Immersive Visualisation In Medical Imaging

A Project Report

Submitted by

ADITYA SALIAN, E034

VVYOM SHAH, E039

SHLOK SHAH, E040

PARIN SHAH, E043

Under the Guidance of

PROF. KAMAL MISTRY

in partial fulfillment for the award of the

degree of

BACHELOR OF TECHNOLOGY

IN COMPUTER ENGINEERING

At

MUKESH PATEL SCHOOL OF TECHNOLOGY

MANAGEMENT AND ENGINEERING, NMIMS UNIVERSITY,

MUMBAI

MARCH 2020

DECLARATION

I, Aditya Salian, Roll No. E034 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.
2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who wrote what. (Source:IEEE, The institute, Dec. 2004)
4. We have made sure that all the ideas, expressions, graphs, diagrams, etc. that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.
5. We affirm that no portion of my work can be considered as plagiarism and we take full responsibility if such a complaint occurs. We understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidents of plagiarism in this body of work.

Signature:

Name: Aditya Salian

Roll No: E034

Place: Mumbai

Date: 28/03/2020

DECLARATION

I, Vvyom Shah, Roll No. E039 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who wrote what. (Source: IEEE, The institute, Dec. 2004)

4. We have made sure that all the ideas, expressions, graphs, diagrams, etc. that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. We affirm that no portion of my work can be considered as plagiarism and we take full responsibility if such a complaint occurs. We understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidents of plagiarism in this body of work.

Signature:

Name: Vvyom Shah

Roll No: E039

Place: Mumbai

Date: 28/03/2020

DECLARATION

I, Shlok Shah, Roll No. E040 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.
2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who wrote what. (Source: IEEE, The institute, Dec. 2004)
4. We have made sure that all the ideas, expressions, graphs, diagrams, etc. that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.
5. We affirm that no portion of my work can be considered as plagiarism and we take full responsibility if such a complaint occurs. We understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidents of plagiarism in this body of work.

Signature:

Name: Shlok Shah

Roll No: E040

Place: Mumbai

Date: 28/03/2020

DECLARATION

I, Parin Shah, Roll No. E043 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.
2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)
3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who wrote what. (Source: IEEE, The institute, Dec. 2004)
4. We have made sure that all the ideas, expressions, graphs, diagrams, etc. that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.
5. We affirm that no portion of my work can be considered as plagiarism and we take full responsibility if such a complaint occurs. We understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidents of plagiarism in this body of work.

Signature:

Name: Parin Shah

Roll No: E043

Place: Mumbai

Date: 28/03/2020

CERTIFICATE

This is to certify that the project entitled “Immersive Visualisation In Medical Imaging” is the bonafide work carried out by Aditya Salian, Vvyom Shah, Shlok Shah and Parin Shah of B.Tech (Computer Engineering), MPSTME (NMIMS), Mumbai, during the VIII semester of the academic year 2019-2020, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

Prof. Kamal Mistry

Internal Mentor

Examiner 1

Examiner 2

Dean

ACKNOWLEDGEMENTS

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report.

We would like to express our special thanks of gratitude to our mentor, Prof. Kamal Mistry who gave us this golden opportunity to do this wonderful project on the topic Immersive Visualization in Medical Imaging. It has helped us in doing a lot of research and we are really thankful to our mentor for the promising domains that we have explored in the course of this project.

We would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame. We are very grateful to those who did their best to help us with our work in this project.

Table of contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Meaning
AR	Augmented Reality
VR	Virtual Reality
XR	Extended Reality
ML	Machine Learning
CNN	Convolutional Neural Network
DSC	Dice Similarity Coefficient
VTK	The Visualization Toolkit
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
OCT	Optical Coherence Tomography
DICOM	Digital Imaging and Communications in Medicine
STL	Standard Triangle Language
2D	2 Dimensional
3D	3 Dimensional

Abstract

The average 5-year survival rate for individuals having been diagnosed with brain cancer is just under 35%. Over 18000 people will have died due to these brain tumours alone in 2019. These statistics show us how important it is for doctors to correctly identify and treat cases of brain cancer.

The second vertical that we want to deal with is the fact that most patients, once diagnosed, do not have a single clue what is exactly wrong with their body. This often leads to anxiety and stress, which can be harmful to them while already dealing with cancer.

Our project aims to leverage the hidden potential of MRI data available to us for detecting and segmenting a tumour automatically. Once segmented, its shape, size and location will be extracted, followed by a 3D projection in an Augmented Reality scene which can be viewed by both doctors and patients.

This is a proof of concept and once proven to be successful, can be implemented for various organs, and other parts of the body. Some such implementations can be for heart diseases, knee injuries.

Chapter 1

Introduction

1.1 Project Domain

Our project addresses the primary domains of Deep Learning, Augmented Reality(AR) and Biomedical Image Processing, which are supported by web technologies. Concepts of these domains have been used to implement this project.

1.1.1 Deep Learning

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised.

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases surpassing human expert performance.

In the medical imaging front, Deep Learning is being widely used to help identify, classify, and quantify patterns in medical images. Specifically, exploiting hierarchical feature representations learned solely from data, instead of handcrafted features mostly designed based on domain-specific knowledge, lies at the core of the advances. In that way, deep learning is rapidly proving to be the state-of-the-art foundation, achieving enhanced performances in various medical applications.

1.1.2 Augmented Reality

Augmented reality (AR) is an interactive experience of a real-world environment where the objects that reside in the real world are enhanced by computer-generated perceptual information, sometimes across multiple sensory modalities, including visual, auditory, haptic, somatosensory and olfactory. The experience so provided is seamlessly interwoven with the physical world such that it is perceived as an immersive aspect of the real environment. In this way, augmented reality alters one's ongoing perception of a real-world environment, whereas virtual reality completely replaces the user's real-world environment with a simulated one.

Augmented reality is used to enhance natural environments or situations and offer perceptually enriched experiences. With the help of advanced AR technologies (e.g. adding computer vision, incorporating AR cameras into smartphone applications and object recognition) the information about the surrounding real world of the user becomes interactive and digitally manipulated. Information about the environment and its objects is overlaid on the real world.

AR is being increasingly used in the healthcare industry as AR can help doctors access the latest and most relevant information about their patients. Moreover, patients can also use AR for self-education and improving the quality of treatment they receive. AR is a very useful tool for patient education, both for treatment and disease prevention. It can provide valuable information concerning a particular illness and treatment for patients and their family members. It is generally believed that human brain evolution involved development, learning and operation in multisensory environments, that is why it is considered that multi-sensory experiences, including AR, can be more effective in transmitting and processing information.

1.1.3 Biomedical Image Processing

Biomedical imaging deals with the capturing and rendering of images for diagnostic and therapeutic purposes. Biomedical imaging includes the analysis, enhancement and projection of images captured technologies like Computed Tomography (CT) scans, ultrasound, Magnetic Resonance Imaging (MRI) and Optical coherence tomography (OCT) to assess the current condition of an organ. Modern image reconstruction and modelling techniques have allowed the processing of 2D signals to create 3D images. The growth of domains like Deep Learning over time have given birth to several algorithms which perform spatial and temporal analysis to detect patterns and characteristics that serve as an indicator for anomalies like tumours.

Biomedical images are rendered on the principle that a beam of waves passing through the body under diagnosis will transmit or reflect back the radiation captured by a detector, which is later processed as an image pattern. However, the wave type differs for different modalities. For instance, Computed Tomography (CT) makes uses of X- Rays, gamma rays and radio frequencies are used for Magnetic Resonance Imaging (MRI) and Single Photon Emission Computed Tomography (SPECT), ultrasound frequencies are used for visualizing internal body structures in Ultrasonography. Biomedical image processing deals with performing feature extraction, morphological operations, segmentation and classification of the information retrieved from the images retrieved from the aforementioned techniques.

1.2 Problem Statement

"To identify extant anomalies within the organ with an aim to reproduce approximate size and relative location of anomaly and to generate a 3Dimensional model of a target organ with the anomaly based on scans of medical reports to visualize in an immersive environment through Augmented Reality."

1.3 Motivation

Our motivation for this project stems from a dire need to abridge the semantic gap between medical practitioners, and laymen like us. We observed that most medical documents like DICOM, CT and MRI scans constitute esoteric jargon which can only be elaborated upon by doctors. Dedicated to tackle this problem, we deliberated upon the potential of Augmented Reality to tackle this existing barrier, and how this technology can disrupt the medical industry because of the immersive experience that it can provide to any user, be it doctor, or the patient.

To comprehend the meaning of medical reports, Augmented Reality can be used to overcome the currently elusive linguistic explanation provided by doctors, and move to a much more graphically interactive solution which the patients can easily understand.

From the doctor's perspective we realized that surgeons are restricted to imagine a patient's anatomy from two dimensional views when most medical images are acquired as three dimensional volumes. Imagining a patient's anatomy and critical 3D interactions of a surgery using 2D images can lead to a non-optimal surgical approach and poor surgical results. Thus, to address this problem faced by the doctors, we plan on making use of this intuitive and interactive environment to provide a platform where medical practitioners can interact with patient specific anatomical features for improved diagnosis and a more thorough surgical planning.

1.4 Project Overview

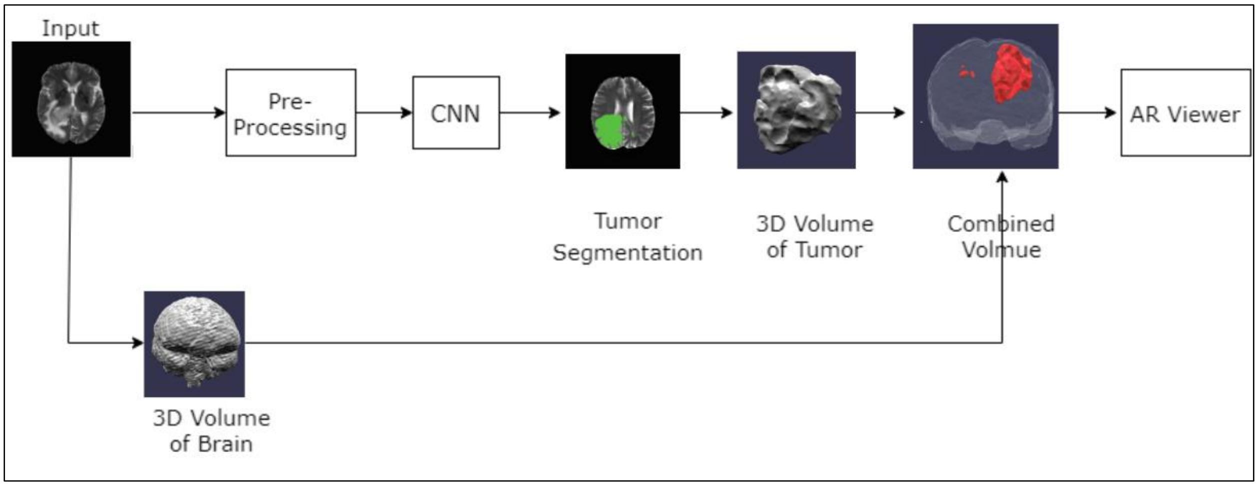


Fig 1.1 Pipeline depicting the overview of the project

Our project addresses the visualization of organs and their corresponding anomalies by making use of Augmented Reality. This decision of ours derives from the reason that augmented reality provides an intuitive and immersive experience to the user as it enables them to interact with the environment. This will enable the doctors to view the reports in 3D as they will not have to rely on 2D sliced images to visualize the target organ. The patients will also be able to interact with the generated model and thus, understand the cause-and-effect relationship of the presence of the anomaly on the target organ.

We have employed a Convolutional Neural Network (CNN) based on the U-Net architecture to work towards detecting and segmenting the anomaly on the target organ, i.e, the brain. Using the segmented anomaly (tumour in this case), we have rendered a 3D volume of the tumour as well as the target organ (brain) of each patient in the form of STL files. These files have been later converted into GLTF files for Android and USDZ files for iOS, to be viewed in augmented reality. The deployment of our project is in the form of a web application built on Flask, wherein the user can upload their MRI scans in the form of MHA files and view the resultant 3D model of their brain in augmented reality. A unique shareable link will be generated for every report, which will enable doctors to share the 3D report with their patients and vice versa. With these functionalities, we encourage the dissemination of information through an interactive and immersive medium provided in the form of augmented reality.

1.5 Project Scope

1.5.1 Collaborative medical operations

This project could be use enable multiple users to collaborate and work together on the model generated based on the report. Doctor-patient and doctor-doctor collaboration are two relevant use cases pertaining to this tool. With mixed reality, doctors are not bound by geographical limitations anymore. They can study the rendered model in real time right on their phone; thus enabling cross collaboration over the same model. This will benefit medical studies when dealing with special cases requiring expert opinions and multiple doctors to be involved.

1.5.2 Medical training

Students must be trained extensively before being allowed to operate or diagnose a patient. This extensive training can be very taxing on resources. By using a 3D model, one can work on specific models of organs so as to act as a substitute to operating on real physical organs. This use case is extremely relevant to trainees or medical students who require rigorous medical experience and training.

1.5.3 Patient awareness

A patient diagnosed with a serious medical condition is often riddled with anxiety due to their lack of ability to understand the cause and effect relationship of their corresponding disease with their bodies. Patients relying on doctors to help them understand this relationship is hindered by the semantic barrier between the two actors. Thus, we believe that a visual medium will address and bridge this semantic gap. By using a 3D model of the target organs, patients can be shown the exact shape, size and location of the anomalies with the target organs. This act itself will provide relief to the patient's suffering as they will be better informed about what is happening in their bodies and less anxious about their medical condition.

1.5.4 Quicker report sharing

A patient report once generated on the server, results in the generation of a unique URL for that report. This URL can be shared instantly over electronic medium enabling access to the report over any smart device. This medium of sharing reports is much quicker than other existing physical and traditional methods.

1.6 Hardware Specifications

RAM: 8 GB

Processor: 10th Generation i5

OS: Windows 10 64bit

Smartphone: ARCore or ARKit enabled

1.7 Software specifications

Python==3.6+

Blender==2.8.0

USD==19.11

Visualisation Toolkit (VTK==8.20)

UFG

SimpleITK==1.2.3

Tensorflow==1.13.1

Keras==2.2.0

Scikit-image==0.15.0

Chapter 2

Review of Literature

2.1 Image pre-processing

2.1.1 Preprocessing Technique For Enhancing The DICOM Kidney Images [1]

Abstract

Kidney carcinoma diagnostics proves to be one of demanding issues in medical imaging. Pre-processing is an essential first step required for diagnostic algorithms to function. It involves image quality enhancement, noise elimination and removal of background parts of the images of the kidney.

Summary

A five step pre-processing technique is proposed in the order of:

i. reading raw data,

ii. stretching the contrast,

iii. noise removal, and

iv. region of interest(ROI) cropping

before implementing global enhancements on the processed image.

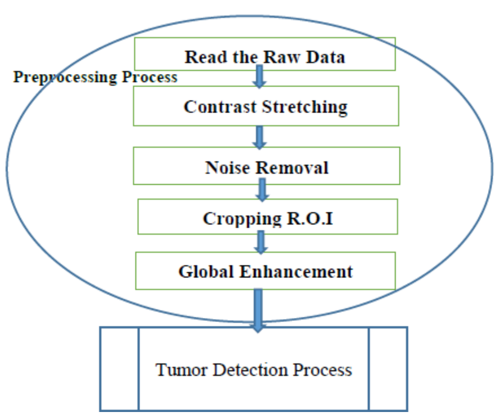


Fig. 2.1 Flow chart for preprocessing techniques

Digital Imaging and Communications in Medicine (DICOM) is the globally accepted file format for medical imaging upon which above said pre-processing techniques were implemented.

