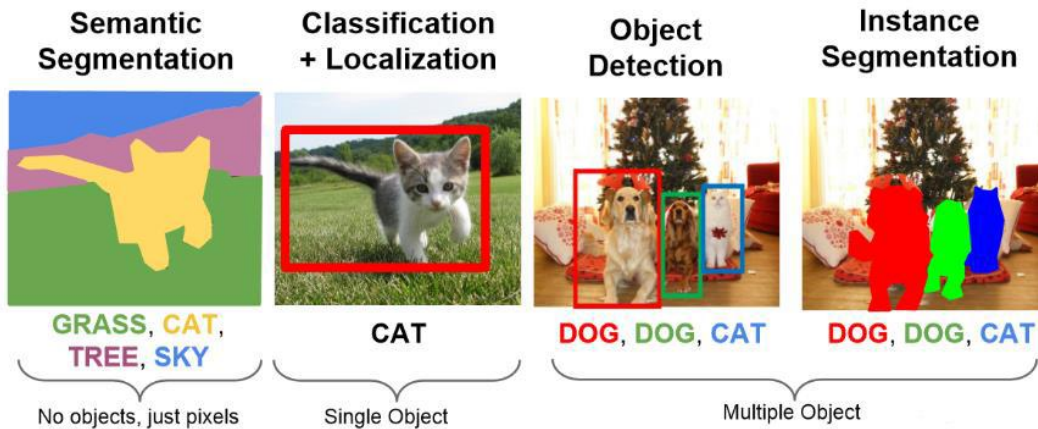


Figure 1.2: Levels of Understanding of an Image by a System

Localization is a bit advancement to classification as We locate the required object in the provided image. Object detection is like combination of both of those because here we perform both the tasks like classifying the object and locating it. Here locating the image is to just come up with a bounding box and hence it is not the case to be used when we need the exact shape of the Image..

2. Literature Survey

Since this brain tumor segmentation is related to the, many researchers get attracted to this work. As part of the initial stage of research in this topic, the researchers use to consider the hand craft different feature extractors and use the outputs of them for the analysis of the brain tumor images. One approach is the method proposed by Nagashree N and Premjyoti Patil [3]. This system's main idea is to work on the encoding and decoding phases of UNet [4] modelling for efficient segmentation of brain images. The input image is divided into several layers called convolutions in this methodology, and the CNN method is used. The process's convolution filter is the feature extraction of individual image layers. In the UNet approach, each layer is represented as a network encoder layer. Alphabet pruning, an AI optimization algorithm for dimensionality reduction, was proposed as a modified form of UNet. The process entails building a tree network out of all the layers of the input image, retaining only the essential images. The remaining image layers are pruned to save time.. The workflow of their approach is as follows:

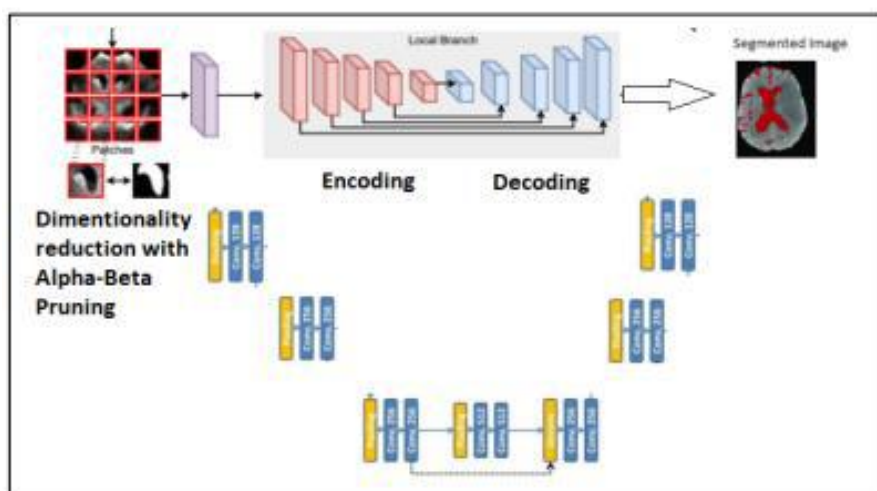


Figure 2.1: Workflow of Nagashree N and Premjyoti Patil

Convolutional neural networks have been widely used in the field of medical picture segmentation since the introduction of deep learning for their great feature extraction capabilities, and have achieved good segmentation performance and robustness. Convolutional neural networks were originally used in brain tumor segmentation by Zikic et al. The network comprises of a convolution layer, a maximum pooling layer, a full connection layer, and a softmax layer. Ronneberger et al suggest the Unet network, which uses an encoder-decoder topology. Convolution with size of 33 and stride size of 1 is used for 4 times down-sampling in the coding phase; deconvolution with size of 22 and stride size of 2 is used for up-sampling in the decoding phase. High-resolution and low-resolution information are equally relevant due to the similarity of medical imaging and the fuzziness of tumor region boundaries. QingJun Ru and GuangZhu Chen [5] propose an improved M-Unet structure to increase the performance of feature fusion and the accuracy of network segmentation. This approach can be improved in the following ways:

1. A multi-scale feature extraction module is added to the Unet network's feature fusion part to better extract the high-level and low-level features of tumor images, while redundant features are avoided from being introduced into the up-sampling feature map, further improving network segmentation performance.
2. To acquire the best network weight, a cosine annealing learning rate attenuation approach is utilised in the training phase to make the network jump out of the local optimal solution.

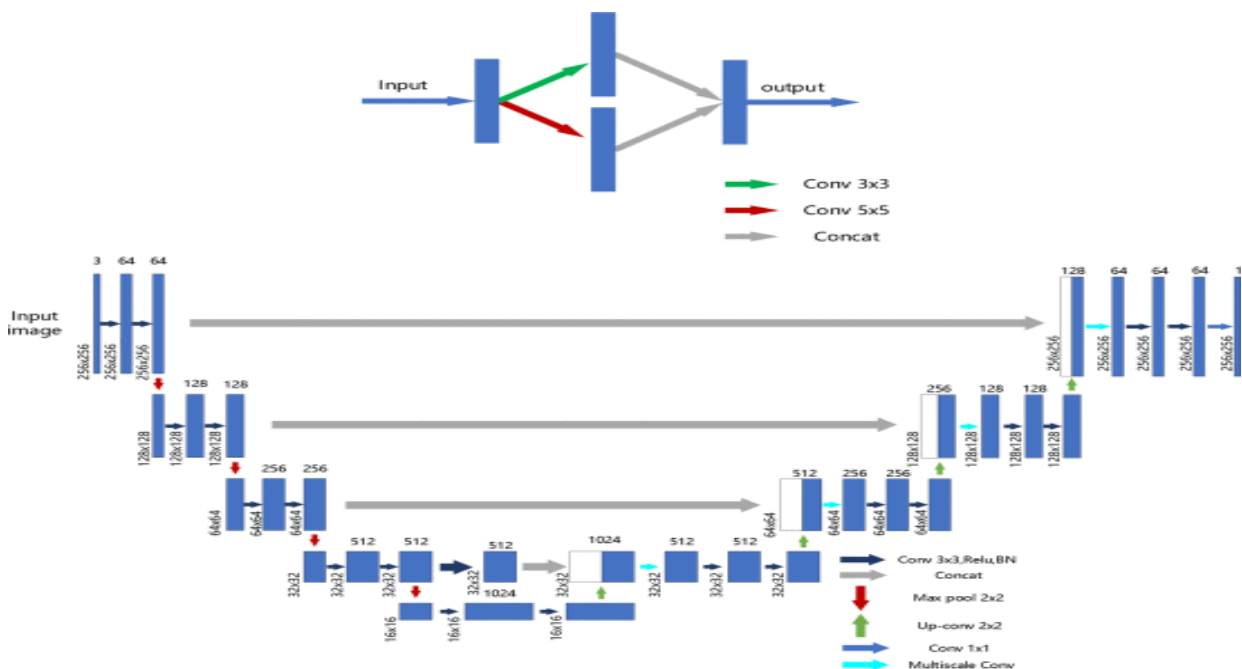


Fig 2.2: Architecture Proposed by QingJun Ru and GuangZhu Chen

3. Methodology

This architecture of the proposed model, training approach and the other technical details are discussed in this section. The proposed architecture is derived from the UNet architecture. Many changes like the number of filters at each layer, introduction of Batch Normalization operations, were made to the original score. The dataset used for this work was lgg MRI segmentation dataset. This dataset [6] comprises of 3762 images of size 256 × 256. Out of the 3762 images, 80% of them were used for the training purpose and the rest of the images (20%) were used for testing purpose.

