Vertebra Segmentation from CT Images Using Volumetric Network (V-Net)

MSc Research Project
Data Analytics

Ninad Mohite
Student ID: x18203591

School of Computing
National College of Ireland

Supervisor: Dr. Manaz Kaleel

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ninad Mohite
Student ID:	x18203591
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Manaz Kaleel
Submission Due Date:	17/08/2020
Project Title:	Vertebra Segmentation from CT Images Using Volumetric Network (V-Net)
Word Count:	4874
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th August 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Vertebra Segmentation from CT Images Using Volumetric Network (V-Net)

Ninad Mohite
x18203591

Abstract

There was a marked rise in the number of people who suffer from vertebra disorders from past many years. For diagnosis of various vertebra or spine conditions a computer-assisted surgical systems, automatic spine or vertebra segmentation resulting from CT images is important. The spine has a complicated structure which is made up of vertebra, inter-vertebral discs, spinal cord and ribs. Hence there is a need for a robust algorithm to segment and create a model of the vertebra. In this study, V-Net is developed for segmentation of vertebra from CT images and is compared with various similar methods. It is a fully convolutional neural network(FCNN) consisting of dice loss layer and convolution tasks with max-pooling layers. The V-Net gave promising results on the limited training data using which it was trained. This method achieved soft dice loss of 0.4742 and dice coefficient of 0.5152 and the specificity and sensitivity of the model are 0.6286 and 0.5857 respectively. The experiment was done with CT images from 25 patients and it demonstrated promising results obtained from the model.

1 Introduction

Accurate segmentation and anatomic diagnosis of vertebrae provide the basis for automated spine examination, such as recognition of fractures with vertebral compression or other anomalies. It is very difficult to find annotated 3D medical volumes and experts are needed to annotate such volumes which results into high cost. Automatic computer based segmentation can help reduce the cost of annotations. Surgeries are depended on computer-based technologies these days and computer technologies play a key role in defining how the surgery is performed (Kumar, 2014). Medical images like Computer Tomography (CT) scans and Magnetic Resonance Imaging (MRI) are used by such computer-based systems using which they extract all the relevant information about the patients to create a 3D structure of the patient. A surgeon can easily manipulate this model and get a view of the vertebra from all the angles and depth for accurate inspection. The surgeon can then assess the condition more accurately to create a more precise diagnosis; such a method is used in the computer-assisted spinal treatment and a system enabling treatment (Shoham et al., 2007). Precise vertebral segmentation may thus improve diagnosis occurring before the surgical operation and the efficiency of the simulation.

A most important element of computer-based surgeries is image segmentation which is used to accurately create a 3D model of a patient. Segmentation of images is essential for extracting features from an image. Based on the problem to be addressed, the segmentation process consists of dividing an item into individual components. Segmentation is stopped on the isolation of the area of interest in a given procedure. Conventional vertebral segmentation methods mainly use mathematical techniques to assess features of cone (Yen et al., 2013) or various model-based to address fitting problems (Korez et al., 2015). Some of these typical methods of vertebral segmentation evaluate mathematically the baseline, edge and vertebral anatomical properties based on a 2D spine cut and are assigned to various targets or shapes. Hence, techniques like these cannot be applied to the images with spine deformities or abnormal cases.

1.1 Research Question

RQ: *"How well can deep learning perform 3D segmentation of vertebrae using CT or MRI images ?"* To improve the diagnosis of various spine related problems.

To solve this research question following objectives are implemented.

1.2 Research Objectives and Contributions

Objective 1: Investigation and critical review of literature related to vertebra segmentation and diagnosis of spine-related problems.

Objective 2: Perform 3D image segmentation of vertebra using VNet.

Objective 2(a): Exploratory data analysis of medical image data by 3D visualization using the marching cubes algorithm.

Objective 2(b): Measuring radio-intensity of CT scan image using Hounsfield Units (HU) to separate vertebra (bones) from soft tissues.

Objective 2(c): Extracting each slice from 3D volume and prepossessing the data to generate masks for ground truth using K-Means clustering.

Objective 2(d): Prepare 3D patches of shape 128*128*128 from 2D slices and saving them as h5 file.

Objective 2(e): Implementation and evaluation of VNet for 3D image segmentation.

Contribution: The main contribution that this project offers is the development and evaluation of VNet on 3D segmentation model using 3D Convolution Neural Network that uses pre-processed patches of images to localize and segment the region of interest (Vertebra). This can be useful for doctors to improve the diagnosis of various spine-related problems by accurately creating a 3D model of the patient. Precise segmentation may thus improve the diagnosis occurring before a surgical operation.

Fully developed segmentation model resulted from objective 2 is a major contribution for supporting medical doctors for accurately creating a 3D model of the patient's vertebra.

The remainder of the technical report is laid out as follows: Section 2 represents the detailed literature review about the challenges faced during diagnosis of spine-related problems along with the work done in 2D image segmentation and the critique of existing 3D image segmentation techniques. Section 3 describes the process flow, research methodology, the technological framework and design. Section 4 describes the implementation, evaluation and results of the project. Section 5 concludes the research by summing up each of the goals, explaining them quickly and suggesting the future scope.

2 Related Work

2.1 Introductoion

In this section, an overview of literature related to vertebra segmentation and similar other segmentation tasks and various techniques and models used in the segmentation process is discussed.