Contrast stretching is effective in highlighting the intensity differences between the object and the background, also resulting in sharper borders. Usually it is stretches the intensity values to desired range of values too.

Gaussian noise and speckle noise was the most abundant form of noise in the CT images read. The noise was filtered out using odd sixe masks whose each point is calculated by the Gaussian distribution function. A Wiener filter was also applied on the images as a pre-processing step.

ROI cropping was performed by cropping the images to fixed coordinates for every slice of the CT images. To perform global enhancement, an algorithm known as Contrast-limited adaptive histogram equalization is used, which operates on regions of the images, referred to as tiles. The histogram of each image is equated to the ideal histogram as computed by the 'Distribution parameter'. The above mentioned steps were found to be producing images of much better performance than other methods while at the same time retaining the structural details.

Conclusion

After applying quantifiable measurement parameters such as PSNR and MSE, the above mentioned steps were found to be producing images of much better performance than other methods while at the same time retaining the structural details.

2.1.2 Pre - Processing Technique for Brain Tumour Detection and Segmentation [2]

The main causes of image imperfections are as follows

1. Low resolution
2. Simulation
3. Presence of image artifacts
4. Geometric Distortion
5. Low contrast
6. High level of noise

Re-sampling is commonly used to produce better estimates of the intensity values for individual pixels. An estimate of the new brightness value that is closer to the new location is made by some mathematical re-sampling technique. Three sampling algorithms are commonly used are, Nearest Neighbor technique, the transformed pixel takes the value of the closest pixel in the pre-shifted array. In the Bilinear Interpolation approach, the average of the intensity values for the 4 pixels surrounding the transformed output pixel is used. The Cubic Convolution technique averages the 16 closest input pixels; this usually leads to the sharpest image.

Contrast enhancement is used to make the image brighter, to improve the visual details in the image. For noise removal, there are several denoising algorithms exists for noise removal each algorithm has its own advantages and disadvantages. Linear filters like Gaussian and wiener filters are conceptually simple, but they degrade the details and the edges of the images. Therefore the denoised image would be blurred. Markov Random Field method is robust against noise and preserves the fine details in the image, but Markov random field algorithm implementation is complex and time consuming. In the case of high redundancy images, using non – local methods we can remove the noise but it eliminate non- repeated details. Maximum likelihood estimation is another method of noise removal by adopting different hypotheses, but it does not retain the edge details.

Mathematical operations such as erosion, dilation, opening and closing are used to various ends throughout the pre-processing pipeline. Skull striping refers to the removal of non-brain structure and unwanted portions of image from scanned image to have the required image for tumour detection. Scanned image consists of brain area, scalp, skull and dura.

The unwanted portions can be separated with the aid of rim of cerebrospinal fluid (CSF). Skull removing can be done with the help of intensity thresholding followed by morphological operations to obtain required brain area for tumor detection.

2.2 Tumour detection

2.2.1 Review of Brain Tumor Detection from MRI [3]

Abstract

The paper reviews various brain tumour segmentation techniques that are implemented to detect brain tumours from MRI images. The authors establish that algorithms for the same exist but a big hurdle lies in calculating the location and size of the tumour.

Summary

Magnetic resonance Imagin(MRI) uses radio wave pulses and magnets to generate images of the body and its organs. MRI is the most perfect technique in radio because with the help of MRI we are able to visualize the details of internal structures. MRI images observe different soft tissues of the human body and is capable to contrast between these tissues.

Translating an image into a meaningful structure is referred to as image segmentation, in order to make it easier for analysis. This finds important use in the medical imaging fields where the image quality is not up to the mark. Algorithms of image segmentation work on two separate logics, image discontinuity and image similarity. Depending on the logic being implemented, either the edges and the corners intensity is changed or the image is partitioned into separate regions.

An accurate tumour detection in the brain is achieved by combining K-Means with a fuzzy C-means algorithm. This integration helps with optimal computation time and accuracy ratio.

Another approach used the PIGFCM algorithm for segmentation. This segments tissues in the brain into gray and white matter and cerebrospinal fluid. This helped detect abnormal tissues such as those of edema and necrosis. This algorithm is based on a fuzzy logic set of algorithms and can be integrated with them too.

2.2.2 Automatic Multimodal Brain Image Classification [4]

Abstract

Multilayer Perceptron (MLP) is identified as an enhanced method for brain image classification. Tumour location and tumour volume measurement procedures are also discussed. A supervised binary classification along with the MLP is used to filter tumour images from non-tumour images. Of the tumour images, the tumour part is separated out using anisotropic diffusion filters(ADF) after which the volume of the tumour is measured with the aid of the boundaries.

Summary

In image processing, feature extraction techniques are employed to extract the feature vectors from complex images. Various techniques provide image feature vectors which can be used to classify and segment the images into different types. Discrete Wavelet Transform(DWT) is one of the methods used for image classification and segmentation.

The system proposed is completely automated, involving four main stages which include feature extraction, classification and segmentation and then 3D modelling.

With feature extraction, DWT is utilised to extract features from the images and then the set of features is reduced based on coefficient values of features. After extraction, the MLP is applied to classify images as tumorous or non-tumorous after which the tumorous images are further enhanced and segmented. In the last stage, the relative and volume of the tumour is calculated and reconstructed.

Feature extraction processing involves applying wavelets for analysis of the image textures depending on their feature vector. The DWT segments the image into different frequencies of HH, HL LL and LH of which LL represents approximate coefficients and the rest represent the details.

The MLP algorithm is used for supervised learning to separate the tumorous and non-tumorous data from the input data set. It is comprised of an input layer, an intermediate hidden layer and the output layer.

The brain tumour segmentation is done by using ADF and active contour models. the image is dispersed using ADF technique, then centroid and origins of all the segments inside the image are used to remove the non-tumorous parts and detect the tumour segment. 2D ADF is used to offset the non-tumorous part. The diffused image is divided into different segments by using simple average threshold.

The proposed system introduces a new method which measures the volume of the brain and tumour very accurately and construct a 3D model based on the tumour volume and location. Brain tumour location is measured by proposing a new technique and dividing the whole brain into eight different parts; Front Right Side, Central Right Side, Back Right Side, Central Back Side, Back Left Side, Central Left Side, Front Left Side and Central Front Side. Fig. 3 describes the brain distribution into eight parts. The location of brain tumour and the coordinates of its centre is measured on the basis of this distribution.

The proposed techniques are tested on MICCAI BraTS 2015 data and an accuracy of 92.59% in the classification of MR images is observed. In tumor segmentation and volume calculations using the proposed methodologies, an accuracy of 90.12% is obtained.

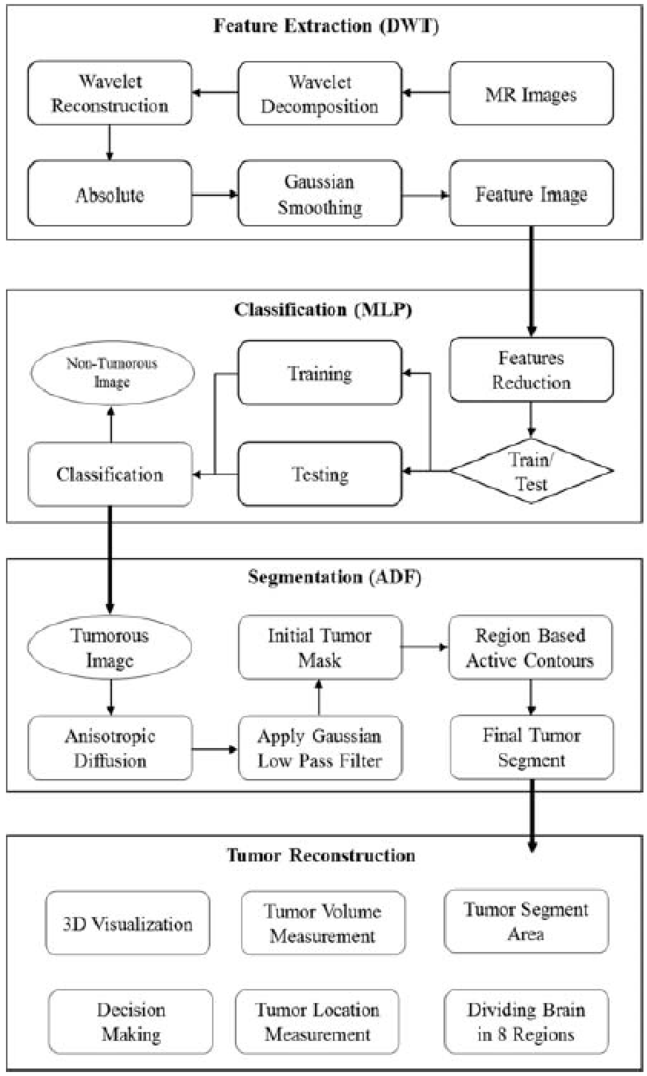


Fig. 2.2 Workflow process in proposed system

2.2.3 Performance analysis of various methods of tumour detection [5]

Brain tumour detection is primarily comprised of 2 parts:

1. Edge detection
2. Tumour segmentation

This paper applies both processes to find malignant or benign tumours based on their respective areas. Before applying various methods to achieve the above goals, the MRI images need to be pre-processed. Noise from the images must be removed for efficient and accurate detection. This is done through various kinds of filtering. There are various kinds of filters available in image processing for removing noise and each filter is having its own characteristics. Linear filters like Gaussian, averaging filters are one of the types used. For example, average filters are used to remove salt noise and pepper noise from the image. Because in average filter, pixel's values are replaced with their neighborhood values. Median filter also removes the noise like salt and pepper noise and also sharpens the image. Weighted average filter can also be implemented easily and give good results. In median filter, value of pixel is determined by calculating the median of the neighbouring pixels.

Followed by filtering, images must now be enhanced to further increase detection accuracy. Sharpening filters depends on spatial differentiation. Laplacian filter is a type of sharpening filter. After applying Laplacian filter to the given image, a new image is obtained which highlights edges and other discontinuities.

Edge detection phase now starts with applying one of the many edge detection masks available to us. For Sobel Edge Detection, Sobel edge detectors mask is used. That Sobel Mask is convolved with the matrix of original image. There are two Sobel masks as one for the horizontal component and the other for vertical component. These two Sobel masks are convolved with original image and finally the horizontal and vertical component images are added to obtain the edge detection image. The same technique is also applied for edge detection using Prewitt and Robert's Edge Detection.

Watershed segmentation can also be done on the basis of intensity values. As every pixel is having different intensity compared to the other pixel. Various watershed algorithms have been proposed nowadays. Once the tumour is segmented, its area can be calculated after eroding and dilating the tumour.

2.2.4 Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review [6]

Abstract

An extensive literature survey done on the effectiveness of CNNs in magnetic resonance imaging has been performed in this paper. This study has three objectives. The primary goalof this paper is to depict how different CNN architectures have evolved, study state-of-the-art strategies and then condense the results obtained using public datasets and examine their advantages and disadvantages. This paper also intends to be the reference of research of research activities in deep CNN for MRI related activities. The third objective of the paper is to suggest the directions of research for future CNN works.

Summary

This paper delineates the work flow working with CNNs for image analysis of 3D volumetric medical image data.

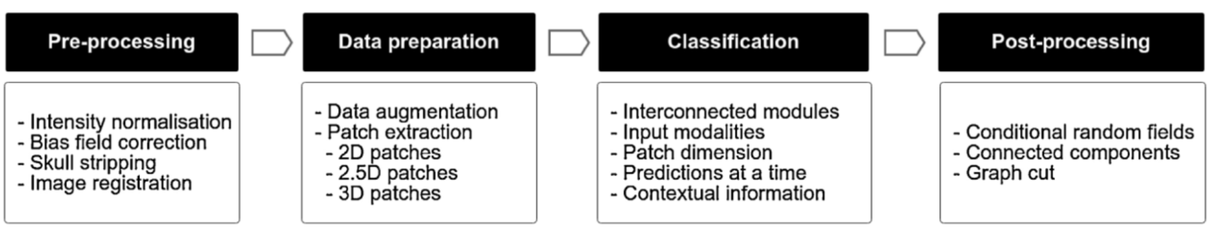


Fig. 2.3 Pipeline of a CNN based image analysis

This paper also classifies the architectures of various CNNs into 5 aspects-

1. Number of interconnected operating modules. On the basis of number of interconnected operating modules, CNNs an be classified into single-path and multi- path architectures. The single-path architectures correspond to the cases in which there is a unique flow of information: the input data is processed by convolutional, pooling and non-linear rectifier layers; the feature maps are then mined in the FC layers, and afterwards used for predicting the label in the output layer. The parallel multi-path architectures are composed of different CNNs

designed to operate in parallel to capture more comprehensive features. Each network uses different versions of the same target area.

2. Number of input modalities. The categories of input modalities are two: single and multi-modality. The former case is certainly more adaptable to different scenarios as a single modality is commonly provided in datasets for tissue and sub-cortical structure segmentation (mainly T1). The latter case contemplates processing different sources of information.

3. Input patch dimension of the network. The cate- gories, in this case, are three: 2D, 2.5D and 3D. Two-dimensional architectures consider patches from a single plane (axial, sagittal or coronal). This architecture is widely used in the literature since it is adaptable to different image domains and segmentation tasks. Tri-planar or 2.5D architectures are provided with patches from the three anatomical planes (i.e. axial, sagittal and coronal), commonly using a multi-path design. The 2.5D information provides a better understanding of the 3D scenario than 2D-based networks since it exploits the 3D nature of MR images and, consequently, brings up contextual information. Three-dimension architectures consist in taking 3D segments di- rectly from the MRI volume. This architecture utilises 3D convolution kernels, which seems to be a more appropriate solution for fully ex- ploiting the spatial contextual information in volumetric data. The main constraint of a 3D CNN approach lies in its expensive computational cost, memory requirements and computational time.

4. Number of predictions at a time. The approaches can also be grouped according to the number of predictions they perform at a time into CNN and FCNN. CNNs correspond to the traditional approach in which a single patch is processed by the network, and a single output is re- turned. Since the idea at the end is to obtain segmented areas, this approach can be slow in practice. FCNNs correspond to the

architectures for which the fully connected layers are replaced by 1×1-kernel (or 1×1×1-kernel) convolutional layers to obtain a dense prediction.

5. Implicit and explicit contextual information. Although some implicit contextual information is encoded in 2D, 2.5D and 3D patches their information is limited to the size of the patches. The bigger the size of the patch, the more information the network can take into account to produce the prediction but the more the parameters to be trained.

2.2.5 Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [7]

Abstract

In this paper, a 3d CNN is proposed which is 11 layers deep for brain segmentation. To overcome the computational burden of processing 3D medical scans, an efficient and effective dense training scheme was devised in this paper which connects the processing of adjacent image patches into one pass through the network while automatically adapting to the inherent class imbalance present in the data. A top ranking performance on MICCAI 2015 dataset was achieved with this model.

Summary

The paper makes the following contributions-

1. An efficient hybrid training scheme is proposed which utilizes dense training. Large training batch sizes are required for training but they result in an increase in memory requirement and computation time. This problem is overcome by using an image greater than the receptive field of the CNN is given as an input which outputs a posterior probability with multiple voxels. Thus, this problem is overcome in this paper.

