Fine-Tuning U-Net for Brain Tumor Segmentation: Custom Layer Training on BraTS 2020 MRI Scans

[1] Rajesh Kannan V, [2] Basil Tamil Selvasn, [3] Suriya V, [4] Gajendran R

[1] Dept. of Computer Science and Engineering Kamaraj College of Engineering and Technology Madurai,India
rajeshkannancse@kamarajengg.edu.in

[2] [3] [4] Dept. of Computer Science and Engineering Kamaraj College of Engineering and Technology Madurai,India

Abstract: This study explores the application of U-Net, a convolutional neural network (CNN) architecture tailored for biomedical image segmentation, in brain tumor segmentation using the BraTS 2020 dataset. U-Net's unique dual-path framework, comprising a contracting path for feature extraction and an expansive path for precise localization, facilitates accurate segmentation by capturing contextual information while preserving spatial details. Leveraging the TensorFlow framework, we develop a U-Net model and incorporate custom layer training processes to enhance its performance. The BraTS 2020 dataset, consisting of multi-institutional pre-operative MRI scans, presents challenges in accurately delineating tumor sub-regions such as the enhancing tumor, peritumoral edema, and tumor core. Our methodology includes pre-processing steps such as intensity normalization, noise reduction, and bias field correction to improve the quality of MRI images before model training. Through rigorous experimentation and evaluation, we demonstrate the effectiveness of our U-Net model in accurately segmenting brain tumor regions, providing valuable insights for clinical decision-making. Furthermore, we discuss the architectural components and functionalities of U-Net, emphasizing its suitability for various clinical scenarios across different medical imaging modalities. Our research contributes to advancing the field of DL-based medical image analysis by showcasing the efficacy of U-Net in addressing challenges associated with brain tumor segmentation. We also highlight the implications of our findings, discuss study limitations, and suggest potential avenues for future research. This study underscores the potential of DL techniques in automating medical image analysis tasks, ultimately improving diagnostic accuracy and patient outcomes in clinical settings.

Keywords: CNN, Deep Learning formatting, U-Net, Segmentation.

I. INTRODUCTION

Healthcare revolves around the well-being of individuals, and the present era witnesses an immense volume of medical data. Leveraging this vast repository of medical data is pivotal for advancing the healthcare industry. However, despite its abundance, medical data poses several challenges. It encompasses a diverse range of formats such as images, texts, videos, and scans. Moreover, variations in data quality arise from the use of different medical equipment. The dynamic nature of medical data further complicates analysis, as it fluctuates over time and in response to specific events. Additionally, the absence of a universal disease pattern due to individual differences adds another layer of complexity. Despite these challenges, medical imaging stands out as a crucial component of medical data analysis. Medical image segmentation serves as a cornerstone in modern healthcare, facilitating precise delineation of anatomical structures and pathological regions from imaging data. With the advent of deep learning techniques, convolutional neural networks (CNNs) have emerged as powerful tools for automated segmentation tasks. Among CNN architectures, U-Net stands out for its unique design tailored to handle medical image segmentation challenges effectively. In this section, we provide an overview of the significance of medical image segmentation and introduce the U-Net architecture as a promising solution for this critical task. The accurate segmentation of medical images is essential for various clinical applications, including diagnosis, treatment planning, and monitoring of disease progression. Traditional segmentation methods often rely on handcrafted features and require extensive manual intervention, limiting their scalability and accuracy. In contrast, deep learning-based approaches leverage the hierarchical representations learned directly from data, enabling more robust and efficient segmentation. U-Net, in particular, addresses the challenges of medical image segmentation by incorporating a contracting path for feature extraction and an expansive path for precise localization. This architecture enables U-Net to capture contextual information while preserving spatial details, making it well-suited for capturing intricate structures in medical images.

II. LITERATURE SURVEY

In "Deep Learning: A Primer for Radiologists "The author Chartrand et al., published in the journal RadioGraphics in 2017 Deep learning (DL) is a subset of machine learning (ML) that has gained significant traction in recent years due to its ability

to learn from large amounts of data and make accurate predictions. This primer aims to provide radiologists with a foundational understanding of DL and its applications in medical imaging.

DL has demonstrated impressive performance in a variety of radiology tasks, including, Image Segmentation in which DL can automatically identify and delineate anatomical structures and lesions in medical images. This is crucial for tasks such as tumor detection, organ segmentation, and image quantification.Image Classification, DL can classify medical images into different categories, such as normal, abnormal, or specific pathologies. This facilitates faster and more accurate diagnoses.DL can be used to reduce noise in medical images, improving image quality and enhancing the visibility of subtle details. This is particularly useful for low-dose CT scans and MRIs. Image Generation: DL can generate synthetic medical images, such as CT scans from MRI data, which can have valuable applications in medical training and data augmentation. Image Registration, DL can be employed to align medical images from different modalities or time points, facilitating comparisons and treatment planning.