This literature is divided into various sections, section 2.2 is a review of the spine and vertebra related problems are discussed along with the techniques physicians used to detect these issues related to the spine. Section 2.3 discusses various 2D and 3D image segmentation approaches used to segmenting medical images. Section 2.4 discusses the previous work related to segmentation of vertebra from spine images and the techniques, approaches used to achieve accurate segmentation of vertebra.

2.2 Review on issues related to spine and vertebra

Pain in the lower back caused by spinal disorders has been identified as a frequent factor for hospital visits in medical routine (Freburger et al., 2009). Computer-assisted spinal treatment and therapy service programs use both computed tomography (CT) and magnetic resonance imaging (MR) technology. For conditions such as spondylosis and spinal stenosis, MRI scan is a preferred diagnostic tool (Ali et al., 2014). Whereas CT images are used for diagnosis of osteoporosis in which calculating mineral density in the bone of vertebral body is done (Aslan et al., 2010). Segmentation is a go to approach along with localization for all such clinical conditions. The literature review present 2D MR based methods and 3D CT/MRI based methods for segmentation of vertebra (Egger et al., 2012). These methods can be categorized into two sub-methods model-based methods and ground truth base methods (Huang et al., 2009, Schwarzenberg et al., 2014).

2.3 Review on 2D segmentation for biomedical data

Using predictive shape models and their variations, vertebra segmentation is mainly tackled as a model fitting problem. Among all active shape model and shape constrained models are the notable ones (Castro-Mateos et al., 2015, Suzani et al., 2015, Yang et al., 2017). Atlases based approaches are also been used (Wang et al., 2016), active contours (Athertya and Saravana Kumar, 2016) and prior shape level sets are used (Lim et al., 2014). To locate bounding boxes of vertebral bodies Zukić et al. (2014) used the Adaboost based Viola-Jones object detector system. Chu et al. (2015) identified the centre of the vertebral body by pixel classification using random forest which is the region of interest using which vertebra is segmented. Suzani et al. (2015) proposed a similar method to regress a distance to the nearest centre of the vertebra using multi-layer perceptron. This distance was used to initialize the shape for segmentation of vertebra.

Machine learning aspect beyond statistical modelling is been included in the abovementioned methods, machine learning was mainly used for identification of vertebra and its segmentation. By using pixel labelling multi-class CNN Janssens et al. (2018) segmented lumbar vertebrae in 2D slices. Janssens et al. (2018) also determined the boundary box for the lumbar area by CNN followed by a CNN to classify pixel labelling to segment a vertebra which is withing a bounding box. Akbaş and Kozubek (2020) proposed a method which uses three convolution layers and max-pooling operations which they termed as simplified encoder-decoder. This work is done by modifying the U-Net model and they even introduced a novel loss function which can be used specifically for biological images.

2.4 Review on 3D segmentation of Vertebra and other biomedical images.

Wang et al. (2019) proposed a 3D UNet architecture which enables the entire network to directly train on 3D data. Similar work was done by Janssens et al. (2018) and proposed a 3D CNN fully connected network architecture to segment and locate lumbar spine. The additional reference information is not necessary for deep learning of 3D data but since the number of feature parameters is large hence there is a huge memory requirement.

The methods discussed above deals with the semantic segmentation of vertebra. Lessmann et al. (2019) describes vertebra segmentation as an image segmentation problem. The proposed method used in this work includes segmentation network along with an extension which is termed as instance memory and a batch for predicting complete data. To adequately accommodate a complete vertebra along with side vertebra patches of shape 128*128*128 were created. Segmenting each vertebra separately from top to bottom is the main idea for patch wise segmentation. Milletari et al. (2016) proposed architecture knows as V-Net which is used for segmentation of medical images along with that they introduced a custom loss function termed as Dice coefficient maximisation.

2.5 Conclusion

Based on the literature review of problems related to spine and vertebra, 2D segmentation and 3D segmentation of spine there is clear evidence that there is a need to address a various limitation in those approaches. Hence, in the work done during this research, the V-Net proposed by Milletari et al. (2016) is used for segmentation of vertebra and loss function soft dice loss is used with dice coefficient as a metric.

3 Scientific Methodology and Approach

3.1 Introduction

Knowledge Discovery in Databases (KDD) and Cross-industry standard process for data mining (CRISP-DM) are the methodologies usually used in research related to data mining. For this research KDD suits well, as model implementation or deployment in the business layer does not apply here. In recent years, data mining and the KDD procedure have been widely implemented in numerous medical branches: hypocellular myelodysplastic syndrome treatment and aplastic anaemia. The KDD cycle is broken down into separate stages, each generating different results. However, the medical domain features further confuse the way KDD phases are handled, data is analyzed, and tests are interpreted. While KDD needs advanced care in the medical domain.