2. Taking advantage of the utilization of small kernels to yield a much more discriminative, yet computationally expensive 3d CNN. Smaller kernels of 3X3X3 size are used as they are faster to convolve with, and contain less weights.

3. Parallel convolutional pathways for multi-scale processing are used. This provides as a solution to incorporate both local and contextual information which improves segmentation results. In order to incorporate both local and larger contextual information into the CNN, a second pathway is added that operates on down-sampled images. Thus, the dual pathway 3D CNN simultaneously processes the input image at multiple scales. The detailed local appearance of structures is captured in the first pathway, whereas higher level features such as the location within the brain are learned in the second pathway.

2.2.6 Brain Tumor Detection and Segmentation in MR Images Using Deep Learning [8]

Abstract

A deep learning-based method that uses different modalities of MRI is presented for the segmentation of brain tumor. The proposed hybrid convolutional neural network architecture uses patch-based approach and takes both local and contextual information into account, while predicting output label. The proposed method contains a preprocessing step, in which images are normalized and bias field corrected, a feed-forward pass through a CNN and a post-processing step, which is used to remove small false positives around the skull portion. Over-fitting is overcome by utilizing dropout regularizer alongside batch normalization, whereas data imbalance problem is dealt with by using two-phase training procedure.

Summary

The paper makes the following contributions-

1. A fully automated algorithm for brain tumor segmentation using deep learning.

2. Combining a two path parallel network with 3 path network to form hybrid network. This hybrid network is both efficient, and effective. In the two path network, the first stream has small kernels to get local information from the image, whereas the second stream has kernels with large receptive field, focusing more on contextual information. 490 features are extracted in the final layer of this network. This four layered CNN is comparatively faster to the 3 path CNN, which uses five layers. A dropout of 0.5 is used to tackle overfitting. The hybrid network in total contains nine convolution layers, but it is still efficient as most of the processing is done in parallel.

3. Local and global information of the image is taken into account while predicting a segmentation label for a pixel. Hence the output of the model is based on contextual and local features.

4. Analyzing the utility and validity of advanced regularization techniques like dropout and novel two phase training.

Preprocessing of data in this paper is done by N4ITK bias field correction to account for bias field distortions and to neutralize the motion heterogeneity which is caused due to the movement of subjects during scans. In post processing, morphological operations are used to remove false positives around the edges of the predicted output. Erosion is used to remove false positive, and then dilation operation is performed to enlarge the output to its original size.

2.2.7 A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scans [9]

Abstract

Deep neural networks have been widely adopted for automatic organ segmentation from abdominal CT scans. However, the segmentation accuracy of some small organs (e.g., the pancreas) is sometimes below satisfaction, arguably because deep networks are easily disrupted by the complex and variable background regions which occupies a large fraction

of the input volume. In this paper, a fixed-point model is developed, which uses a predicted segmentation mask to shrink the input region. This is motivated by the fact that a smaller input region often leads to more accurate segmentation.

Summary

This paper highlights the segmentation of relatively smaller objects presents in the medical images. Pancreas consist of less than 0.5 percentage of the entire medical image, and it is observed that FCNNs yield less accurate results when segmenting such small organs because it is easily disrupted by the background regions. Thus, a predicted segmentation mask is used to shrink the region of input. The iterative shrinking of the input region can be best depicted using the following diagram from the paper.

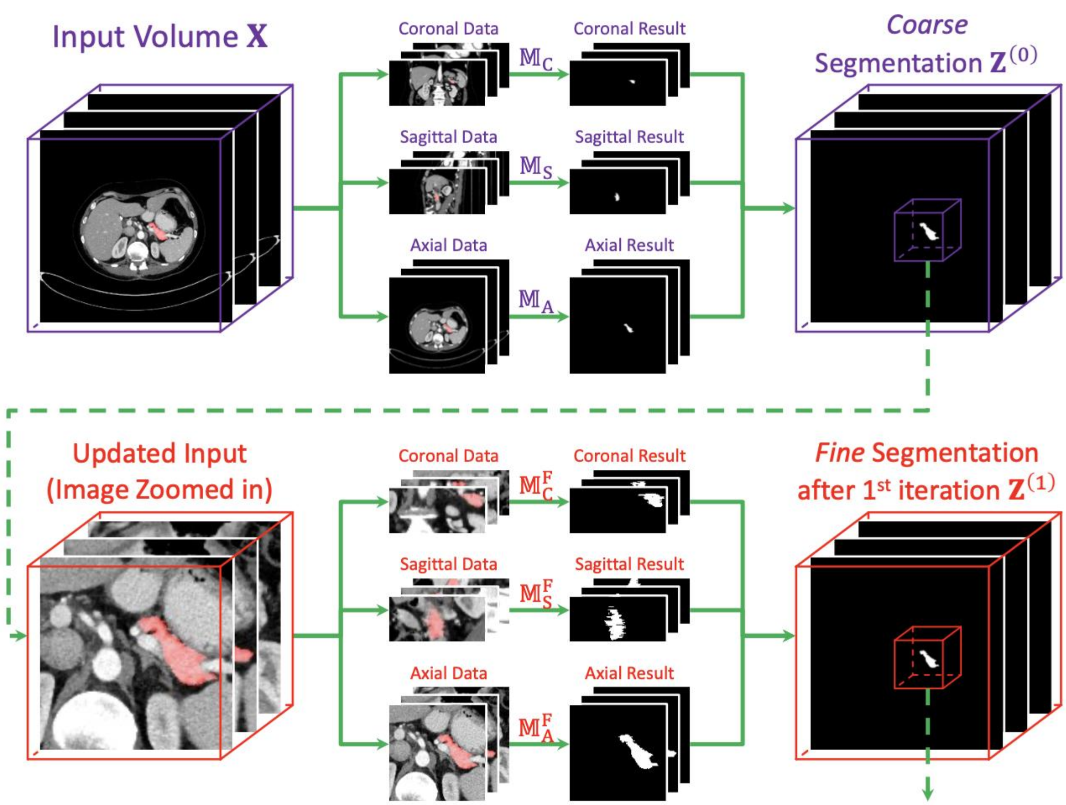


Fig 2.4 Illustration of testing process, 1 iteration shown

The paper also proposes an algorithm called Fixed-Point Model for Segmentation. In this algorithm, the image volume is read sliced along the coronal, sagittal and axial planes in the first iteration and the predicted mask is sliced. Then the smallest 2D box to cover all predicted foreground pixels in each slice, is found and a 30-pixel-wide frame around it is added. This process is repeated iteratively until a fixed number of iterations T is reached, or the similarity between successive segmentation results is larger than a certain specified threshold.

2.2.8 A semi-automatic method for segmentation and 3D modeling of glioma tumors from brain MRI [10]

Abstract

An efficient way to volumetric render glioma tumours that have undergone segmentation from the two dimensional MRI scans by replacing manual segmenting with user interactive control automation. To automatically segment the tumour portions, morphological filters were utilised on T1 FLAIR and T2 modalities of the MRI scan. A software package was used through which a 3D model of the tumour with accurate tumour border is generated.

Summary

To extract and visualise objects in three dimension is one important step for diagnosis, planning and delivery of treatment. The varying nature of tumour shape, size, location and texture makes its segmentation a challenging task. Along with morphological tools, MATLAB and 3D Doctor software package is utilised. The dataset used in this process was confirmed to have the presence of the tumour abnormality.

Skull stripping as a step of pre-processing was performed. With the skull stripping, parts of the skull and intracranial tissues that do not belong to the brain, example skin, muscles, fat etc were removed. As part of skull stripping, erosion and dilation functions are implemented in order to maintain the integrity of the boundaries of the brain. This was followed by intensity normalisation and thresholding to develop a binary image, where non-tumour pixels were labelled 0 and rest sustained their intensity value.

To generate a 3D model, an empty space is chalked out where each tumour pixels (x, y) coordinate pair is marked. For the z coordinate, the slice number representing the distance between two slices is assumed. Adjoining pixels in the 3D space are connected together and this process repeats until all slices are traversed. The generated model is then verified using 3D DOCTOR.

An advantage of this method is that it preserves the shape and gray levels of the original image scans which helps in texture analysis and classification. Another advantage is that it is an automatic process which allows interactive user control, unlike manual process that were followed earlier by doctors.

2.2.9 Detection and 3d Reconstruction of Brain Tumor From Brain [11]

Abstract

This paper not only develops method to segment brain tumour but also calculates volume of the tumour along with its shape and size in three dimension.

Summary

This paper utilizes different image processing methods to detect brain tumours. After inputting the images, they are passed through a high pass filter using a fspecial function. The histogram is then equalized and thresholding is performed to distinguish objects from their background. Morphological functions such as erosion, is then performed leading to segmentation through connected component labelling. Volume 3D creates the volume render object from input data using the vol3d() function that is based on texture mapping technique in openGL.

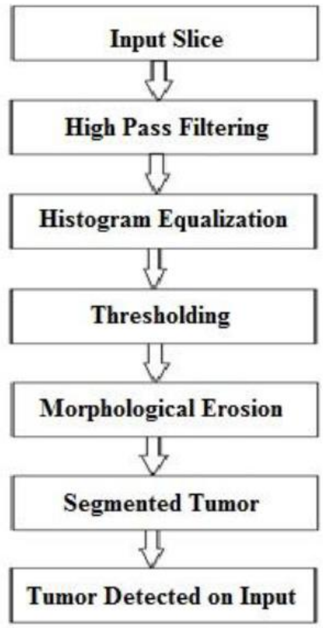


Fig 2.5 Flowchart for tumour detection

2.3 Augmented Reality application

2.3.1 Augmented Reality for Breast Tumours Visualization [12]

Abstract

This system uses a marker-based image-processing technique to render a 3D model of the tumours on the body. Mammograms are considered as one of the best screening tools for years, but the 2D view of them is not as effective as 3D views. The question is: “Is there a better way of showing 3D views of the breast internal composition, especially lumps and its tumours to doctors so that they can have a better visualization?” This question is answered by proposing an MB-AR application that shows a 3D model of the breast tumours on a plastic model of the body. Such visualization of breast tumours is very important for doctors, oncologists, and surgeons. In addition, better visualization can lead to a better screening, less false positive and avoiding unnecessary biopsy. It can also be helpful for planning and navigation during surgery.

Summary

The system has four modules. The overall system takes the video from the scene and it produces overlay images on top of real video by rendering the prepared 3D model. The four modules in order are: 1. Pre-processing, 2. Marker finding, 3, Pose Estimator, Graphic generation.

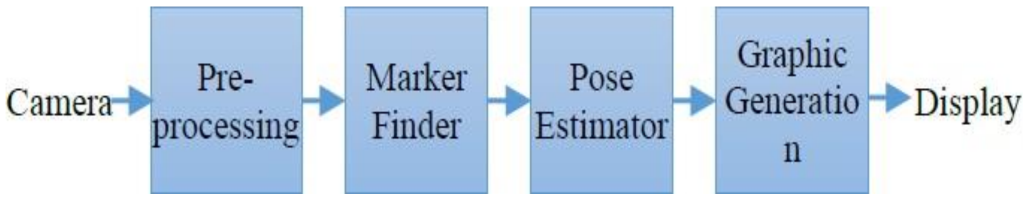


Fig. 2.6 Architecture of the system

Pre-processing

The camera's image is the input to this module and it produces a binary image as output. The first step is converting the input image to a grayscale image. The next step is converting the inverted grayscale image to a binary image.

Marker finder

A marker is a predefined object in the scene which can be identified easily. Usually, markers are asymmetric rectangular black and white object which are added to an object in the scene. Since they are black and white; they can achieve the maximum contrast in an image.

Markers are defined as squares because it is easier to define squares with polygons and find them. The important point is that the selected pattern should not be symmetric so that the algorithm can find the rotation of the marker.

The first step in finding markers is finding the all closed contours in the binary image. This is done by topological structural analysis algorithm. For detecting the marker, there are two steps: (i) obtain the frontal view of the rectangle by removing the perspective projection, and (ii) search for the marker pattern in the frontal view of the polygon. The pattern of each marker is represented by a matrix of 1 (for white and 0 (for blacks).

Position Estimator

The last step is estimating the angle that camera is looking at the scene. By extracting this information, the 3D model can be rendered in the correct angle. This problem includes estimating 3D information using 2D information. It is preferred to work with the homogeneous coordinate system instead of Cartesian coordinate system when estimating the homography transform. Suppose $q_c = [x, y, z]^t$ denotes position of a point in 3D Cartesian system; then, its homogenous equivalent is obtained as $q_h = [x, y, z, 1]^t$

Graphic Generation

This module renders a pre-stored 3D model and adds texture to the model for displaying.

2.3.2 Augmented Reality Based Brain Tumor 3D Visualization [13]

Abstract

A 3D brain tumour visualization method is proposed in this paper that functions in real time by using facial features as markers of the subject in the scene. A new method of camera calibration using the size of the face of the subject is implemented and the pose is computed by processing 3D data and its 2D projections. Using these computations, the reconstructed brain tumour is displayed on the subjects' face.

Summary

When 2D images are used to portray anatomical structures, mental mapping is needed to parallel the anatomy of the body. Another added issue is the reversal of the left-right orientation in medical imaging because of which mapping can become difficult. Hence, with an AR application, an intuitive guide is developed to visualize the anatomy of humans.

A marker-less system is proposed by the paper, as placing markers on patient bodies is demeaning and irresponsible. In marker-less tracking, the camera pose is determined from naturally occurring features, such as edges and textures of an anatomical object (e.g. eye corners, nose-tip). A pre-constructed model of the brain will be imposed on the patient body.

The proposed model of the system consists of three components: Camera calibration, Pose estimation and augmentation.

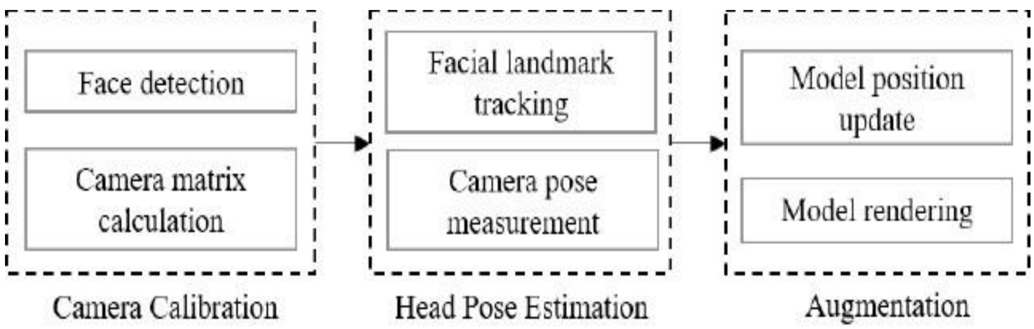


Fig. 2.7 Architecture of proposed system

Camera Calibration is based on Tsai algorithm and the Viola-Jones face detector. The most challenging task for pose estimation is reference point recognition. The five reference points looked up are the ends of the eyes, ends of the lips, tip of the nose and tip of the chin of the subject's face; using Dlib's facial landmark detector. The 3D points of the corresponding reference points are used to compute their 2d equivalent values of the face. For the face pose estimation.

The 3D model of the brain is rendered onto the scene using Unity. A difference in colour and in material is used to identify parts of the skull and soft tissue of the brain.

