**MRI IMAGE PROCESSING FOR DETECTING TUMORS AND DAMAGE**

**Abstract**

With the increased interest in computer-aided image analysis methods, there is a greater need for objective methods of algorithm evaluation. Validation of in vivo MRI studies is complicated by a lack of reference data and the difficulty of constructing anatomically realistic physical phantoms. The authors present here an extensible MRI simulator that efficiently generates realistic three-dimensional (3D) brain images using a hybrid Bloch equation and tissue template simulation that accounts for image contrast, partial volume, and noise. This allows image analysis methods to be evaluated with controlled degradations of image data.

**Introduction**

Information is conveyed through images. The main aim of all image processing techniques is to recognize the image or object easier and with hope of some good accuracy. All the images used in today’s world are in the digital format. In the medical field, these images show the physical attributes distribution of the persons organs (most commonly the brain). Medical imaging modalities as in MRI, CT scan mostly depend on computer technology to generate or display digital images of the internal organs of the human body which helps the doctors to visualize the inner portions of the body, helping with discrete diagnoses of the persons disease. CT scanner, Ultrasound and Magnetic Resonance Imaging took over conventional x-ray imaging for allowing the doctors see the body's third dimension. By adding the third dimension, doctors can scan the whole organ and display it as multiple layered images, thus creating an “in hand” feel of the organ, enabling them to check every part and diagnose accurately.

**Magnetic Resonance Imaging (MRI)**

Protons and neutrons of the nucleus of an atom have an angular momentum which is known as a spin. These spins will cancel when the number of subatomic particles in a nucleus is even. Nuclei with odd number will have a resultant spin. This forms the basis of MRI. An MRI scanner uses powerful magnets to polarize and excite hydrogen nuclei (single proton) in human tissue, which produces a signal that can be detected, which can be encoded spatially, resulting in images of the body. The MRI machine emits a radio frequency (RF) pulse that specifically binds only to hydrogen. The system sends the pulse to that specific area of the body that needs to be examined. Due to the RF pulse, protons in that area absorb the energy needed to make them spin in a different direction. This is what resonance of MRI means. The RF pulse makes the protons spin at the Larmor frequency, in a specific direction. This frequency is found based on the tissue being imaged and the strength of the main magnetic field. MRI uses three electromagnetic fields: static field, which is a very strong static magnetic field that polarizes the hydrogen nuclei; gradient field, which is a weaker time-varying field used for spatial encoding; and a weak radio frequency field, for manipulation of the hydrogen nuclei to produce measurable signals, which are collected through radio frequency antenna.

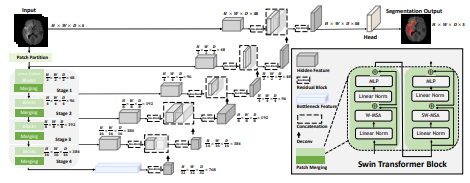
**Image Segmentation**

Image segmentation can be defined as the partition or segmentation of a digital image into similar regions with a main aim to simplify the image into something that is more meaningful and easier to analyze visually. Image segmentation is the main significant process in most of the medical image analysis. Image segmentation methods can be classified as thresholding, region based, supervised and unsupervised classification techniques. Various approaches have been carried out in the field of brain tumor detection. Sindhushree. K.S, et al[14] have developed a brain tumor segmentation method and validated segmentation on two dimensional MRI data. Also, detected tumors are represented in 3-Dimensional view. High pass filtering, histogram equalization, thresholding, morphological operations and segmentation using connected component labeling was carried out to detect tumor. The two dimensional extracted tumor images were reconstructed into three dimensional volumetric data and the volume of the tumor was also calculated. M.C. Jobin Christ and R.M.S. Parvathi[6] proposed a methodology that integrates K Means clustering with marker controlled watershed segmentation algorithm and integrates Fuzzy C Means clustering with marker controlled watershed segmentation algorithm separately for medical image segmentation. The proposed methodology is a two stage process. First K-means clustering (Fuzzy C Means) is used to get a primary segmentation of the input image, and secondly marker controlled watershed segmentation algorithm is applied to the primary segmentation to get the final segmented image. P.Vasuda, S.Satheesh [12], proposed a technique to detect tumors from MR images using fuzzy clustering technique.