3.2 Knowledge Discovery and Data Mining Methodology

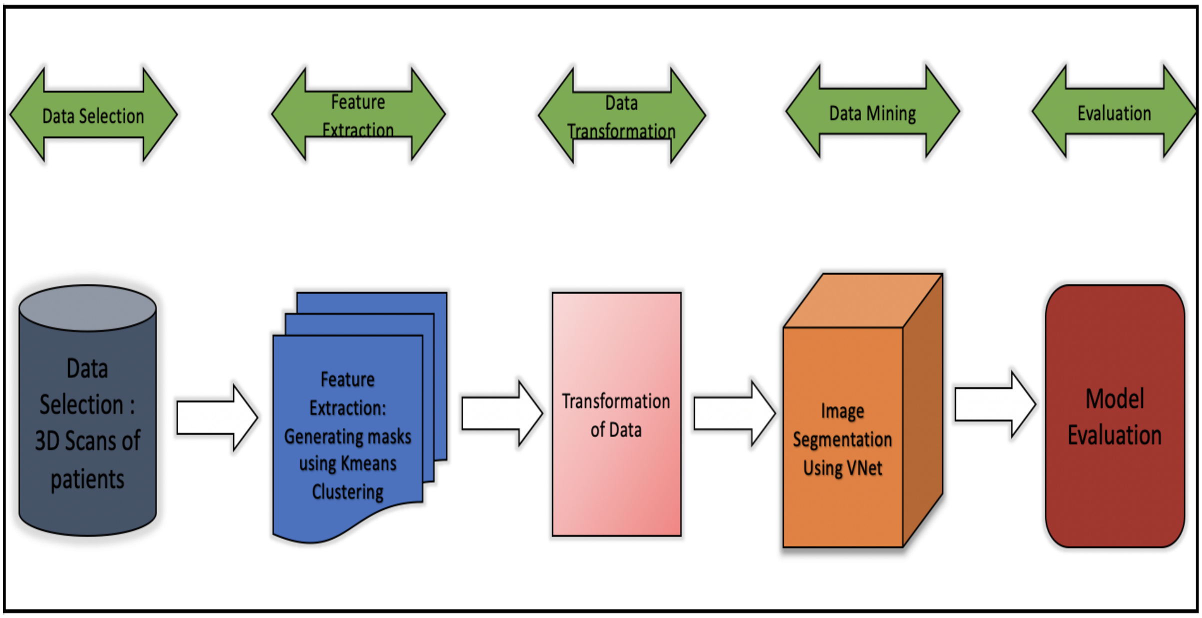

Figure 1: Methodology for 3D Segmentation of vertebrae

The Scientific Methodology used in this research is KDD (refer Figure 1) for 3D segmentation of vertebrae. This methodology consists of following stages:

1. **Data Selection :** Data from spine web [1] is used which is an open sharing forum for those involved in spinal imaging work and image processing and the data is available to use publicly. All volumes have a size ranging from 180*256*256 to 200*256*256 voxels, with voxel spacing of 1.5mm*0.9mm*0.9mm. The data set is in .mhd format. Each 3D scan is made up of a varying number of 2D slices of ranging from 180 to 200. Scans of 25 patients are collected and preprocessed. Spine web is a free and open-source repository for research in spinal data and image analysis. The spine web is open to everyone to work with its data and held challenges hence the data set used in this research is ethically sourced.

2. **Feature Extraction :** The data is made up of various slices which when combined makes a 3D model. From each slice of the data ground truth (mask) is generated using KMeans clustering.

3. **Data Transformation :** Each slice and respective ground truth is stored in .bmp format and then it is converted to patches of shape 128*128*128 to h5 file which will constitute of respective image and its mask.

4. **Data Mining :** Developing and training V-Net architecture using the 3D patches.

5. **Model Evaluation :** Model is evaluated and interpreted using the dice coefficient as the main evaluation measure.

[1] http://spineweb.digitalimaginggroup.ca/

3.3 Design Flow and Architecture

Figure 2 shows the technical architecture and design flow which is made up of two tiers. In a two-tier architecture tier, 1 represents a presentation layer which runs on the client and the tier 2 and data structure gets stored on the server also called a business logic tier.