2.3.3 Augmented Reality In Laparoscopic Surgical Oncology [14]

Abstract

This paper realizes how Augmented Reality can increase the surgeon's intra-operative vision by providing a virtual transparency of the patient. AR in minimally invasive surgery techniques is divided into two main processes: the 3D visualization of the anatomical or pathological structures appearing in the medical image, and the registration of this visualization on the real patient. 3D visualizations can be performed directly from the medical image without the need for a pre-processing step due to volumetric rendering. However, better results are obtained with surface rendering after organ and pathology delineations and 3D modelling. Registration can be performed interactively or automatically.

Summary

The 3d visualization of patients is achieved by using pre-operative data like MRI and CT scan reports. This visualization has two types- volumetric rendering and surface rendering.

Volumetric rendering is computed automatically from raw DICOM data of 3D medical CT images. The method consists in visualizing simultaneously all voxels of the selected images by replacing their initial density by a colour and a degree of transparency defined through a transfer function. This transparency allows therefore to view or to hide contrasted anatomical or pathological structures in 3D. The main benefit of this 3D visualization is the lack of preprocessing, no delineation being required to obtain useful results for diagnosis and surgical planning. However, there are several limitations that can be prohibitive depending on the medical application. Firstly, without manual interaction users cannot visualize independently the organs they want to see if they have the same gray level in the DICOM data. Secondly, it is not possible to compute any organ volume since independent structures are not delineated. For the same reason, it is not possible to simulate an organ resection without cutting neighbouring structures.

Surface rendering, on the other hand needs first a structure of interest delineation followed by a mesh generation processing providing the surface of delineated structures. This preprocessing is the main drawback of the technique due to the possible difficulty and duration of such delineation. Once all important organs and pathologies have been segmented, it is possible to visualize them in 3D, to change their surface transparency and to compute their volume. The initial information being reduced to important structures to be visualized only, the rendering is faster on a usual computer.

Registration methods are the most difficult to incorporate as they require the superimposition of additional information on the correct image and in the correct step.

Any discrepancy in the registration process means that the guidance information is wrong and can lead to a dangerous surgical movement. Therefore, the registration accuracy is one main concern in AR guiding systems to ensure safety. There are two types of registration techniques-

1. Interactive registration augmented reality- Interactive augmented reality (IAR) systems generally use a standard screen associated to a camera view or light projection on patient. The first method to perform the registration is to simul- taneously visualize on the same image the patient and the pre- operative 3D model, and to manually modify its position, scale and orientation so that landmarks which are visible in both the 3D model and the real patient view are properly superimposed. The second interactive method is also based on landmarks that can be identified on both pre-operative 3D model and real

patient. Users have firstly to determine the 3D position of these landmarks in the preoperative image (using a medical imaging software). Secondly, they need to point intra-operatively the corresponding landmarks visible on the patient using a tracked pointer. This technique does not allow for real-time modification of the registration, the pointing process being too long to apply. This technique should then be used essentially for fixed structures having no movement during surgery. The results of the interactive registration augmented reality can be viewed in the following diagram-

Fig. 2.8 External camera based display IAR uses visible landmarks such as ribs and iliac crest in abdominal surgery (left) or clavicle and sternum in neck surgery

2. Automatic registration augmented reality- In order to reach full automatic augmented reality capability in endoscopic surgery, two main issues have to be considered: physiological organ (or patient) motion and surgical instrument inter- actions on organs. Moreover, in case AR information is provided on a view from a moving camera or display (typically an HMD or an endoscopic camera), it is usually requested to track their position/ point of view. To provide automatic AR, real- time the localization of the structures of interest need to be known. There are two approaches to tackle this issue.

The first method is to intra-operatively acquire a 3D image of the patient zone which contains the structures of interest, and to render this image to allow AR visualization. The 3D image acquisition can be performed using either 2D or 3D US probe or MRI device.

The second approach consists in registering a pre-operative 3D image which contains all important structures on intra-operative information acquired either by the endoscopic camera or another imaging modality that can be available in the OP room.

2.4 Virtual Reality application

2.4.1 Virtual Reality based Immersive Visualization for DT-MRI Volumes [15]

Abstract

This paper developed a Virtual Reality environment to visualize volumetric datasets acquired via DT-MRI. They developed an environment with two application areas in mind namely studying changes in white-matter structures after gamma-knife capsulotomy and pre-operative planning for brain tumor surgery. After the successful building of the environment, the feedback that they received suggested that their developed system helps the user better interpret the large and complex geometric models, and facilitates communication among a group of users as compared to desktop displays.

Introduction

Datasets produced by DT-MRI are volumetric, with six values at each spatial location, hence there is a significant challenge to visualize and understand them. Thus there is a need for a tool that can help visualize the internal structures. The resulting models developed by these images is complex and difficult to interpret using traditional single-screen displays. An immersive virtual environment has significant advantages. The improvement in about 200% when compared to the static images.

VR Setup

The implemented environment is a CAVE environment that consists of an 8 x 8 x 8 foot with rear-projected front and sidewalls. The user wears a pair of LCD shutter glasses. The system was used with the goal of making the user feel comfortable and allowing the user to get immersed in the visualization. In this system, the model remains stationary and the user moves around allowing him to view the model from different angles and ways. This allow the user to interact more naturally with the models than while using a monitor.

Data Preparation

They displayed the DT-MRI and T2 weighted images of the human brain. They used tricubic B-spline functions to interpolate the second-order diffusion tensor field and then interpolated the eigenvector field by calculating the eigenvector of the interpolated tensor at any location.

Geometric Models of DT-MRI Data

They generated stream tubes and stream surfaces to represent linear and planar structures. To increase interactivity, they selected a subset of the possible geometric models and decimated them for display.

Performance Considerations

There was a need to balance between fidelity and interactivity. There has to be a proper tradeoff between the details displayed and the rendering time for the best interactivity. They found out that the values of frames per second and the number of polygons had to be adjusted based on different scenarios and the type of users interacting with the interface.

Interactions

Models stay still in the system, the user walks around and adjusts his head position to observe the data from various perspectives. The user can use the wand to indicate the region of interest and change the sections via buttons on the wand. Feedback: Three MDs and two senior medical students used the environment and provide the following feedback. To begin with they suggested that the system is easy to use and it provides a better understanding of the data. Secondly, it took only a few minutes for the users to get used to the environment and the features. Thirdly, the user moved around to view the visualization from different perspectives and once the user found the area of interest he explored the model in detail by standing still.

Conclusion

This paper provided us with a definitive answer that such a system is suitable and provides better interactivity to the doctors and allows them to visualize the data easily. The methods described in the paper suggested that movement is key as it allows the user to feel natural while interacting with the system and thus makes the overall experience better.

2.4.2 GPU Based High-Quality Volume Rendering for Virtual Environment [16]

Abstract

Visualization of medical data sets acquired by CT and MRI has become important as it offers a three-dimensional view of the underlying data. The paper developed an enhanced virtual reality system based on GPU based ray-casting. It focused on seamless integration of perspective high-quality direct volume rendering into virtual reality.

Introduction

The main concern of medical visualization is to aid doctors in their daily work. VR requires a high level of interactivity so that the user has a feeling of being immersed in the system and can manipulate the model in a natural way. Image quality also plays an important role in any kind of medical diagnosis.

GPU Ray-Casting

Ray-casting renders an image of a volume by casting rays from the viewpoint through each pixel, and performing resampling and compositing of successive sample contributions along with these rays. GPU Ray-Casting allows the enhanced and fast generation of these volumes.

VR Interaction with the Volume Data

The main aim of the system was to help doctors in medical diagnosis and surgical planning. Furthermore, allowing effective collaboration of the users was key. The slice of the medical data can only be interpreted by experts whereas the volume-rendered model can be interpreted by non-experts also. Navigation was achieved by allowing the user to move freely around in space. Clipping of the model allowed the user to view the internal structure of the organ. Lighting intensity can be changed according to the users' discretion.

Hardware Setup

The system consisted of a Barco Baron Table along with shutter glasses and an ART Optical Tracking System with three cameras and 0 a single PC running the hybrid application.

Conclusion

The paper achieved a seamless integration of the volumetric data into the polygonal rendering of the VE by calculating pixel accurate intersections of both in image space. They used a variety of methods to increase user interaction and the preliminary demonstrations resulted in positive feedback.

2.4.3 VR Techniques in Multi-Field Volume Rendering [17]

Abstract

The paper proposes a new approach to visualizing multi-field MRI or CT datasets in an immersive environment. There are two steps being performed namely classification and exploration. Classification is done at desktop whereas the exploration step is done on a powerwall.

Introduction

Direct volume rendering is important and flexible technique for visualizing 3D volumetric data. An immersive environment with head and hand tracking enhances the visualization by providing the user with a truly 3D view of the dataset. The immersive environment with its high spatial acuity allows for collaborative surgical planning among multiple domain experts.

Collaborative Immersive environment

Three important criteria were kept in mind while developing the system. First, collaboration is key for successful visualization. Second, user interfaces must be carefully designed to meet the users' key requirements. Finally, ergonomic factors are an important issue for immersive environments. The system thus designed accommodated multiple users using two large stereoscopic displays. This workspace allows users to interact with the visualization using the most appropriate modality.

Immersive Volume Rendering

This task was divided into two separate tasks namely classification and optical property specification. The classification step involves identifying the regions of the data domain, or feature space, that correspond to unique materials or material boundaries. Once these regions have been determined, all that remains for the user to do is assign color and opacity, making the classified materials corresponding to features of interest visible and unimportant materials transparent.

Implementation

Hardware used consists of two rear-projected passive stereo displays and a table in front of them. Software is built on top of several existing visualization and scientific frameworks.

Focus

The focus of this study was two-fold. First, they intended to demonstrate the effectiveness of multi-modal MRI data classification using multi-dimensional transfer functions. Second, they intended to identify the ways in which their immersive visualization system can assist in diagnosis and treatment planning.

Conclusion

The paper presents a novel immersive visualization workspace layout that emphasizes tight collaboration between domain and visualization experts. The results so achieved show that the immersive environment allows the user to interpret the data easily and effectively.

2.4.4 Immersive Visualization Environment for the Exploration of Medical Volumetric Data [18]

Abstract

This paper describes a real-time visualization system for analyzing medical volumetric data in various virtual environments. They use 3D texture hardware for accelerating the rendering process. Direct volume rendering is used for visualizing the details of medical volumetric data sets without intermediate geometric representation.

Introduction

The current practice of doctors viewing the medical slices in front of white diffusing light is commonplace and accepted but it does not show the complete data provided by the systems. Virtual Reality systems can solve the problem of volume rendering. The paper develops a cross-platform medical data visualizer that is capable of dealing with various display modalities. It incorporates a GPU based acceleration techniques of volume rendering. They also develop a user interface for a virtual environment. The whole system is developed using the object-oriented programming paradigm that allows for portability.

System Design and Implementation

MedVis has a desktop version and a CAVE environment.

Volume rendering is used as it bypasses the intermediate geometric representation and directly renders the volumetric data set based on its scalar value. This allows a radiologist to visualize the fine details of medical data as he/she is the one who by law must make the final interpretation.

Software Architecture

DICOM images are read by the reader and passed to the renderer. The interactor handles the user input and exchanges the states with the renderer to adjust the rendered images.

Fig. 2.9 Software Architecture of MedVis environment.

Conclusion

This paper presents the Medical Visualizer, VR system for visualizing volumetric medical data in various VR systems ranging from non-immersive to immersive systems. It develops a tool in a modular way so that new components can be added.

Table 2.1 Advantages and Disadvantages of the VR based solutions.

Paper Title	**Advantages**	**Limitations**
An immersive virtual environment for DT-MRI volume visualization applications: a case study	• Interactive environment that allows multiple users at once. • Performance improvement of 200% as compared to static models.	• Navigation through the model requires physical movement. • CAVE hardware including projectors and headsets required
Medical Applications of Multi-field Volume Rendering and VR Techniques	• Allows to view the spatial features head and hand movement. • Domain experts can interact with each other	• Projectors, 3D displays and headsets are required. • Feature extraction is not possible.
MedVis: A Real-Time Immersive Visualization Environment for the Exploration of Medical Volumetric Data	• Modular application thus adding new features or using different environment is easy. • Ease of interaction. • Real time response by changing the color of region of interest.	• CAVE hardware required that includes a projector, headsets and wand.
GPU based high quality volume rendering for virtual environment.	• Hand tracking and medical instruments tracking is implemented • No prior segmentation or pre- processing is required.	• Interaction between various users not possible. • HMD are required

Chapter 3

Analysis & Design

3.1 Review of Various Architectures for Segmentation

We have conducted a thorough literature survey to determine the architecture most suitable for our project. This endeavour has led us to compare and contrast the performance of various different architectures on the MICCAI BRATS dataset. With the performance metric being the Dice Score, this table illustrated in table 3.1 enlists the performance of various architectures on the MICCAI BRATS 2018 dataset.

Table 3.1 Review of the DSC of Various Architecture on the MICCAI BRATS Dataset

Architecture	**Whole Tumour**	**Tumour Core**	**Enhanced Tumour**
3D CNN	0.84	0.66	0.63
CRF-RNN	0.84	0.73	0.62
FCN	0.86	0.73	0.62
ResNet	0.85	0.77	0.64
U-Net	**0.88**	**0.81**	**0.76**

3.2 Architecture

The architecture can be divided into two parts as shown in Fig 3.1, the server and the web app. The web app primarily serves as the interface for the user, where files are uploaded and the resultant 3D model can be viewed. It also serves as the gateway to create an AR instance on

ARCore and ARKit enabled smartphone devices, to view the same 3D model in Augmented Reality.

After accepting the input from the user, the image files are passed to the server which has multiple different components. First the image files are passed to a Convoluted Neural Network with a U-Net architecture, details of which have been explained below. The output of this module is the segmented tumour, which is then passed to the surface mesh creation unit. The surface mesh unit renders volume to the 2D slices of the tumour and to the brain. These volume rendered objects are passed to the blender API to generate a merged tumour and brain model, with accurate size, location and orientation adjustments. This final model is then rendered and passed back to the web app where it is displayed. Each model has a unique URL attached to it which can be shared for viewing over any number of devices.

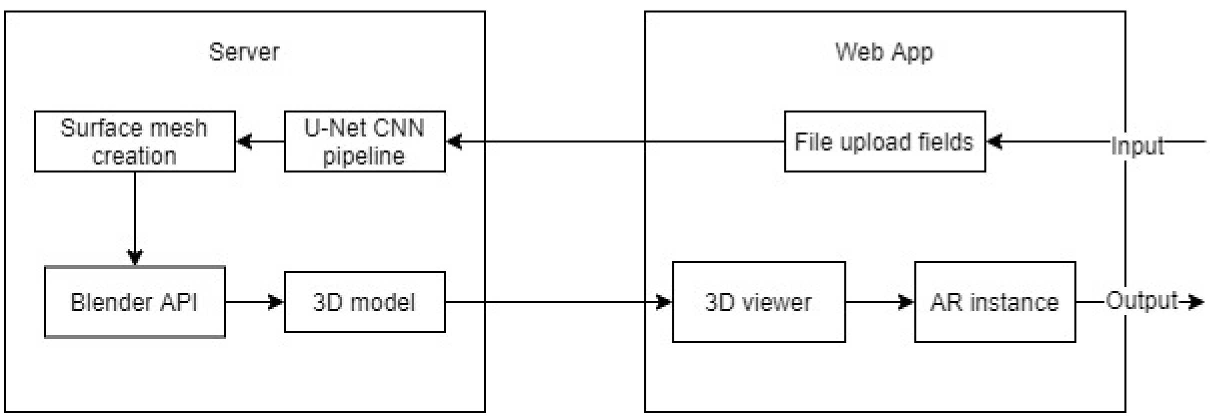


Fig 3.1 Architecture of the system

3.3 UML diagrams

3.3.1 Activity Diagram

As seen in Fig 3.2, the process begins with the user uploading .MHA files into the system which goes into the CNN. As a result, the segmented tumour is stored as a numpy array. It is necessary to convert it back to .MHA file format in order to perform volumetric rendering using the VTK module. The corresponding object we obtain is a 3D surface mesh in .STL format. This is loaded into the blender API where we perform texture and orientation correction. These are then exported as .gltf and .usdz file formats which are required for viewing in AR on Android and iOS powered mobile devices, respectively.