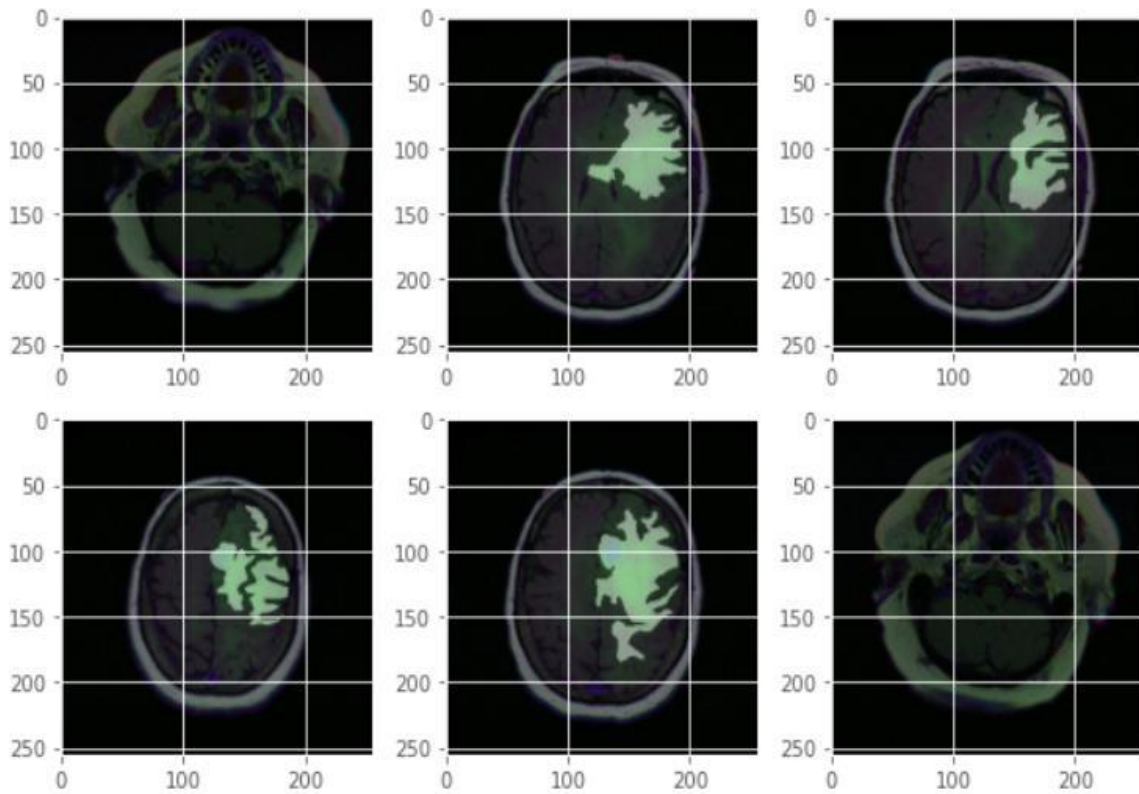


Figure 3.1: Sample Images

As shown in the figure 3.2, the training images and the corresponding masks were loaded and as a visualization technique. In the proposed methodology, we are using random 2828 images for training and 708 for validation and 393 for testing.

The performance of the model is monitored in form of the metric Intersection of Union (IoU) , which was the most common metric used for the segmentation tasks. Call backs like Early Stopping and Model Check point are further implemented on the basis of average IoU of every 50 batches. These callbacks helps to save the best model and stop the training process if there was no further improvement. At the end, save the model weights so that to use them later. There are many other countless efforts like augmentation, using different architectures were made but none of them proved to be successful.