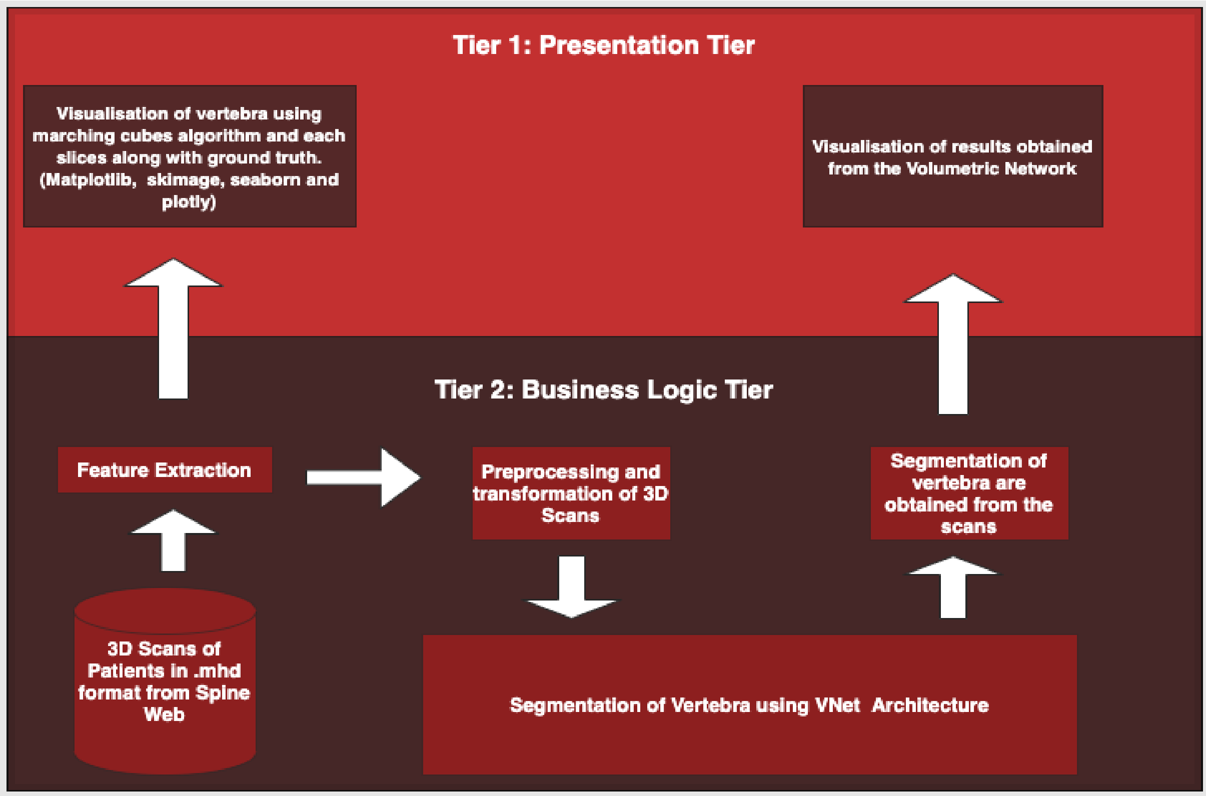

Figure 2: Design Flow

3.3.1 Presentation Layer

The Presentation Layer is the first layer of the architecture and in this tier data, understanding and visualization along with the presentation of the results obtained from V-Net are represented. For visualization, a 3D plot is made using the marching cubes algorithm and each slice of the image is visualized.

The result obtained from image segmentation using V-Net can be used by the doctors for in-depth study and early diagnosis and cure.

3.3.2 Business Logic Layer

In business logic layer is the backbone of the overall system where features are extracted from the 3D Scans of patients creating the ground truth for the segmentation using KMeans clustering. The features are then preprocessed and patches are created of size 128*128*128.

The preprocessed images and respective masks are then stored as .h5 file and passed through the V-Net model for segmentation. The results obtained from the segmentation are then visualized and presented in the presentation layer.

3.4 Technical Requirement, Feature Selection and Data Preparation

Detailed understanding of prerequisites required to implement this project is given in this section. This section is subdivided into three sections as follows:

3.4.1 Specifications of Technology Used

An architecture of image segmentation model is the extensive integration of Python's multiple packages and advanced libraries specifically used for the development of deep learning projects. Preprocessing of 3D scans is done locally using Jupyter notebook and various python libraries. The preprocessed data is then uploaded on a Google drive which is mounted on google colab a cloud service provided by Google to train deep learning model which uses Tesla K80 GPU.

3.4.2 Feature Selection

The main feature which is to be considered is the vertebra for its segmentation. To extract features .mhd files are loaded using SimpleITK. The slices are then truncated using the upper and lower limit followed by the normalization. It helps to define a complicated region of the vertebra without having to draw complex borders manually. This method extends a given region depending on the limits defined by a threshold range (minimum and maximum pixel values) to truncate the image.

3.4.3 Data Preparation

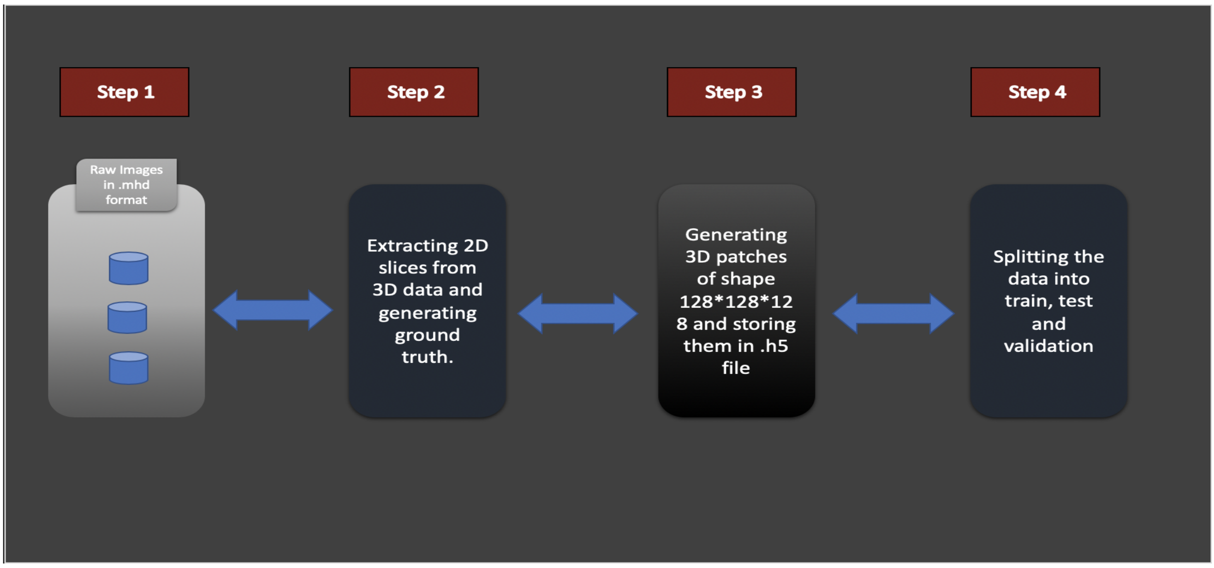

Figure 3: Data Preparation Process

Data preparation is a 4 step process where in step 1 raw data in .mhd format is acquired from the source. Each slice of 3D data is saved as a .bmp file and ground truth is generated using KMeans clustering. Then the original slices and respective ground truths are converted to 3D patches of 128*128*128 which is referred from (Lessmann et al., 2019) and are stored in .h5 file followed by splitting the data into train test and validation set for segmentation process.