In the field of medical image analysis, particularly in neuroimaging, the accurate segmentation of brain lesions from MRI scans stands as a critical challenge in clinical diagnosis and treatment planning. Manual segmentation methods are time-consuming and prone to human error, necessitating the development of automatic approaches for faster and more consistent lesion delineation. Addressing this need, a recent study introduces a deep learning pipeline comprising two key components tailored to 3D medical images: a Multi-scale 3D Convolutional Neural Network (CNN) and a Fully Connected Conditional Random Field (CRF) post-processing technique. The multi-scale 3D CNN employs a dual pathway architecture to simultaneously capture local details and larger contextual information within the brain anatomy. Following initial segmentation by the CNN, the fully connected CRF refines the results by incorporating spatial relationships between voxels, thereby enhancing segmentation accuracy. Noteworthy contributions include an efficient training scheme that addresses class imbalance and an analysis of network depth's impact on segmentation performance. Evaluation on challenging lesion segmentation tasks, including Traumatic Brain Injuries, Brain Tumors, and Ischemic Stroke, yielded state-of-the-art results, surpassing previous methods on benchmark datasets like BRATS 2015 and ISLES 2015. Furthermore, the incorporation of the fully connected CRF post-processing step notably reduced false positives, underscoring the method's potential for clinical applications in neurological disorder diagnosis and treatment planning.

III. MEDTHODOLOGY

3.1 Data Collection and Preprocessing

This stage involves gathering the necessary data to train the machine learning model. The data should be relevant to the task at hand and of sufficient volume to allow the model to learn effectively. The raw data collected may not be directly suitable for training the model. Data preprocessing entails cleaning, transforming, and preparing the data to ensure it adheres to the model's requirements and improves its accuracy. This may involve tasks like data cleaning, normalization, and feature engineering. The preprocessed MRI scans are fed into a deep learning-based segmentation model, such as UNet, to segment the brain tumors from the background. The segmentation model generates pixel-wise segmentation masks delineating the tumor regions, including the tumor core, enhancing tumor, and peritumoral edema

3.2 Feature Extraction and Model Traning
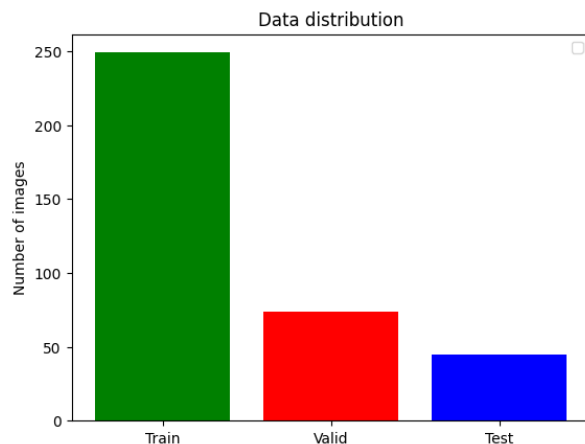
After segmentation, features are extracted from the segmented tumor regions to characterize their shape, texture, and intensity properties. Feature extraction techniques may include histogram-based features, texture analysis using gray-level co-occurrence matrices (GLCM), and morphological features. Once the data is prepared, the model is trained using an algorithm that learns from the data. The model's parameters are adjusted iteratively to minimize its error on the training data. After training, the model's performance is evaluated on unseen data, which is typically a portion of the collected data set that was not used for training. This provides an unbiased assessment of the model's generalizability and ability to perform well on real-world data.
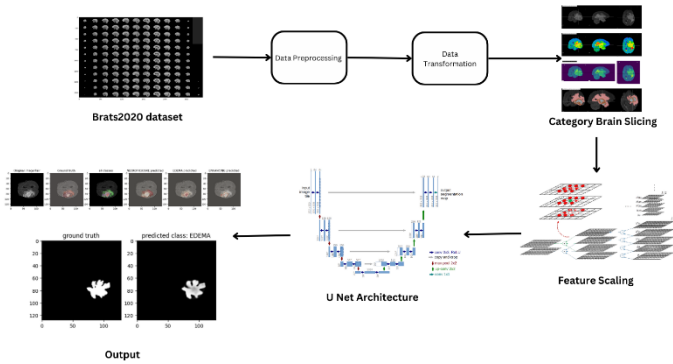

Fig 3.2

IV. IMPLEMENTATION

UNet is a convolutional neural network (CNN) architecture commonly used for semantic segmentation tasks, particularly in biomedical image analysis. UNET is a U-shaped encoder-decoder network architecture, which consists of four encoder blocks and four decoder blocks that are connected via a bridge. The encoder network (contracting path) half the spatial dimensions and double the number of filters (feature channels) at each encoder block. Likewise, the decoder network doubles the spatial dimensions and half the number of feature channels.