4. Network Architecture

The proposed network is derived from the UNet architecture. UNet, being a popular approach to be used for segmentation task, applies classification on each and every pixel in the given input image and thereby producing a mask of same size as input.

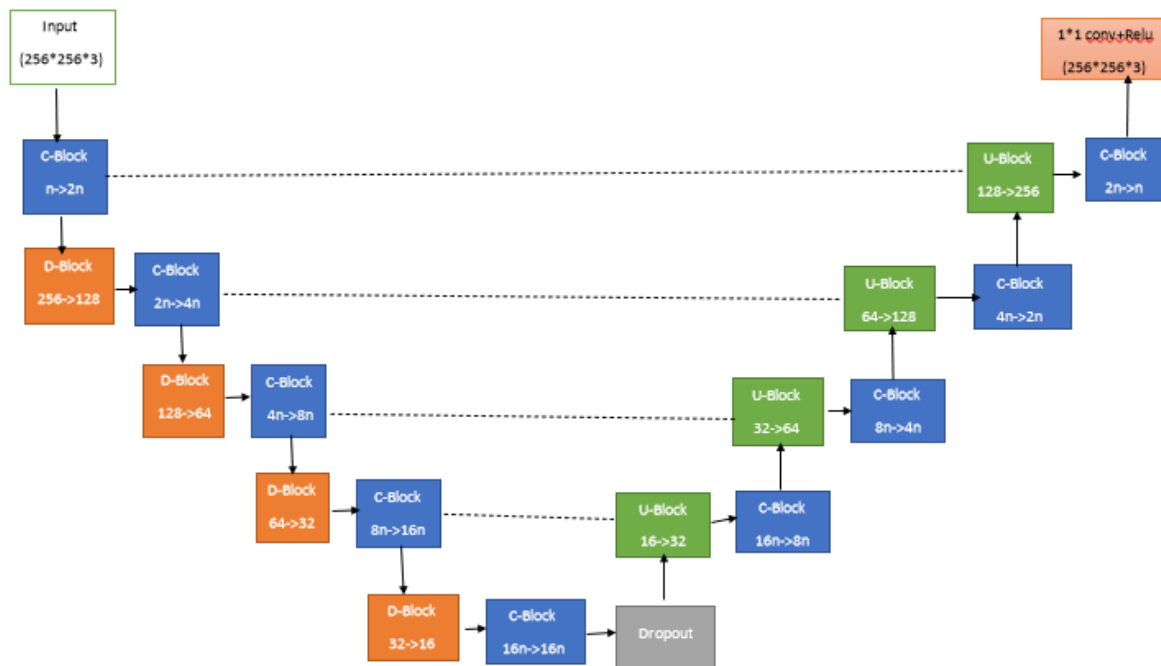


Figure 4.1: Proposed Architecture of varied UNet

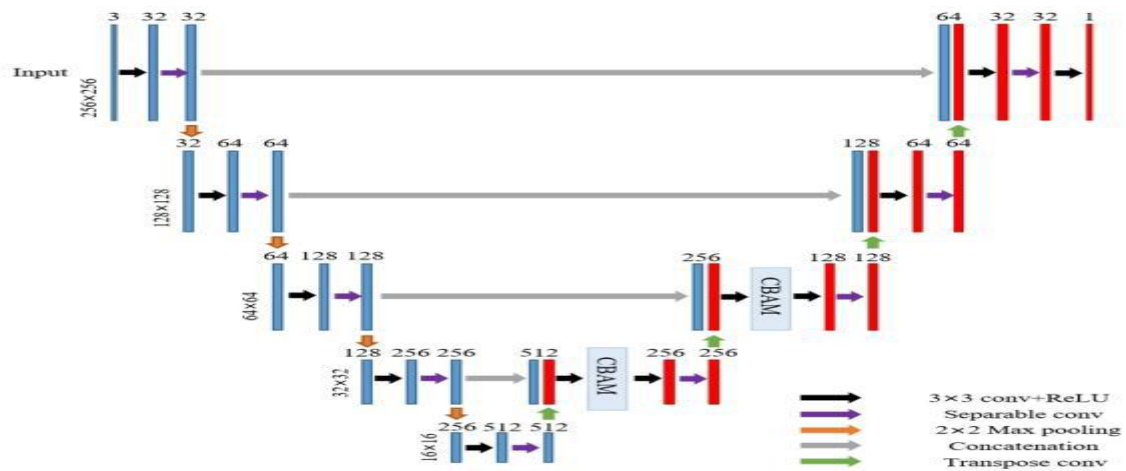


Figure 4.2: Proposed Architecture of varied UNet

The network shown above takes an input of size $(256 \times 256 \times 3)$. It features an encoder-decoder structure, and in the encoder part, we apply pooling techniques to reduce the size of the image. so that we can extract the information of "what" is present in the image. We increase the picture size in the decoder phase to extract the information of "where" in the given image. Each block in the contracting path has the following structure:

convolution1 \rightarrow convolution2 \rightarrow max_pooling \rightarrow Batch_Normalization

Two convolutional layers are used in the first block, followed by pooling and batch normalization processes, and the channel count is increased from one to eight. Since the process of convolution increases the depth of the image, there are four such blocks in the entire contraction path, and by the end of all four blocks, the channel count has increased to 64 channels. The image size is reduced to by the conclusion of the four blocks (8,8) thanks to the max pooling procedure, which decreases the image size.. From here, the expansive path will begin, in which the image size is steadily raised through upsampling while the channel count is reduced.

Convolution_2d_transpose → concatenate → convolution1 → convolution2

Transposed convolution is employed here as an up sampling technique to increase image size. On the initial image, a padding operation is done, followed by a convolution action. There are four such blocks here, just like in the contraction path, and by the end of these blocks, we will have the original size image. The final prediction is obtained by applying a 1D convolution with 1 kernel and a sigmoid activation on the output of the last block. The 1D convolution reduces the number of channels necessary for the network output, while the sigmoid activation function maps every pixel in the output block to the range of the required network output (0, 1). The results will be rounded to the nearest integer.