3.5 Conclusion

During the implementation of the solution, the scientific methodology approach is covered in this section, the project process flow, the technical architecture and design, as well as the method for data preparation are used.

4 Design Specification of VNet architecture used for vertebrae segmentation

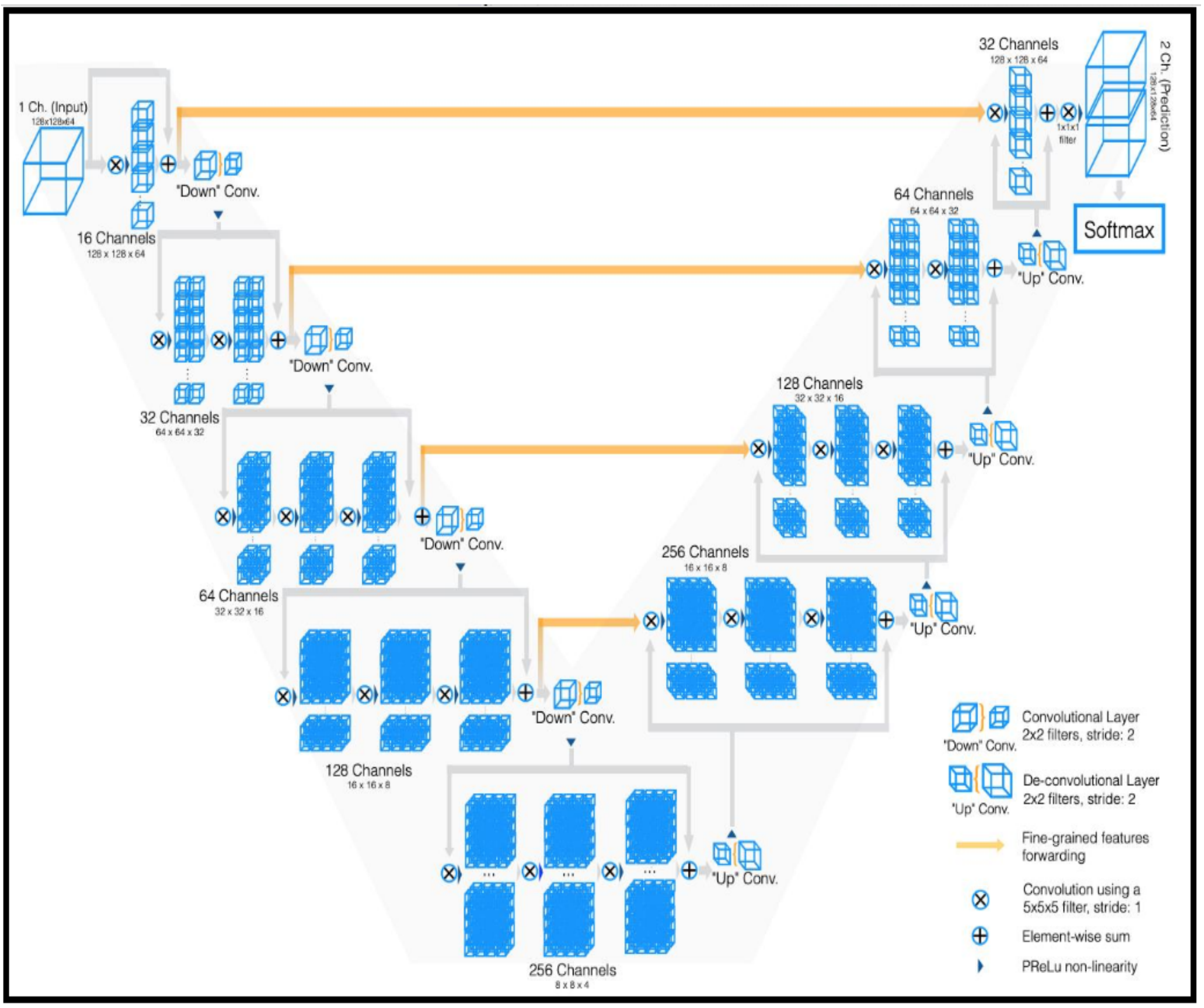

Figure 4: VNet architecture Schematic (Milletari et al., 2016)

V-Net architecture is briefly discussed in this section. Much of the diagnostic imagery used in clinical practice is composed of 3D volumes, such as prostate-representing MRI measurements (Milletari et al., 2016), although other methods can only handle 2D images.

The 3D image segmentation which is based on 3D volumes is done using a completely convolutional neural network like V-Net.

Figure 4 shows the architecture of VNet where the left part is the compression part and right part is the decompression part. The network's left side is divided into various stages which operate at different resolutions. One to three convolutional layers are there in each stage. A residual function is learnt at each stage. At every stage, the input is used in the convolutional layers and applied to the output stage's convolutional layer by passing it through non-linearities which ensures consistency over non-linear networks. To extract necessary information and to output volumetric segmentation, features are extracted along with spatial support of feature maps which are of the lower resolution. At each point, to increase the size of features deconvolutional layer is used and a residual function learned which is similar to the left part of the network.

There is some location loss during compression in the left part of the network, to deal with its horizontal connection is used. Hence, the features extracted during compression are forwarded to the right part of the network using horizontal connections. This helps in improving the quality of the final segmentation by providing location information. Padding is applied to every performed convolution. Features are computed from a larger size than that of the vertebra, hence the innermost layer in the network does computation of the entire volume.