The encoder network acts as the feature extractor and learns an abstract representation of the input image through a sequence of the encoder blocks. Each encoder block consists of two 3x3 convolutions, where each convolution is followed by a ReLU (Rectified Linear Unit) activation function. The ReLU activation function introduces non-linearity into the network, which helps in the better generalization of the training data. The output of the ReLU acts as a skip connection for the corresponding decoder block. Next, follows a 2x2 max-pooling, where the spatial dimensions (height and width) of the feature maps are reduced by half. This reduces the computational cost by decreasing the number of trainable parameters.

These skip connections provide additional information that helps the decoder to generate better semantic features. They also act as a shortcut connection that helps the indirect flow of gradients to the earlier layers without any degradation. In simple terms, we can say that skip connection helps in better flow of gradient while backpropagation, which in turn helps the network to learn better representation.

The bridge connects the encoder and the decoder network and completes the flow of information. It consists of two 3x3 convolutions. Multimodal segmentation in medical image analysis refers to the process of segmenting structures or regions of interest within images obtained from multiple imaging modalities. This approach is particularly useful in healthcare applications, where different imaging modalities, such as MRI, CT, PET, or ultrasound, provide complementary information about anatomical structures or pathological conditions. In the context of brain tumor segmentation, multimodal segmentation involves integrating information from different MRI sequences (e.g., T1-weighted, T2-weighted, FLAIR, and T1-weighted with contrast enhancement) to improve the accuracy and robustness of tumor segmentation. Each MRI sequence highlights different aspects of the tumor, such as its morphology, texture, intensity, and surrounding edema, making multimodal segmentation essential for capturing the full extent of the tumor and its heterogeneous characteristics. Each MRI sequence provides unique and complementary information about the brain anatomy and pathology. For example, T1-weighted images are good for visualizing structural details, T2-weighted images highlight edema and soft tissue abnormalities, FLAIR sequences suppress cerebrospinal fluid (CSF) signal to enhance contrast, and T1-weighted with contrast enhancement highlights areas of active tumor enhancement.

V.   RESULTS AND FUTURE WORK

The two graphs on the left show training loss and validation loss. Training loss is a measure of how well a model performs on the data it is trained on. Validation loss is a measure of how well a model performs on data it has not been trained on. In general, the lower the loss, the better the model is performing. The two graphs on the right show training accuracy and validation Dice coefficient. Dice coefficient is a metric used to evaluate the performance of image segmentation algorithms. It is a measure of how well two sets of data overlap. In this case, the two sets of data are the ground truth (the actual segmentation of the brain tumor) and the segmentation produced by the model. A Dice coefficient of 1 indicates perfect overlap, while a Dice coefficient of 0 indicates no overlap. The graphs show that both training loss and validation loss decrease as the number of training epochs increases. This indicates that the model is learning from the training data.

The graphs also show that training accuracy and validation Dice coefficient increase as the number of training epochs increases. This indicates that the model is getting better at segmenting brain tumors.
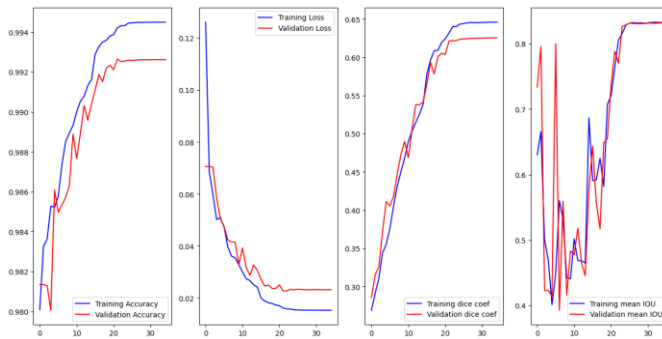