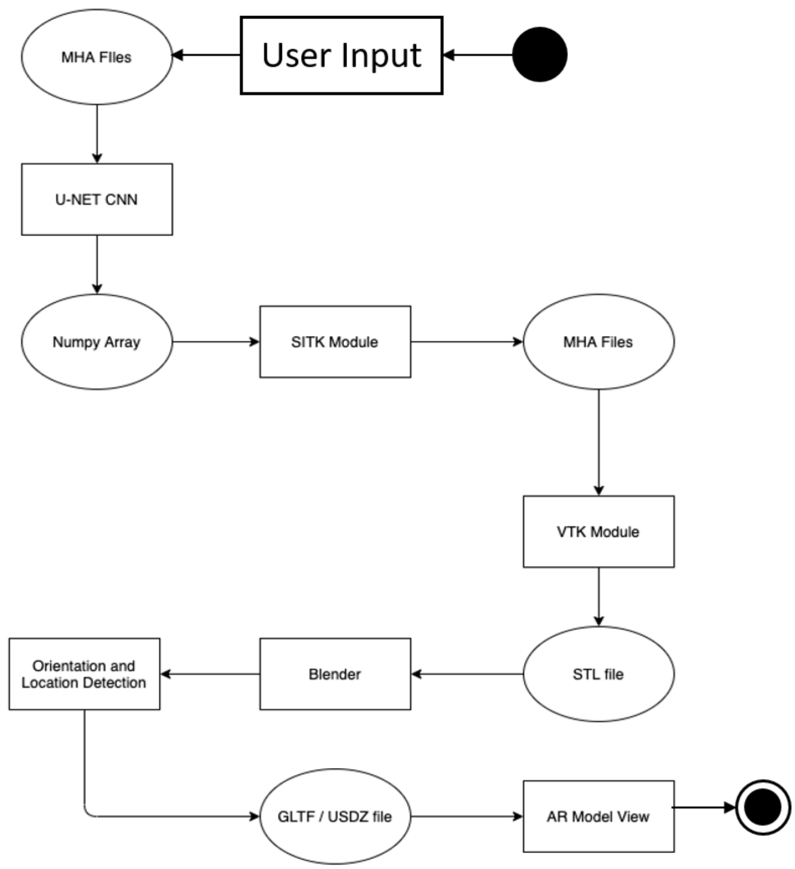


Fig 3.2 Activity Diagram of implemented system

3.3.2 Use Case Diagram

The diagram depicted in Figure 3.3 depicts the use case diagram of our project. The primary actors of our application are the doctor and patients, and the secondary application of our system is the server. All use cases of this project have been enlisted in the diagram in a chronological order. The primary actors can interact with Reports3D to upload their MRI files, download the 3D model rendered by the web app and interact with the model using their mobile handheld devices. The secondary actor, which is the server, uses our web application to preprocess the MRI file uploaded by the primary actor, segment the tumour from the brain and render a 3D volume of the tumour and brain for the primary actors to download and interact with.

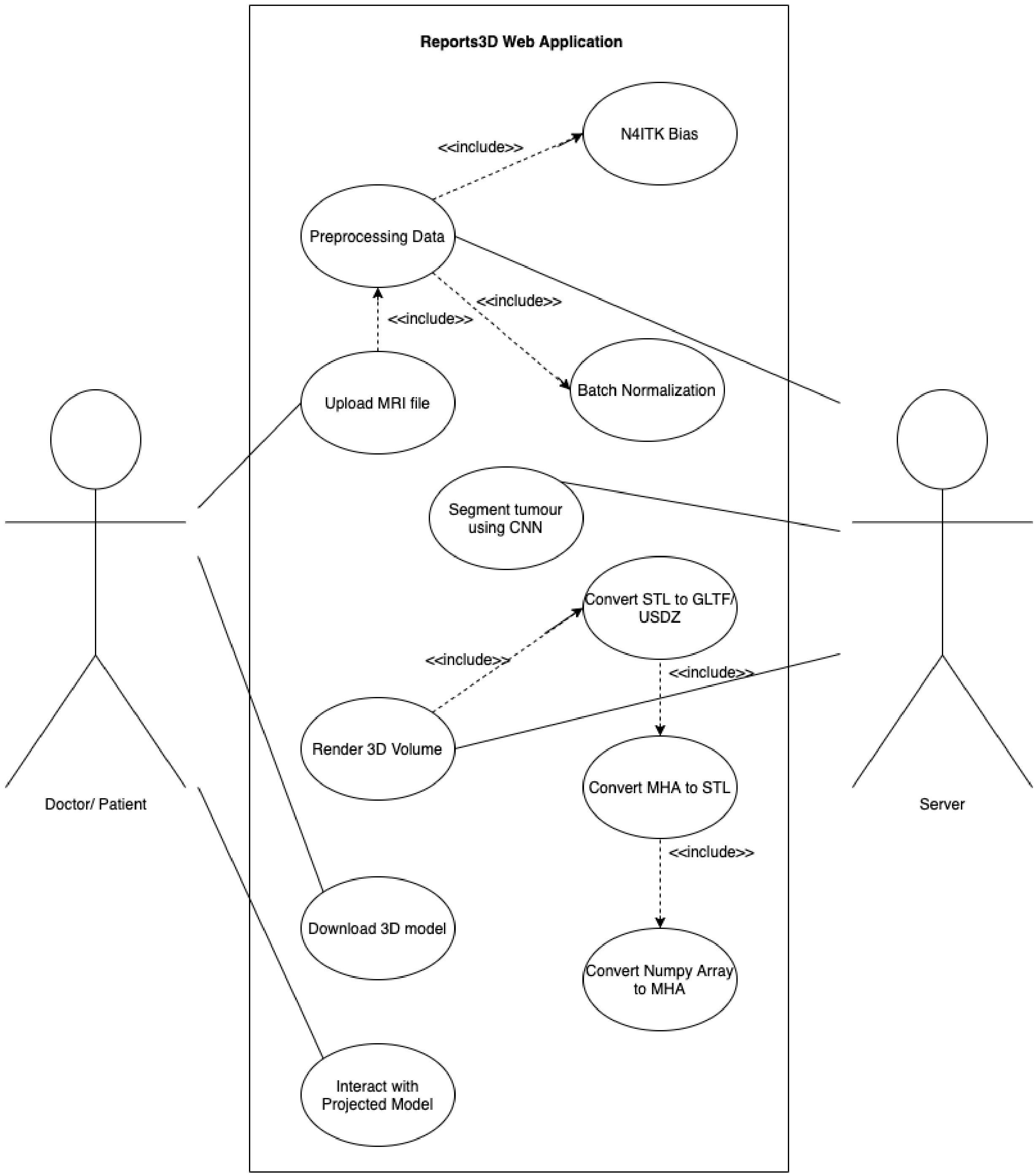


Fig 3.3 Use Case Diagram of implemented system

3.4 User implementation

The primary target users for the app are doctors who will be using this system to help patients visualise their medical ailments and abnormalities in a graphic 3D manner. The proposed flow of using the app is as follows:

1. User uploads medical scan images on the web interface
2. The file is sent to the server for processing
3. Server returns with 3D model of the tumour embedded in brain for each patient, a unique resultant URL is generated.
4. 3D model is displayed in the web app, with allowances to rotate, translate and scale.
5. After being granted device permissions, the 3D model can be viewed in AR directly from the web browser.
6. The aforementioned URL can then be shared universally and viewed on any web browser.

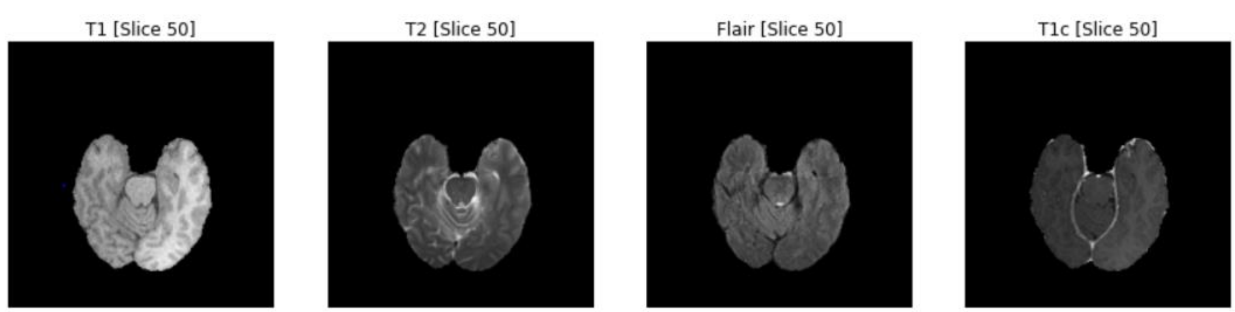
Chapter 4

Implementation

4.1 Procuring Dataset for Target Organ

Our vision to build immersive and interactive reports for medical reports was channeled into developing an initial proof of concept model for the same. After due deliberation, we decided upon using the brain as our target organ. Thus, with the brain as our target organ, we have decided to use the MICCAI BRATS Dataset of 2018 to segment tumours and construct models from.

The MICCAI BRATS Dataset of 2018 contains clinically acquired pre-operative multimodal MRI scans of glioblastoma (HGG) and low grade glioma (LGG) in the form of training, testing and validation datasets. The MRI scans in the dataset was available to us as NIfTI files (.nii.gz) which describe 4 modalities- a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. Moreover, the overall survival data, defined in days, is included in a .csv file with correspondences to the pseudo-identifiers of the imaging data which also includes the age of patients, as well as the resection status. The four modalities can be seen in figure 4.1



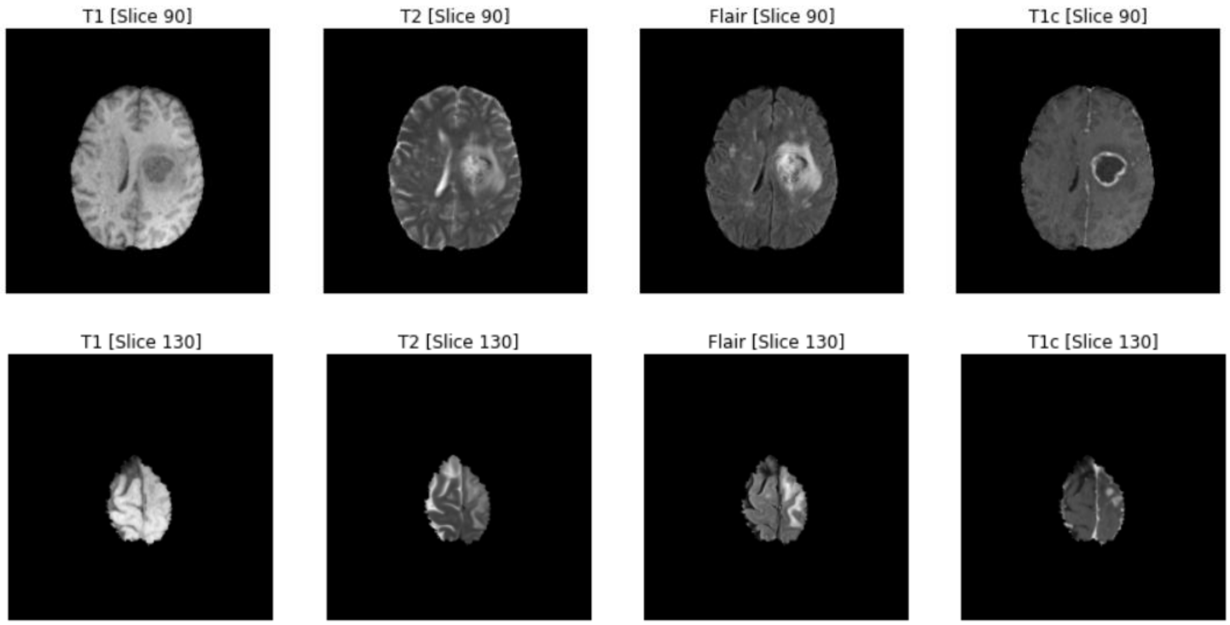


Fig 4.1 Depiction of four modalities of a particular patient from the dataset

4.2 Preprocessing

The dataset procured from MICCAI BRATS 2018 is delivered with skull stripping performed already on all files. With skull stripping, the bone part along with the soft tissue, skin, muscle etc. is removed from the images. This makes tumour detection much easier.

It has been observed that the intensity values across MRI slices vary greatly, therefore normalization is applied in addition to the bias field correction to bring mean intensity value and variance close to zero and one respectively; observed in figure 4.2.

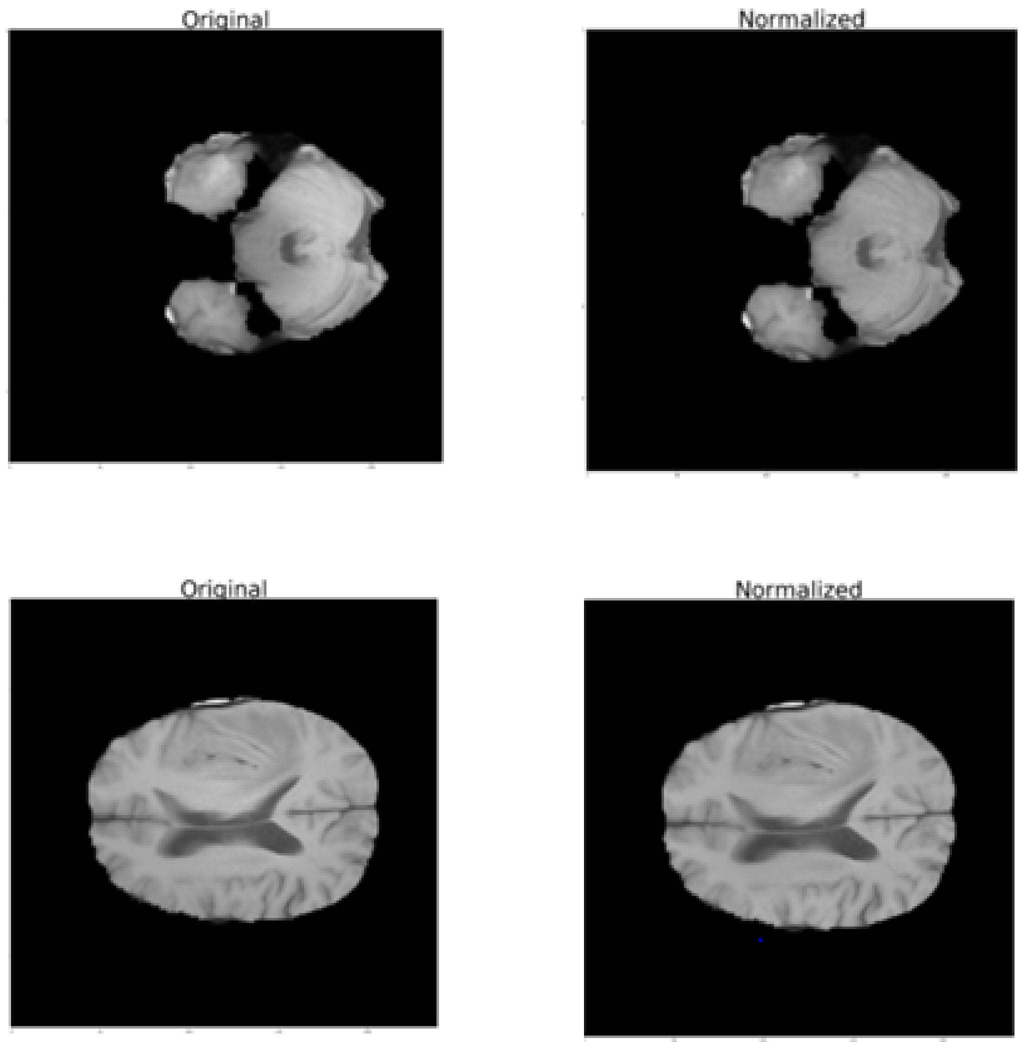


Fig 4.2 Difference in the slices pre and post normalization

No additional steps for resampling of images have been carried out since the dataset of images falls within acceptable image size limits of the U-Net architecture that has been implemented. Otherwise too, resampling of data should be done carefully at the cost of data loss.