5 Implementation, Evaluation and Results of V-Net image segmentation model

5.1 Introduction

This section provides detailed implementation of V-Net along with the evaluation of the final results of the model (Objective 2). This section also gives a detailed understanding of 3D data sets and how to handle and preprocess 3D medical scans and various terms related to it. Implementation of objective 2 is a complex process and hence it is carried out in sub-objectives.

5.2 Data Understanding and Preprocessing

This section provides insights on complete understand and preprocessing of the 3D scans.

5.2.1 Data Understanding

Understanding the data is the most important step in any machine learning problem. The data set used is in .mhd format. The library used to read the data is SimpleITK. The first thing to be considered is whether the Hounsfield Unit (HU) is properly scaled and represented in the data. It is used by radiologists to interpret CT scans which is a quantitative measure of radiodensity.

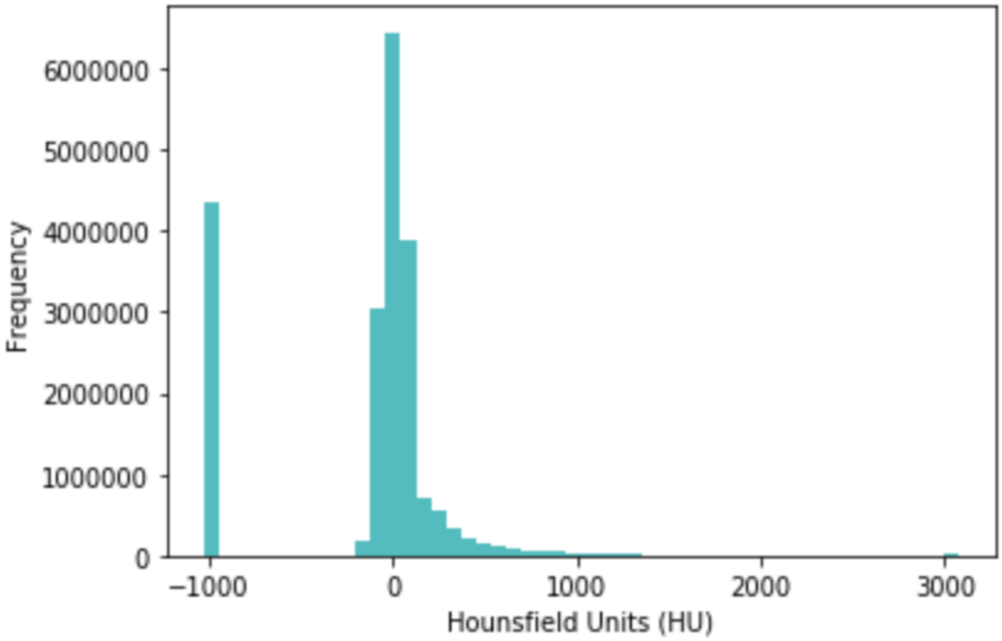

Figure 5: Hounsfield Units Histogram of a 3D Scan

The histogram in figure 5 suggests that there is lots of air (HU = -1000) and water (HU = 0) in the scan. There's a large amount of soft tissue mostly muscles and fats and there is a decent amount of bones seen between 700 to 3000 which is our region of interest as vertebrae are made up of multiple bones. This observation suggests that significant preprocessing will be needed to separate vertebra from other components.

5.2.2 Data Preprocessing, Generation of Mask (Ground Truth) and 3D Visualization

1. **3D Visualization :** To make data suitable as an input for CNN the data needs to be preprocessed. After having information about HU of a scan there resampling of data is done. Although we can extract each slice the thickness of the slice is unknown. Resampling and using the metadata from .mhd file slice thickness was figured out to plot the data in 3D. Figure 6 shows the interactive 3D plot of a slice rendering muscles and other tissues.

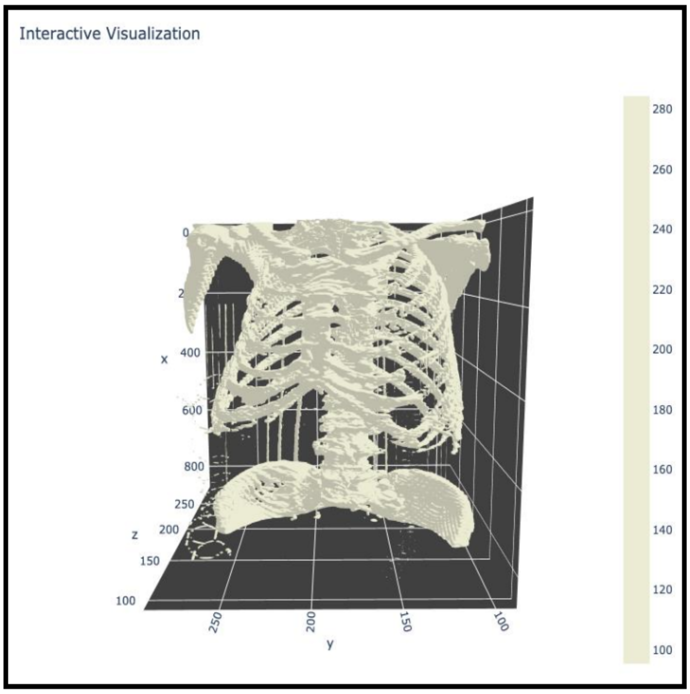

Figure 6: Interactive 3D plot of Scan

It is beneficial to have isotropic data, as it gives us a sense of the Z dimension. Which implies that there is enough information to map the CT scan picture in 3D space. By focusing on rendering muscles and soft tissues. Using matplotlib high quality static 3D plot is created and an interactive rendering is created using Plotly. Marching cubes algorithm is used to create a 3D mesh.