The model weights are retained at the end of the training and used in the testing procedure. We use the sharpening technique in the final stage of the testing procedure, after the output mask prediction. As a post-processing approach, sharpening allows for a greater view of the salt deposits present in the projected mask, resulting in a higher IoU score. Low pass and high pass filters are commonly used on photographs to improve their viewing capabilities. Smoothing is the term used to describe the use of a low pass filter, whereas sharpening is the term used to describe the use of a high pass filter. Low frequencies are frequently attenuated by a high pass filter, which allows high frequencies to flow through. As a result, the salt pixels in the expected mask pass through the filter, yielding a superior outcome. The kernel for sharpening in our suggested methodology is represented by the following array.

5. Results and Discussions

In the testing phase, we load the remaining 20% data with images and masks. The weights that are saved earlier are loaded into the model and the model is used for the prediction of the masks for the given images. By comparing the anticipated and original masks, we can now determine the IoU (Intersection over Union) value. The average of IoU was then calculated for a range of criteria ranging from 0.5 to 0.95, with a 0.05 step between each is reported. IoU on a threshold tells that a particular IoU value has crossed that threshold. For Example, a predicted output mask is considered to be valid over a threshold of 0.7 if the value of IoU is above 0.7 for that particular mask. The following is the interpretation of IoU between genuine segmentation pixels, Y, and a similar set of predicted segmentation pixels,

$$IoU(Y, \hat{Y}) = \frac{Y \cap \hat{Y}}{Y \cup \hat{Y}} = \frac{\sum_{i=1}^n \min(Y_i, \hat{Y}_i)}{\sum_{i=1}^n \max(Y_i, \hat{Y}_i)}$$

Fig 5.1: IOU Formula

Which can also be expressed as a function of the Y-Y confusion matrix

	$\hat{Y} = 1$	$\hat{Y} = 0$
$Y = 1$	TP	FN
$Y = 0$	FP	TN

Fig 5.2: General Confusion Matrix

IoU is then calculated as: (TP=True positives, FP=False positives, etc.)

$$IoU(Y, \hat{Y}) = \frac{TP}{TP + FN + FP}$$

Fig 5.3: IoU Formula in terms of confusion Matrix

Hence the results of the different methods on various thresholds and the average IoU over all the thresholds from 0.5 to 0.95 with a step value of 0.05 and the loss of training is reported in the following table.