One important step in analyzing an MRI scan is to remove its N4 Bias, a low frequency current field in an MRI machine that causes intensity variations causing the machine to be misled into incorrectly segmenting parts of the brain. This N4ITK bias correction has been accounted and accommodated for in the U-Net architecture that has been implemented.

4.3 Segmentation of Tumour

We have employed unsupervised learning in the form of CNNs in this project for the detection of anomalies because the large variability, location and size, shape and frequency of the tumour make it difficult to devise segmentation rules for supervised learning. Due to this high variability, the most accurate segmentation of these anomalies can be achieved if they are segmented manually by humans. We realized the power of CNNs in pattern recognition, and thus, decided to employ CNNs to detect anomalies by training them on the MICCAI BRATS 2018 dataset procured by us.

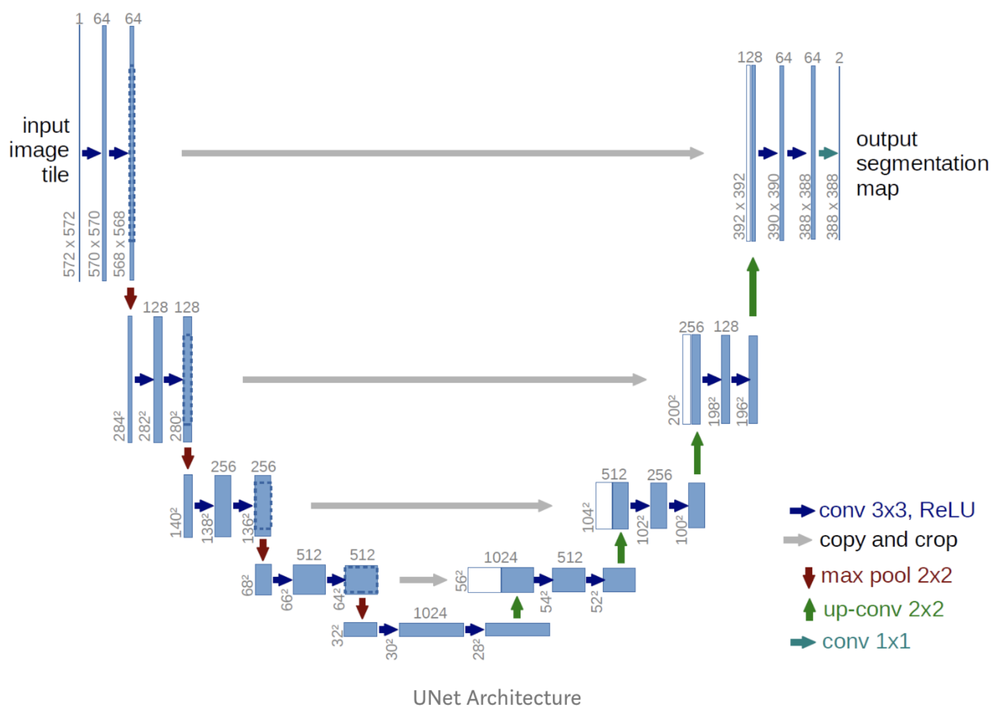


Fig. 4.3 U-Net Architecture

Our network is based on the U-Net architecture, which consists of two major parts- the encoding or contracting path; and the decoding or expansive path, as seen in Fig 4.3. The contracting path consists of the traditional convolutional process whereas the expansive path consists of transposed 2D convolutional layers. We have added a batch normalization layer after every convolutional layer.

The contraction path comprises several contraction blocks, with each block taking an input and applying two 3X3 convolution layers followed by a 2X2 max pooling. The number of feature maps doubles with every layer which enables the network to learn more complex structures effectively. The bottommost layer of the contracting path is the intermediate stage between the contracting and the expansive paths, and it uses a 3X3 convolutional layer followed by a 2X2 up convolutional layer.

The expansive path, similar to the contracting path, has several expansion blocks, with each block passing the input to two 3X3 CNN layers followed by a 2X2 upsampling layer. In juxtaposition to the contracting path, the number of feature maps in this path are halved after every block. Additionally, every time the input is also appended by feature maps of the corresponding contraction layer. This action would ensure that the features that are learned while contracting the image will be used to reconstruct it.

Initially, we used the U-net model to perform the segmentation of the full tumour, tumour core and the enhancing tumour. We got promising results for the segmentation of the full tumour, however our results for segmenting the tumour core and the enhancing tumour were not satisfactory. After due deliberation, we diagnosed that this is shortcoming was due to the fact that the tumour core and the enhancing tumour are too small in comparison to the entire brain (0.75% and 0.45% per slice respectively). To address this situation, we calculated the center point of the full tumour and then cropped out the training data for tumour core and enhancing tumour using the training data. Figure 4.4 depicts the shape and size of our predicted tumour in juxtaposition with the ground truth. We have used the T2 and Flair modalities for full tumour segmentation instead of using all the modalities while training, which has significantly reduced our training time. Thus, we have two input channels as we use only T2 and Flair modalities for full tumour segmentation.

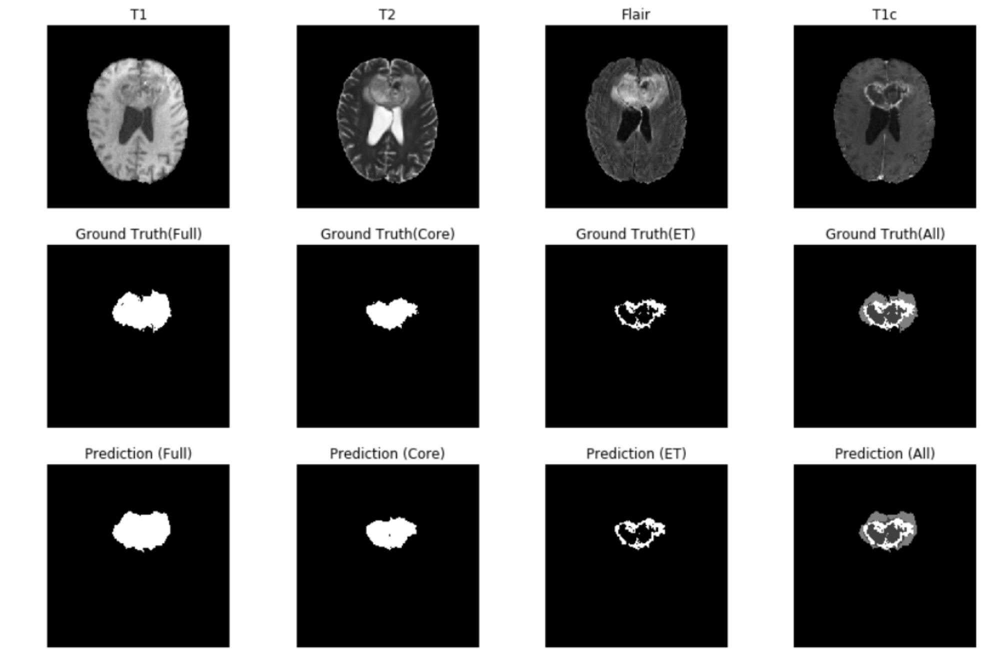


Fig. 4.4 Comparison of the ground truth with predicted images by our network

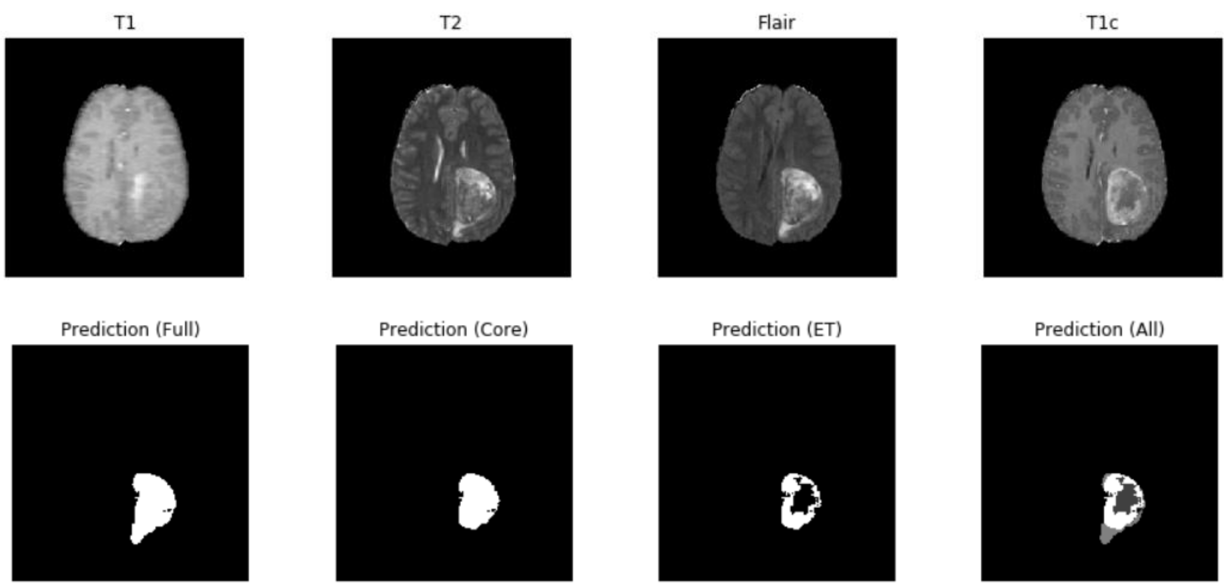


Fig. 4.5 Result on test set by our network

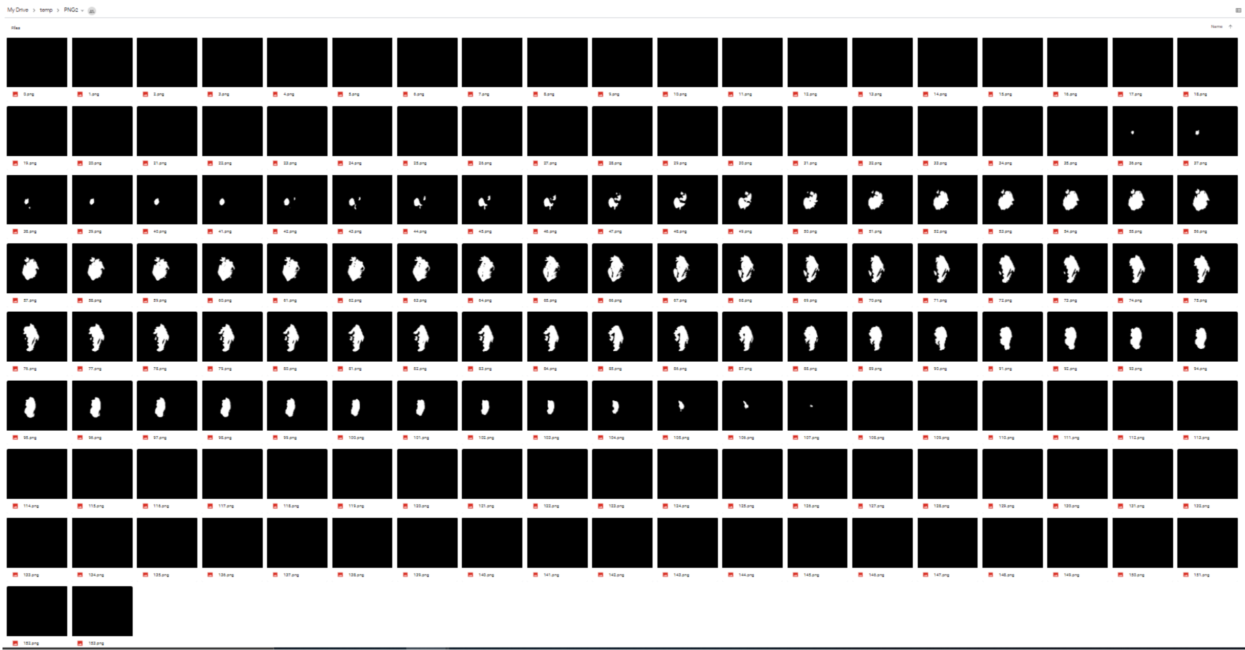


Fig. 4.6 Screenshot of folder containing images of segmented slices by the network

4.4 3-Dimensional volumetric rendering

Once we have the tumour segmented by the UNet model that is presented in the previous section, we focus on creating a 3D model that is to be displayed. The model so created is unique to each patient. This 3D model needed to be in two different formats namely GLTF (for android users) and USDZ (for iOS users). The detailed steps performed for the creation of the model are enlisted below:

1. Numpy array to MHA
   The UNet model performs segmentation on the given input file and the result is a NumPy array. We convert this NumPy array to an MHA file using the SITK write image module. The result is an MHA file consisting of the segmented tumour.

2. MHA to STL

   Once we have the MHA file, we create surface meshes for both the detected tumour and the patient's brain. A surface mesh stores boundary information about the 3D model. This is stored in the form of an STL file. These make sure that every 3D model is uniquely created for every patient. The end result of this step is two STL files, the first one being the tumour and the second one being the brain of the patient. This is done using the following algorithm.

   FlyingEdges3D is an implementation of the 3D version of the flying edges algorithm. It is designed to be highly scalable for large data. It is a 4 pass algorithm and is available as a part of the VTK library. After creation of the surface mesh, we have applied a smoothening filter to remove rough edges generated by interlocking polygons of the mesh.