**Challenges of Brain Tumor**

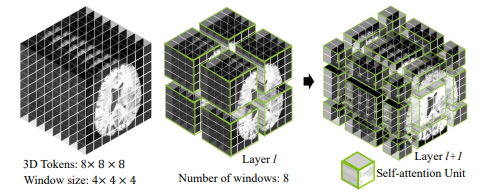
The brain is the anterior most part of the central nervous system. Brain tumor is an intracranial solid neoplasm. Tumors are created by an abnormal and uncontrolled cell division in the brain. In this work, we have used axial view of the brain image (2D) from MRI scan because MRI scan is less harmful than CT brain scan. A patient is subjected to different diagnostic methods to determine the cause of the symptoms mentioned by him. Techniques like performing a biopsy, performing imaging, like taking an MRI or CT scan of the brain will be done. In biopsy, pathologists take a specimen of the brain tissue under consideration for checking the presence of tumor. A pathologist looks at the tissue cells under a microscope to check for presence of abnormality. Though biopsy will show the presence of tumor and its pathology, when doctors go for surgery, they must know the tumor extent and the exact location of it in the brain, which can be found by taking MRI scan of the patient, as it doesn’t involve the use of harmful radiations when compared to CT scan. A traditional method in hospitals is to segment the medical image manually, but this depends on how well the physicians can perceive the image, so they can get the required region extracted out, which is made difficult because of minute variations and resemblance between the original and affected biological part in the image. The shortage of radiologists and the large volume of MRI to be analyzed can make these readings labor intensive and cost expensive. It also depends on the expertise of the technician examining the images. Estimates also indicate that between 10% and 30% of tumors are missed by the radiologists during the routine screening. During the acquisition of medical images, there are possibilities that the medical image one gets might be degraded because of problems that can occur during the acquisition stage, making it so that the original image may not be suitable for analysis. Image segmentation also has some problems as well. This algorithm uses fuzzy C-means but the major drawback of this algorithm is the computational time required. Classifiers are also known as supervised methods since they require training data that are manually segmented and then used as references for automatically segmenting new data. The use of the same training data for classifying a large number of images, may lead to biased result. Supervised segmentation method requires considerable amount of training and testing data which comparatively complicates the process[13].

**SWIN Transformers for Semantic Segmentation**



**Encoder**

This image has been taken from the paper. It illustrates the architecture of Swin UNETR (U-Net Transformers). The input to the Swin model X is a token with a patch resolution of (H’, W’, D’) and dimension of H’ x W’ x D’ x S. We first utilize a patch partition layer to create a sequence of 3D tokens with dimension of [] x [] x [] and project them into an embedding space with dimension C. The self-attention is computed into non-overlapping windows that are created in the partitioning stage for efficient token interaction modeling. In the other figure taken from the paper below, we show the shifted windowing mechanism for subsequent layers.



Overview of the shifted windowing mechanism. Note that 8 × 8 × 8 3D tokens and 4 × 4 × 4 window size are illustrated.

Specifically, we utilize windows of size M x M x M to evenly partition a 3D token into [] x [] x [] regions at a given layer “L” in the transformer encoder. Subsequently, in layer L + 1, the partitioned window regions are shifted by ([], [], []) voxels. In subsequent layers of L and L + 1 in the encoder, the outputs are calculated as:

= W-MSA (LN ()) +

= MLP (LN ()) +

= SW-MSA (LN ()) +

= MLP (LN ()) +

Here, W-MSA and SW-MSA are regular and window partitioning multi-head self-attention modules respectively. and denote the outputs of W-MSA and SW-MSA. MLP and LN denote layer normalization and Multi-Layer Perceptron respectively. For efficient computation of the shifted window mechanism, we leverage a 3D cyclic-shifting and compute self-attention according to:

Attention (Q, K, V) = Softmax ()V.

In which Q, K, V denote queries, keys, and values respectively; d represents the size of the query and key. The Swin UNETR encoder has a patch size of 2 × 2 × 2 and a feature dimension of 2 × 2 × 2 × 4 = 32, considering the multi-modal MRI images with 4 channels. The size of the embedding space C is set to 48 in our encoder. Furthermore, the Swin UNETR encoder has 4 stages which comprise of 2 transformer blocks at each stage. Hence, the total number of layers in the encoder is L = 8. In stage 1, a linear embedding layer is utilized to create x x 3D tokens. To maintain the hierarchical structure of the encoder, a patch merging layer is utilized to decrease the resolution of feature representations by a factor of 2 at the end of each stage. In addition, a patch merging layer groups patches with resolution 2 × 2 × 2 and concatenates them, resulting in a 4C-dimensional feature embedding. The feature size of the representations is subsequently reduced to 2C with a linear layer. Stage 2, stage 3 and stage 4, with resolutions of x x x x and x x respectively, follow the same network design.

**Decoder**

Swin UNETR has a U-shaped network design in which the extracted feature representations of the encoder are used in the decoder via skip connections at each resolution. At each stage i (i ∈ {0, 1, 2, 3, 4}) in the encoder and the bottleneck (i = 5), the output feature representations are reshaped into size x x and fed into a residual block comprising of two 3 x 3 x 3 convolutional layers that are normalized by instance normalization layers. Subsequently, the resolution of the feature maps is increased by a factor of 2 using a deconvolutional layer and the outputs are concatenated with the outputs of the previous stage. The concatenated features are then fed into another residual block as previously described. The final segmentation outputs are computed by using a 1 × 1 × 1 convolutional layer and a sigmoid activation function.

**Loss function**

We use the soft Dice loss function which is computed in a voxel-wise manner as shown by the equation below:

L (G, Y) = 1 -

where I denotes voxels numbers; J is classes number; and denote the probability of output and one-hot encoded ground truth for class j at voxel i, respectively.

**Implementation Details**

Swin UNETR is implemented using PyTorch and MONAI and trained on a DGX-1 cluster with 8 NVIDIA V100 GPUs. The table below details the configurations of Swin UNETR architecture, number of parameters and FLOPs. The learning rate is set to 0.0008. We normalize all input images to have zero mean and unit standard deviation according to non-zero voxels. Random patches of 128 × 128 × 128 were cropped from 3D image volumes during training. We apply a random axis mirror-flip with a probability of 0.5 for all 3 axes.