Model	Loss	Testing IOU at Different Thresholds				Average Testing IOU
		0.5	0.7	0.8	0.85	
Unet (Proposed Architecture)	0.81	0.84	0.82	0.80	0.79	0.84

Table 5.1: Final Results

Model	Dice Coefficient
Unet(proposed architecture)	0.914
M-Unet(QingJun Ru, GuangZhu Chen) [5]	0.873
Alpha Beta Pruned Unet (Nagashree N,Prem Jyoti Patil) [3]	0.901

Table 5.2: Comparison of results with citations

The Input image along with the true mask and the predicted mask are plotted here for various images.

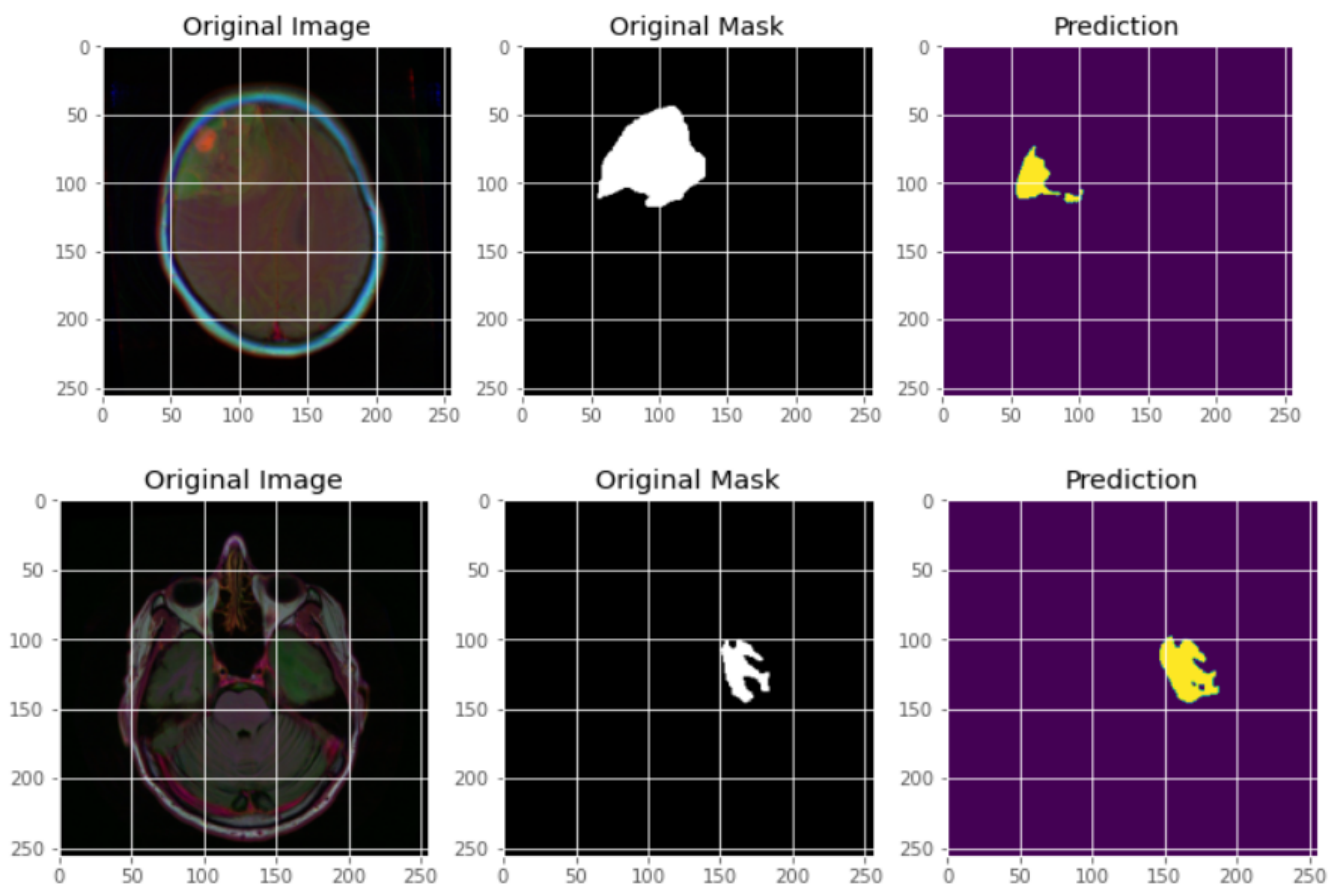


Figure 5.4: Predicted Masks comparison with the original mask

6. Conclusion

We perform batch training in the proposed manner, where each randomly formed batch is submitted to a variant of UNet, a popular Segmentation model. We added batch normalizations after each convolution layer in this model in the hopes that a

deeper network will assist extract better features, which proved out to be accurate. Instead of accuracy, we opt to utilise the measure Intersection over Union (IoU) [1], This is less influenced by the inherent class imbalances in foreground/background segmentation tasks. We get an averaged IoU of 84.3 and a dice coefficient value of 91.4 using the provided methods. The proposed model will be improved in the future by employing different filter sizes and including all modalities of MRI images in tumor segmentation. By raising the mini-batch size from 16 to 64 and the max-epoch from 80 to 120, the segmentation result will be improved even more.

7. References

- [1] Rezatofighi, Hamid, et al. "Generalized intersection over union: A metric and a loss for bounding box regression." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
- [2] I Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
- [3] Alpha Beta Pruned UNet-A Modified UNet Framework to Segment MRI Brain Image to Analyse the Effects of CNTNAP2 Gene towards Autism Detection. In 2021 3rd International Conference on Computer Communication and the Internet (ICCCI) (pp. 23-26). IEEE.-240). IEEE.