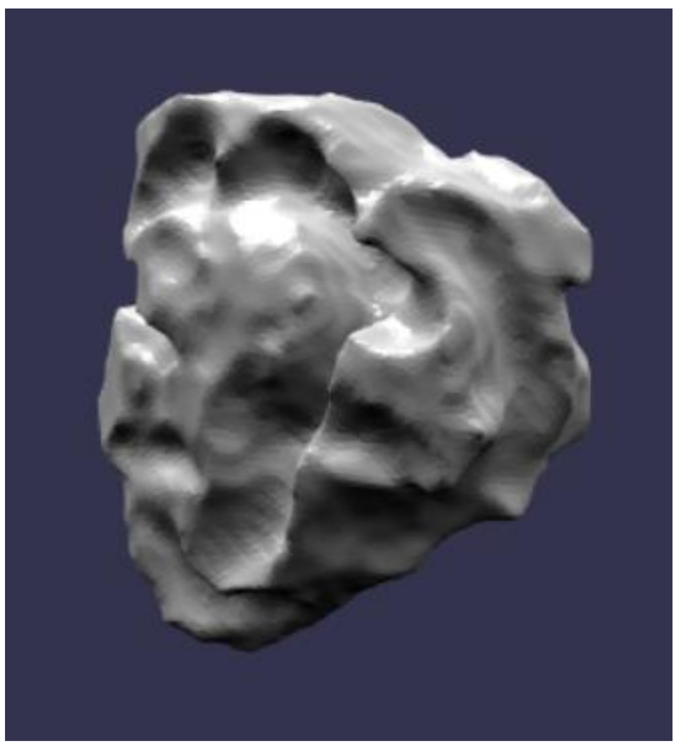


Fig 4.7 STL file of segmented tumour

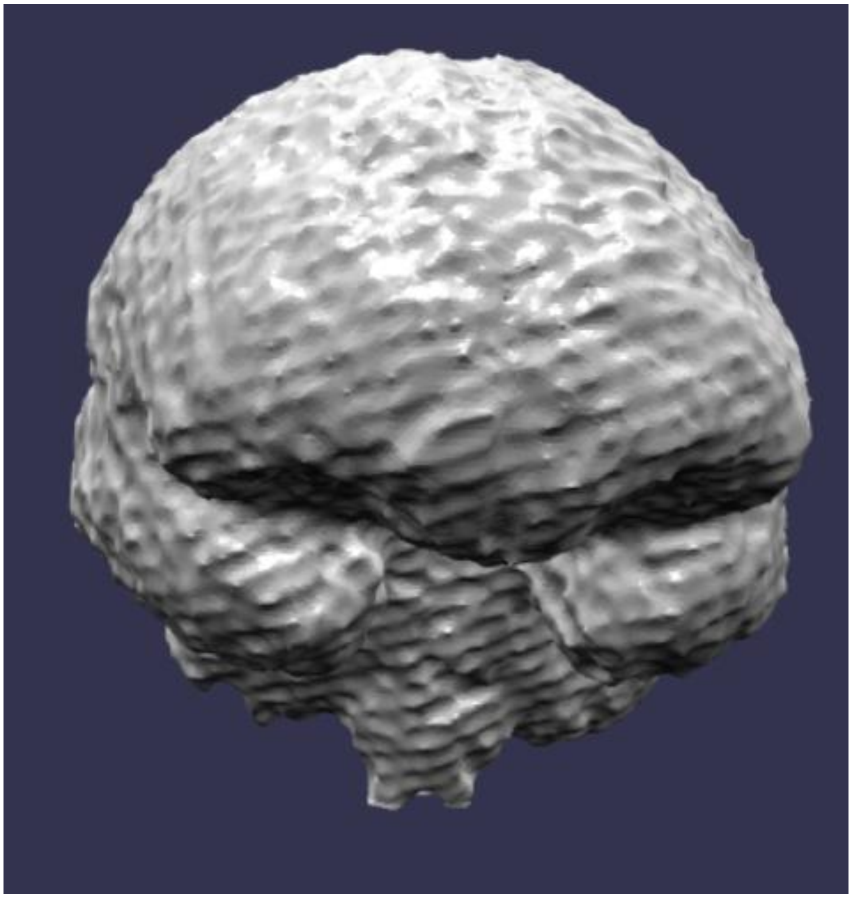


Fig 4.8 STL file of brain

3. STL files to Blender

   Once we have the two STL files, we use the Bender API to combine these two STL files into a single file that can be displayed. The major hurdle encountered here was how do we place the tumour at the exact location inside the brain so as to get an accurate representation. After much discussion, we came up with a novel method to achieve near-perfect accuracy. We added planes to the MHA files of the brain and tumour so as to represent origin. Next, we superimposed both of these planes to combine both STL files. In doing so the resulting combined model had close to perfect placement of tumour in the patient's brain.

   As the surface meshes created go beyond 600,000+ polygons, it was imperative for us to decimate and reduce them to about 10% of the original count. This is because the modelviewer module cannot handle high poly-count data very well and lags if given anything above 60000 polygons.

The brain is being decimated to about 15000 polygons, while the tumor is not being decimated as it is the main focus from a doctor’s POV.

Once we combined the STL files we exported the same into GLTF format to be used for displaying in AR.

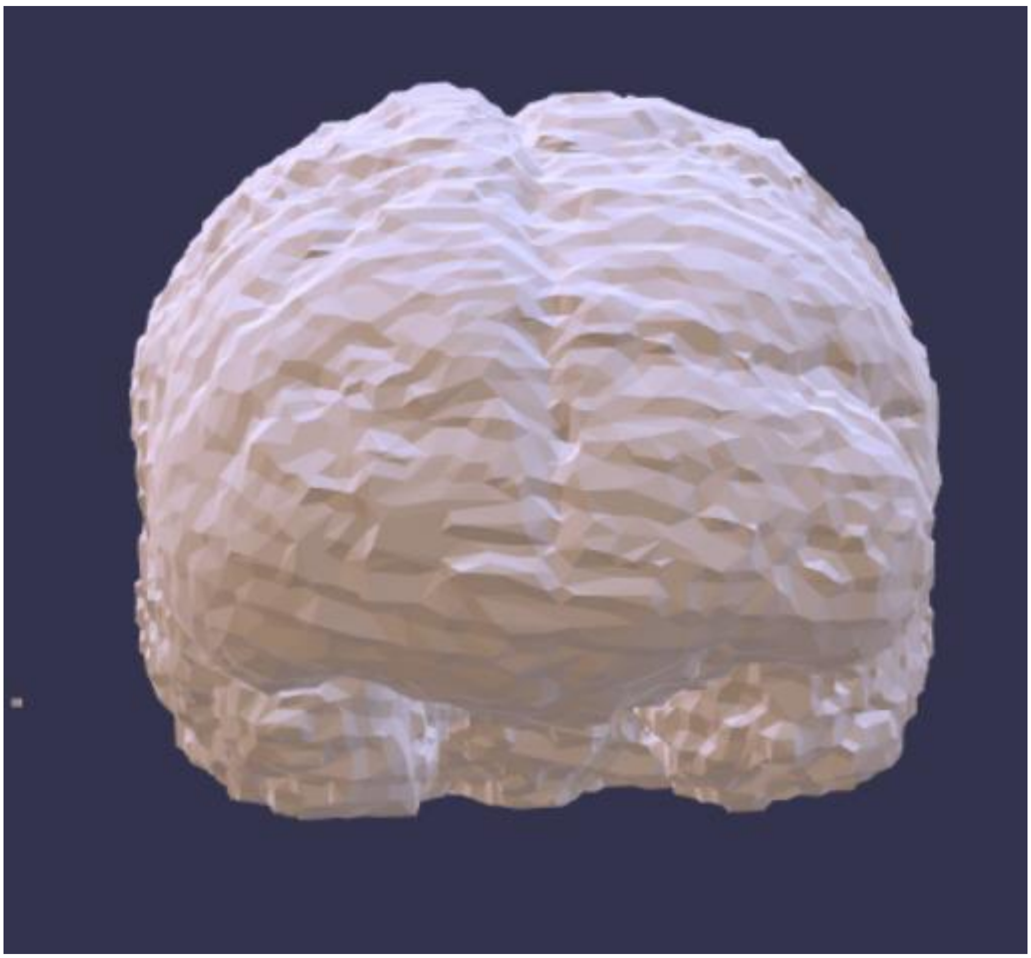


Fig 4.9 GLTF file of the patient before editing

4. GLTF editing

Once we exported the GLTF file from Blender API, we noticed that the file did not maintain its transparency and hence we had to edit it manually. The goal was to convert the outer brain mesh to near-transparent so that one can easily view the tumour inside the brain’s outer structure.

GLTF file is essentially a JSON format and we set the alpha value of the brain to 0.1 which allowed us to achieve the desired result. For better viewing experience we changed the tumor’s color to red.

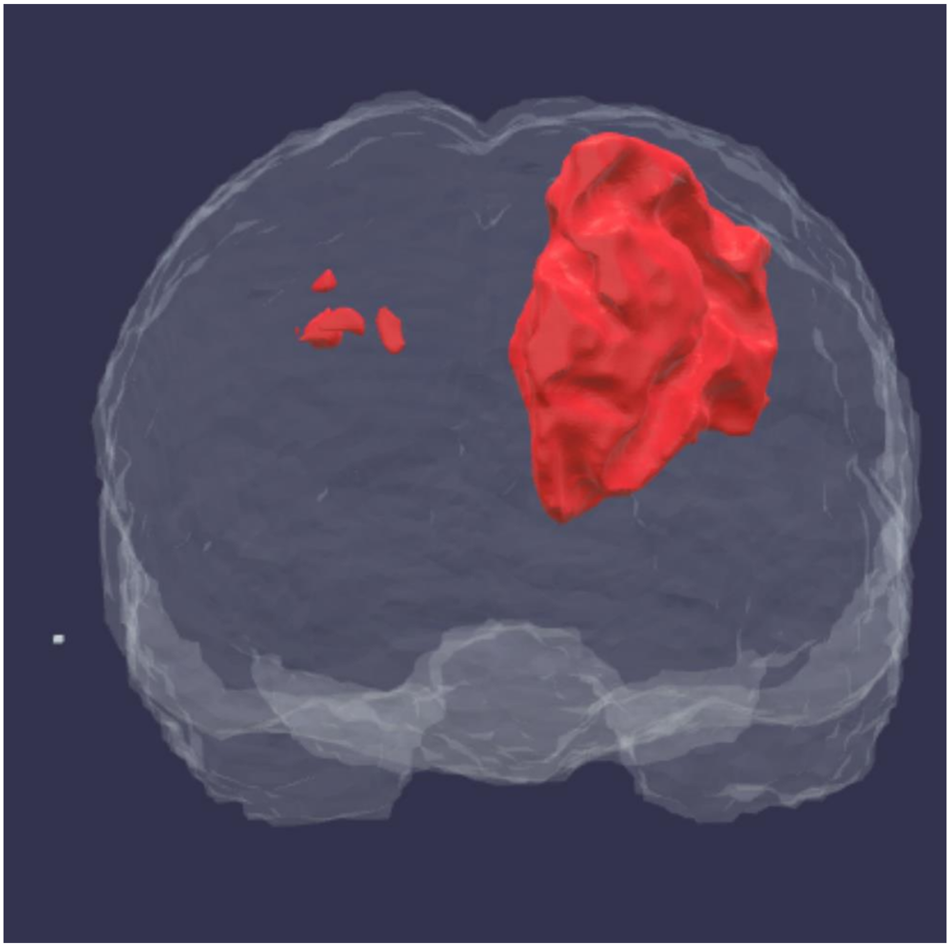


Fig 4.10 GLTF file of the model

5. GLTF to USDZ

GLTF is a transmission format for 3D assets that is apropos to web and mobile devices by removing data that is not important for efficient display of assets. iOS does not support GLTF format thus, we convert the GLTF file into USDZ file format. We conducted this conversion by using Google's usd_from_gltf tool. This conversion emulates the functionality of GLTF files in iOS.

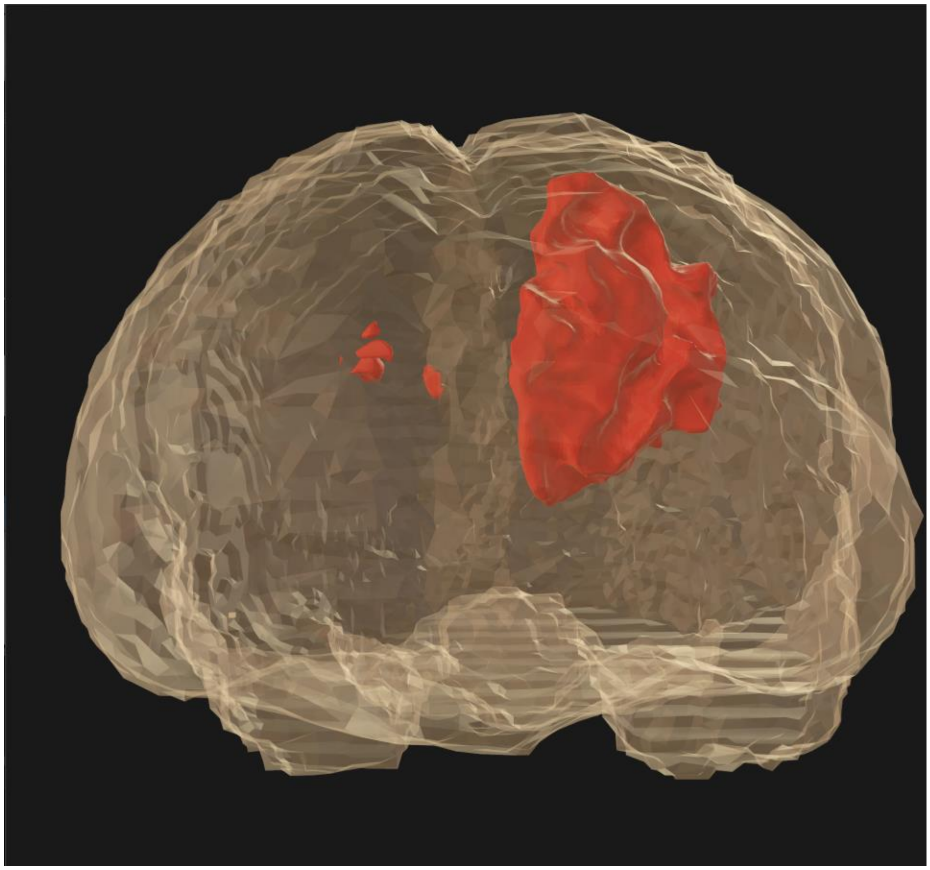


Fig 4.11 USDZ file of the model

4.5 Web Application

Once the pipeline was ready we proceeded to create a web application that serves the following purposes. To begin with, it will help the end-user understand what the product is and how to use it. Secondly, it will provide the end-user with a graphical user interface via which the user can upload the files and receive the shareable link. Lastly, it also depicts some of the models that were developed through the system for better understandability.

The frontend was designed using HTML, CSS and JavaScript and the backend was written in Flask.

The entire pipeline which is described above was written as a single function that takes as input the files of respective modalities and presents as output a modelviewer scene that has the generated 3D model.