2. **Generation of Ground Truth**

   Each slice is preprocessed by auto-detecting the boundaries surrounding the ROI to generate masks. Figure 7 shows the generation of mask for each slice.

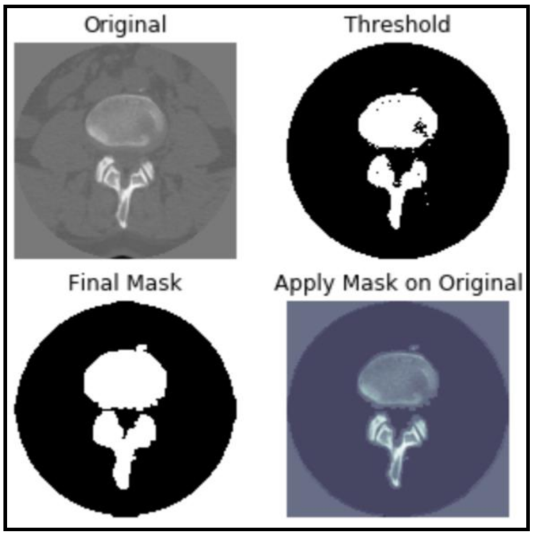

Figure 7: Example of Extraction of mask from a slice

Following steps are taken to generate masks from a slice:

- Pixel value is standardized by subtracting the mean value and then dividing by standard deviation.
- By creating KMeans clusters a proper threshold value is identified.
- To remove tiny features like noise erosion and dilation is used.
- Each distinct label is identified as an image label.
- Creating a bounding box for each label to understand which one is vertebra and which label represents other features.
- Generate mask for vertebra and apply it on the original image for verification.

3. **Creation of 3D patches from 2D slices**

   The idea of patch wise segmentation is referred from Lessmann et al. (2019) where it is suggested the patch size cover entire vertebra along with its centre is 128*128*128. Original slices and respected ground truths are stored as a .bmp image file and then converted into patches of 128*128*128 since the patch wise segmentation process is used. The patches of images and masks are stored in .h5 format which is a hierarchical data format containing arrays of data.

   To prepare 3D patches number of patients, height and width of each slice and the desired shape of the patch are taken as an input. The images are then read using OpenCV and are stored in image array and mask array variables. Image depth is calculated along with other parameters like step width, step height, stride width, stride height, and width and height of patch block. The generated patches of images and masks are then saved as .npy files which are then stored in HDF file format for easy handling. Entire preprocessing of data along with visualization was done locally and the preprocessed data was then uploaded on google drive for training V-Net model on google colab.