Fig 5.1

Dice loss is a crucial metric utilized in tasks like image segmentation, especially in scenarios involving imbalanced datasets. It serves to evaluate the performance of machine learning models, particularly in tasks such as identifying brain tumors in MRI scans. In this context, the model segments the image into pixels and aims to classify each pixel as either part of the brain tumor (foreground) or not (background). The Dice loss assesses the extent to which the model's segmentation aligns with the true tumor location. Derived from the Dice coefficient, the loss quantifies the similarity between the model's predicted segmentation and the ground truth. By calculating the intersection and union of the two sets of pixels representing the actual and predicted tumor segmentations, respectively, the Dice coefficient indicates the level of overlap. A value of 1 signifies perfect agreement, while 0 indicates no overlap. Consequently, Dice loss, being 1 minus the Dice coefficient, offers a measure of dissimilarity between the predicted and actual segmentations, with lower values denoting a closer match and thus better performance. Implementing a custom callback to adjust the learning rate during training.

The callback has a parameter dwell. If dwell is set to True, the callback monitors the validation loss. It keeps track of the lowest validation loss thus far achieved in run through each epoch and stores this as the lowest loss and also stores the weights for that epoch as the best weights. At the end of an epoch the validation loss for that epoch is compared with the lowest loss. If the validation loss at the end of the current epoch is less than the lowest loss than it becomes the lowest loss and the weights of the current epoch become the best weights If the validation loss at the end of the current epoch is greator than the lowest loss this implies you have moved to a location in Nspace(N is the number of trainable parameters on the validation cost function surface that is less favorable(higher cost) than the position in Nspace defined by the best weights
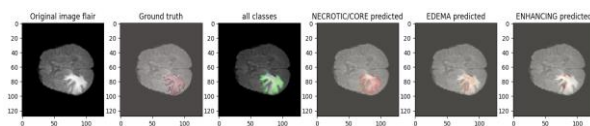

Fig 5.2

Integrating information from multiple MRI sequences, multimodal segmentation improves the sensitivity and specificity of tumor segmentation. It allows for a more comprehensive assessment of tumor boundaries, enhancing the accuracy of tumor delineation compared to using a single MRI sequence alone. Brain tumors can exhibit heterogeneous characteristics across different imaging modalities and tumor subtypes. Multimodal segmentation helps account for this variability by capturing complementary features from different MRI sequences, enabling more robust and accurate segmentation across diverse tumor

types and imaging conditions. Accurate segmentation of brain tumors from multimodal MRI scans is crucial for clinical decision-making, treatment planning, and monitoring of patients with brain tumors. It provides clinicians with valuable insights into tumor morphology, growth patterns, and treatment response, facilitating personalized treatment strategies and improved patient outcomes
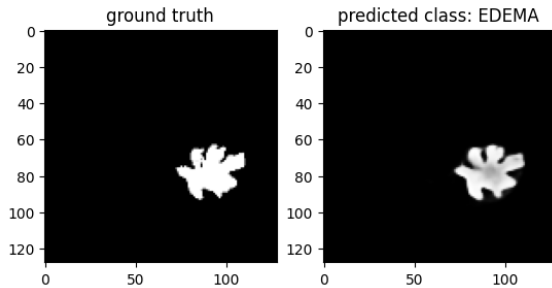

Fig 5.3

A.   Future Works

As the healthcare landscape continues to evolve, the future holds exciting possibilities for further enhancing the application of Artificial Intelligence (AI) in the classification of kidney-related conditions in CT images. Expanding the scope to include multiple imaging modalities, such as MRI and ultrasound, will facilitate comprehensive diagnoses by capitalizing on information from various imaging, Rigorous clinical trials and collaborations with healthcare institutions and professionals will validate the AI system's real-world performance and safety.

REFERENCES

[1]   Rao, C., & Liu, Y. (2020). Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. Computational Materials Science, 184, 109850.

[2]   Myronenko, A., & Winkler, S. (2016). Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation

[3]   Angermeier, C., Khan, N., & Maier-Hein, K. (2018, February). MRI tumor segmentation with densely connected 3D CNN. In Proceedings of the SPIE Medical Imaging (Vol. 10574, p. 105740O)

[4]   Ge Xing, Xiaotian Qiao, Shanshan Wang, Li Wang, Yifan Peng. "SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation" (arXiv: https://arxiv.org/abs/2401.13560)

[5]   SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation: https://github.com/ge-xing/SegMamba

[6]   Xia, Y., Liu, J., Li, X., Zhang, Y., & Sun, L. (2023). CD-Net: Cascaded 3D Dilated convolutional neural network for pneumonia lesion segmentation. Computers in Biology and Medicine, 177124900112.M. Young