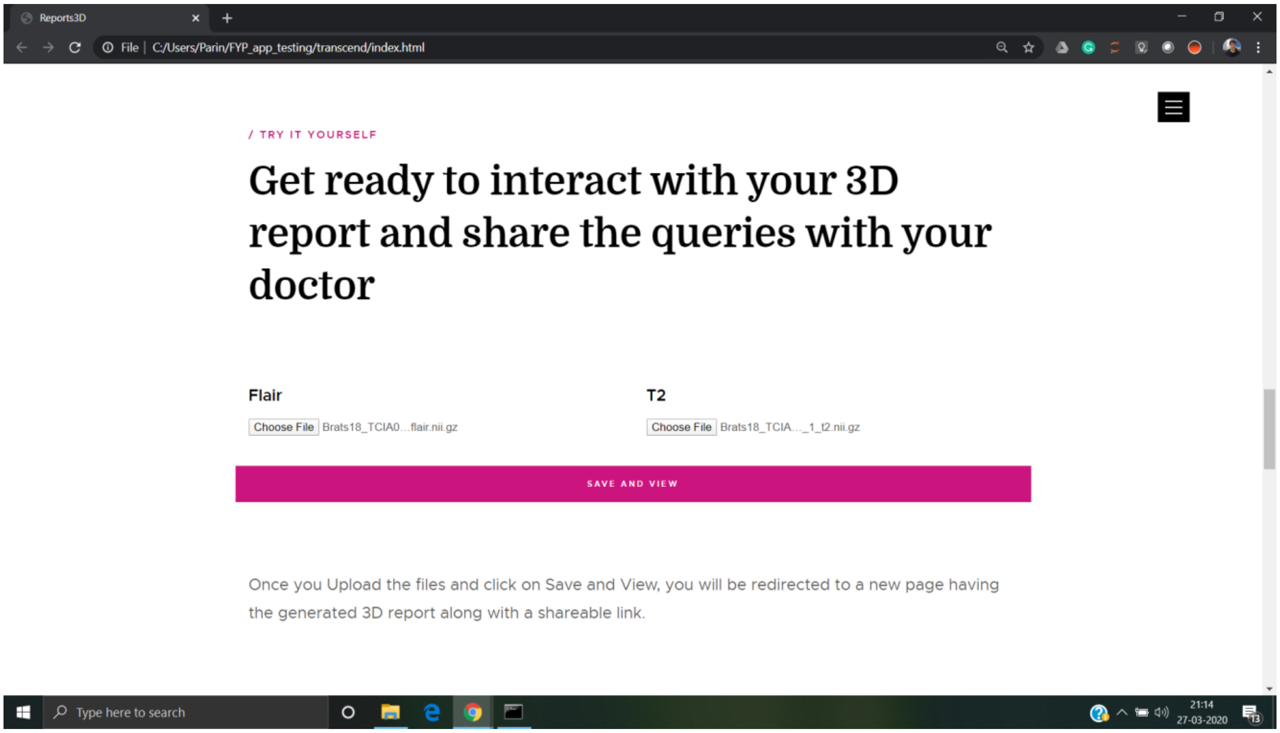


Fig 4.12 Web Application screenshot displaying the GUI to upload the files

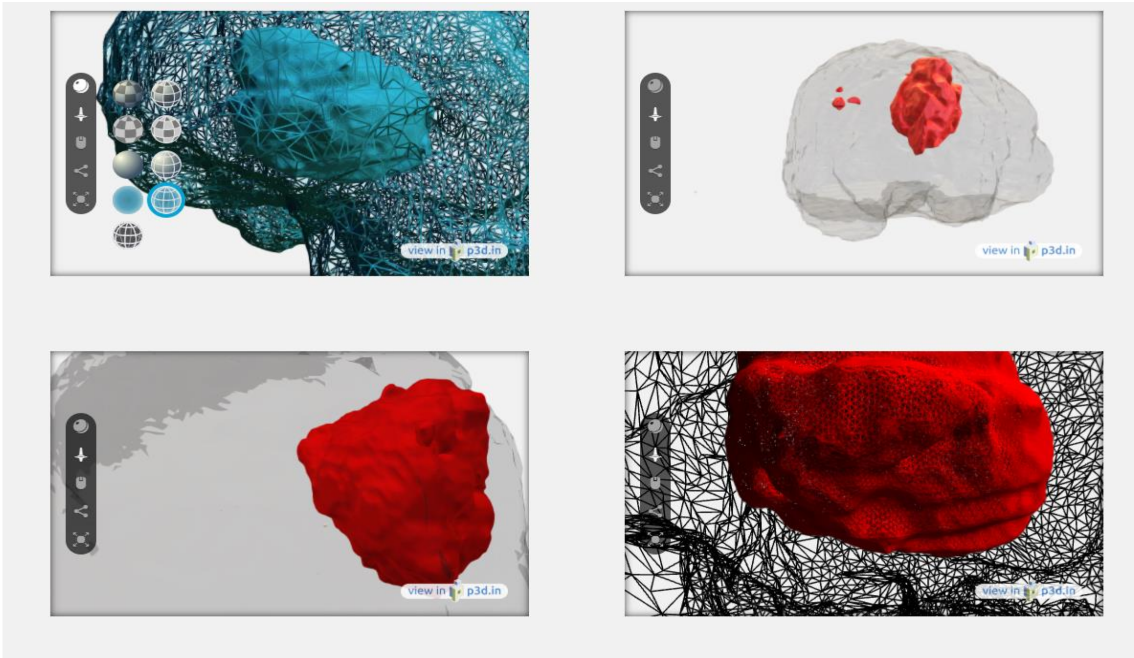
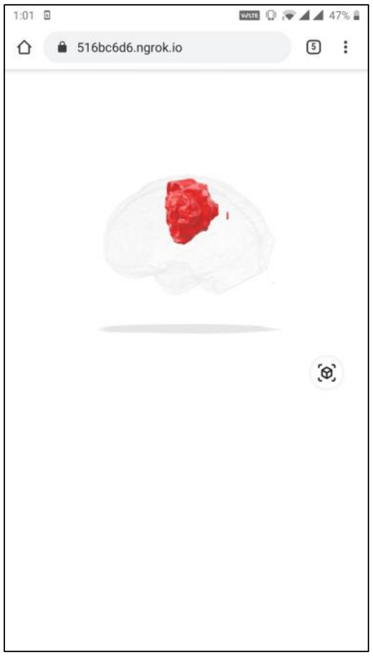


Fig 4.13 Web Application screenshot displaying the sample generated models

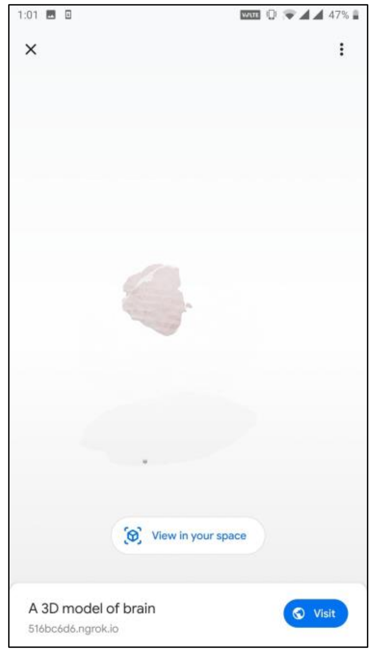
4.6 Augmented Reality scene

The final output of the pipeline is a USDZ and GLTF file that is unique to the patient. These files are fed into the modelviewer scene which allows for them to be viewed in Augmented Reality. The modelviewer scene is depicted below, it has the 3D model which can be viewed as is in the web app. This is especially great for viewing in phones that do not have AR support. In the modelviewer scene there is a button in the bottom right corner that allows the user to view the model in AR. When clicked the user is taken to a new page where the user can proceed to click the “view in your space” button. Once clicked the user is prompted to move the phone around the physical space to allow the calibration of the ground plane using the phone’s camera. Once calibrated the 3D model will appear in the physical space as shown in the figure below. The user can then move, rotate and scale the model in the physical space. The user can also take a screenshot of the scene and share it instantly.

To allow another user to view the report, the link only needs to be shared and the same process is followed on the shared user’s phone.



(a)



(b)

(c)

Fig 4.14 (a) (b) Screenshot displaying the interface that allows the user to view the scene in AR, (c) Screenshot of a model in physical space

Chapter 5

Results

A novel system to visualize medical images of brain tumours, in an immersive augmented reality setting, was successfully developed as the result of efforts put into this project. This proof of concept system can segment tumours captured in MRI scans of patient brains, create a 3D model of the patient brain and tumour, accurately identify location, size and orientation of the tumour in the brain and display it in augmented reality with the help of a web application.

The segmentation of tumour regions from brain scans is carried out with a U-Net architecture CNN. This method produces an average Dice Similarity Coefficient(DSC) of 87% for the full tumour segmentation. The DSC is a statistical validation tool based on the spatial overlap of two segmentations, to measure their reproducibility and accuracy. In this work, the DSC of the segmented tumour is calculated against the ground truth of the tumour region, provided with the dataset.

Table 5.1 DSC scores for different regions of the tumour

	Mean Dice Similarity Coefficient	Median Dice Similarity Coefficient
Full Tumour	0.87	0.90
Tumour Core	0.76	0.84
Enhancing Tumour	0.71	0.80

The novelty of this research is enhanced by the method of accurately pinpointing the location size and the orientation of the tumour within the brain of the patient. No such method exists currently, as found by the team during their study of literature on the topic.

Further, this is the first of its kind to utilize augmented reality immersion to visualize the 3D models of the brain in the tumour. Previous works have all relied on virtual reality, which requires expensive headgear and user familiarity with the platform. With the inclusion of augmented reality, this project requires only ARCore and ARKit powered smart phones, which are available dime a dozen now. The user only needs to access one customized URL and can then view the final results in their web browser on their phones.

Chapter 6

Conclusion and future work

After an extensive survey of existing research literature in the project domains, we can conclude that there are models in implementation that can detect tumours and other abnormalities from medical images and also compute the volumetric compositions but no such system exists that can do all of that and then display it as a 3D model in Augmented Reality(AR). This lack has been covered by application now. With the advent of ARCore and ARKit in almost all new generation smartphones, this tool is applicable universally without the need of any additional hardware.

To further expand on the applications of this system, support for reports of multiple target organs can be developed. Each organ has its own imaging scanning procedures, for which the system will have to be modified but the gains outweigh the work by lengths. Our vision with this project is that of a full body patient report, all available on one screen.

Further additions can be made to the system such as simulation flow of body fluids in and around the organ which can be animated as per the effects of the abnormality present. For instance, a blockage in the heart can be better visualized by animating the blood flow in an AR model and the difference in flow when the block is eliminated. This system will find its use in innumerable facets of medical applications, ranging from training to diagnosis and surgery.

Exploring the capabilities of extended reality(XR), further collaborative options can be supplemented into the system. These tools include but are not limited to:

i. Real time sharing or interaction across multiple devices for multiple doctors, doctors across borders separated from their patients and so on. All sets of doctors and/or users can observe the same model and observe the changes made by one another in real time. This gives greater power for better, quicker and multiple consultations.

ii. Cross-sectional views of organs and tumour tissues within it. This will allow doctors to understand the gravity of the abnormality to get a greater extent by facilitating clearer views and perspectives that are unable to be achieved from standard medical reports.

iii. Simulation of organ functioning with and without the presence of the abnormality in it. For instance, blood flow in the heart could be simulated with and without an artery blockage in it, and portrayed in the augmented reality scene.

iv. There lies a scope to create an entire range of XR solutions with the development of Virtual Reality(VR) solutions with the same functionalities as mentioned earlier.

v. AR surgery planning where doctors can wear a AR headset and can collaboratively work on a single 3D model to plan ways in which the surgeries can be performed and the steps to be taken to perform the same. This can also be performed using VR to provide an even immersive experience.

REFERENCES

1. A Labeeb, Yasser & Morsy, Muhammad & Abo-Elsoud, Mohy, “Preprocessing Technique for Enhancing the DICOM Kidney Images”, 2018.
2. Sheela.V., K and Dr. S. Suresh Babu. “Pre-Processing Technique for Brain Tumor Detection and Segmentation.” (2015).
3. Deepa and A. Singh, "Review of brain tumour detection from MRI images," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom),* New Delhi, 2016, pp. 3997-4000
4. G. Latif, M. Mohsin Butt, A. H. Khan, M. Omair Butt and J. F. Al-Asad, "Automatic Multimodal Brain Image Classification Using MLP and 3D Glioma tumour Reconstruction," *2017 9th IEEE-GCC Conference and Exhibition (GCCCE),* Manama, 2017, pp. 1-9.
5. Handore and D. Kokare, "Performance analysis of various methods of tumour detection," *2015 International Conference on Pervasive Computing (ICPC),* Pune, 2015, pp. 1-4. doi: 10.1109/PERVASIVE.2015.7087002
6. Jose Bernal, Kaisar Kushibar, Daniel S. Asfaw, Sergi Valverde, Arnau Oliver, Robert Martí, Xavier Lladó, “Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review”, Artificial Intelligence in Medicine, Volume 95, 2019, Pages 64- 81, ISSN 0933-3657.
7. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B.: Efficient multiscale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78 (2017).
8. Sajid, S., Hussain, S. & Sarwar, A. Arab J Sci Eng (2019) 44: 9249. https://doi.org/10.1007/s13369-019-03967-8
9. Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K. Fishman, Alan L. Yuille, “A Fixed- Point Model for Pancreas Segmentation in Abdominal CT Scans”, 1612.08230, arXiv, cs.CV, 2016.

10. Resmi, S. and Thomas, T. (2012) A Semi-automatic method for segmentation and 3D modeling of glioma tumours from brain MRI. *Journal of Biomedical Science and Engineering*, **5**, 378-383. doi: 10.4236/jbise.2012.57048.
11. Sindhushree. K. S, Mrs. Manjula. T. R, K. Ramesha, 2013, Detection And 3d Reconstruction Of Brain tumour From Brain Mri Images, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 02, Issue 08 (August 2013).
12. Qian Shan, Thomas E. Doyle, Reza Samavi, Mona Al-Rei, "Augmented Reality Based Brain tumour 3D Visualization, Procedia Computer Science", Volume 113, 2017, Pages 400-40.
13. M. A. Ghaderi, M. Heydarzadeh, M. Nourani, G. Gupta and L. Tamil, "Augmented reality for breast tumours visualization," *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 2016, pp. 4391-4394.
14. Nicolau, S., Soler, L., Mutter, D., & Marescaux, J. (2011). *Augmented reality in laparoscopic surgical oncology. Surgical Oncology, 20(3), 189–201.* doi:10.1016/j.suronc.2011.07.002
15. S. Zhang, C. Demiralp, D. F. Keefe, M. DaSilva, D. H.Laidlaw, B. D. Greenberg, P. J. Basser, C. Pierpaoli, E. A.Chiocca, and T. S. Deisboeck. An immersive virtual environment for DT-MRI volume visualization applications: a case study. In VIS '01: Proceedings of the conference on Visualization '01, pages 437–440, 2001.
16. A. Kratz, M. Hadwiger, R. Splechtna, A. Fuhrmann, andK. B¨uhler. GPU-based high-quality volume rendering for virtual environments. In Proceedings of international workshop on Augmented environments for medical imaging and computer aided surgery, 2006.
17. J. Kniss, J. P. Schulze, U. Wossner, P. Winkler, U. Lang,and C. Hansen. Medical applications of multifeld volume rendering and VR techniques. In Proceedings of the Joint Eurographics-IEEE TCVG symposium on Visualization, pages 249–254, 2004

18. Shen, Rui & Boulanger, Pierre & Noga, Michelle. (2008). MedVis: A Real-Time Immersive Visualization Environment for the Exploration of Medical Volumetric Data. 63 - 68. 10.1109/MediVis.2008.10