- [4] Kermi, A., Mahmoudi, I., & Khadir, M. T. (2018, September). Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. In International MICCAI Brainlesion Workshop (pp. 37-48). Springer, Cham.
- [5] Ru, Q., Chen, G., & Tang, Z. (2021, August). Brain Tumor Image Segmentation Method Based on M-Unet Network. In 2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 243-246). IEEE.
- [6] Datasets Used in The Discussed Model [Online]. Available: <https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation>.
- [7] Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 2481–2495. [Google Scholar] [CrossRef] [PubMed]
- [8] Guo, Meng-Hao, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R. Martin, Ming-Ming Cheng, and Shi-Min Hu. "Attention mechanisms in computer vision: A survey." Computational Visual Media (2022): 1-38.
- [9] Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1520–1528.
- [10] Wu, X.; Liang, L.; Shi, Y.; Fomel, S. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 2019, 84, IM35–IM45
- [11] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
- [12] McCaffrey, J. (2014). Understanding Neural Network Batch Training: A Tutorial. Visual Studio Magazine.
- [13] Daimary, Dinthisrang, et al. "Brain tumor segmentation from MRI images using hybrid convolutional neural networks." Procedia Computer Science 167 (2020): 2419-242
- [14] Rehman, Mobeen Ur, et al. "Bu-net: Brain tumor segmentation using modified u-net architecture." Electronics 9.12 (2020): 2203
- [15] Deb, Daizy, and Sudipta Roy. "Brain tumor detection based on hybrid deep neural network in MRI by adaptive squirrel search optimization." Multimedia tools and applications 80.2 (2021): 2621-2645

8. Biographies



Heena Kousar

Department of Computer Science & Engineering
Vignan's Foundation for Science, Technology & Research



Arimanda Chaitanya Sri

Department of Computer Science & Engineering
Vignan's Foundation for Science, Technology & Research



Saranu Charitha Sri

Department of Computer Science & Engineering
Vignan's Foundation for Science, Technology & Research