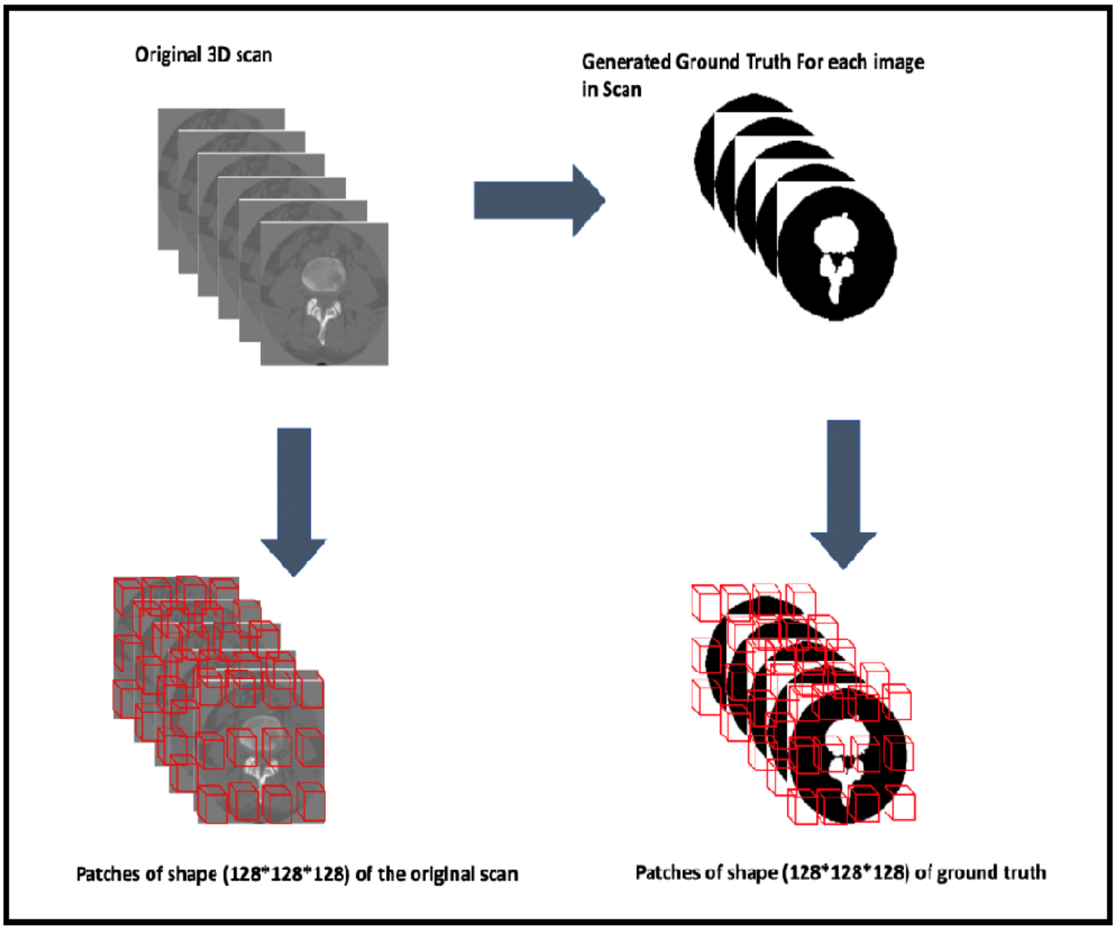

Figure 8: Generating 3D patches

Figure 8 illustrates the extraction of 3D patches from the original scan and ground truth where red cubes show 3D patches of shape 128*128*128. Each patch along with the patch of its respective ground truth is fed to the network. Initially, ground truth is generated which segments vertebra using K-means clustering which is explained in detail in section 5.2.2 subsection 2. The function generates the sub-images and masks with desired patch blocks size (in this case 128*128*128), taking the image, stride as function parameters. It then calculates the width, height and depth of the image and sets the stride width and height accordingly. The image patches and its respective mask patches are fed as in input to the network.

5.2.3 Implementation of V-Net

V-Net architecture as shown in figure 4 is been implemented using Keras and TensorFlow. The design includes, as shown in figure 4 a series of down-convolutions linked by max-pooling operations, accompanied by a series of up-convolutions linked by upsampling and concatenation operations. In the upsampling section of the network, each of the down-convolutions is often explicitly linked to the concatenation operations. V-Net's "depth" is proportional to how many down-convolutions are needed. The depth in the implemented architecture is 4 as 4 down-convolutions extend down the left side to the very bottom of the V.

Starting by building the downward path in the network here (the V-Net's left side). If we go along this path, the input (height, width, length) gets smaller, and the number of channels increases. For constructing depth 0 which is referred to the depth of the first down-convolution in the V-net. For increasing depth and for each layer inside that depth the number of filters is defined. The formula for estimating the number of filters to be used is:

$$filters_i = 32 \times (2^i)$$

Where i is the current depth.
So at depth $i = 0$:

$$filters_0 = 32 \times (2^0) = 32$$

6 Evaluation and Results

This section provides an evaluation of the results from the V-Net model used for segmentation of vertebra. Along with the statistical techniques, model outputs and matrix used for the evaluation of the segmentation model. Evaluation of the segmentation model is based on dice coefficient, sensitivity and specificity.

6.1 Dice coefficient and Soft Dice Loss

6.1.1 Dice coefficient

Choosing the loss function is one of the most important elements of any deep learning model. The cross-entropy loss function is a popular choice which many are familiar with. However, due to excessive class imbalance, this loss function is not optimal for segmentation tasks (typically there aren't that many positive regions).

The Dice similarity coefficient is a much more common loss for segmentation tasks which is a measure of how well two shapes overlap. The dice index ranges from 0 to 1 where 0 is a complete mismatch and 1 is the complete match. Dice coefficient for two sets is generally defined as :

$$\text{Dice Coefficient}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}.$$

In the above equation we can consider A and B as sets of voxels where A is predicted region of vertebra and B is the ground truth. The model maps each voxel to 0 or 1 where zero means background voxel and 1 means that the voxel is a part of segmented region. The formula for dice coefficient is:

$$\text{Dice Coefficient}(f, x, y) = \frac{2 \times \sum_{i,j} f(x)_{ij} \times y_{ij} + \epsilon}{\sum_{i,j} f(x)_{ij} + \sum_{i,j} y_{ij} + \epsilon}$$

Where x is the input image, $f(x)$ is the predicted value (model output), y is the label (ground truth) and ϵ is a number which is added to avoid division by 0. The matrix used for the model is the Dice coefficient.

6.1.2 Soft Dice Loss

The Dice Coefficient is not the best for training even though it makes sense. Because discrete values like 0 and 1 is been taken by dice coefficient. We will not be able to backpropagate through the outputs where the model gives outputs as probablity where every pixel is vertebra or not. Hence soft dice loss is used as a loss function which is analogue of dice loss and takes real input values. The formula for soft dice loss is:

$$\mathcal{L}_{Dice}(p, q) = 1 - \frac{2 \times \sum_{i,j} p_{ij} q_{ij} + \epsilon}{\left(\sum_{i,j} p_{ij}^2\right) + \left(\sum_{i,j} q_{ij}^2\right) + \epsilon}$$

Where p is the prediction from the model, q is the ground truth, in practice, each q_i will either be 0 or 1 and ϵ is a number which is added to avoid division by 0.

Below figure shows the result of soft dice loss and dice coefficient on validation data set.

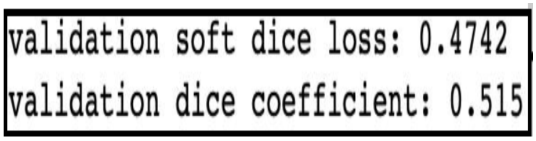

Figure 9: Dice Loss and Soft Dice Loss (Validation)

Figure 9 shows the output from the validation data set where the model shows promising results on the limited amount of data. The model was trained only on 1000 slices because of the lack of resources and the immense training time using more than 14GB of Nvidia Tesla K80 GPU. The dice coefficient value is 0.515 which means there is 51.5% average overlap between the samples and the average soft dice loss is 0.4742.

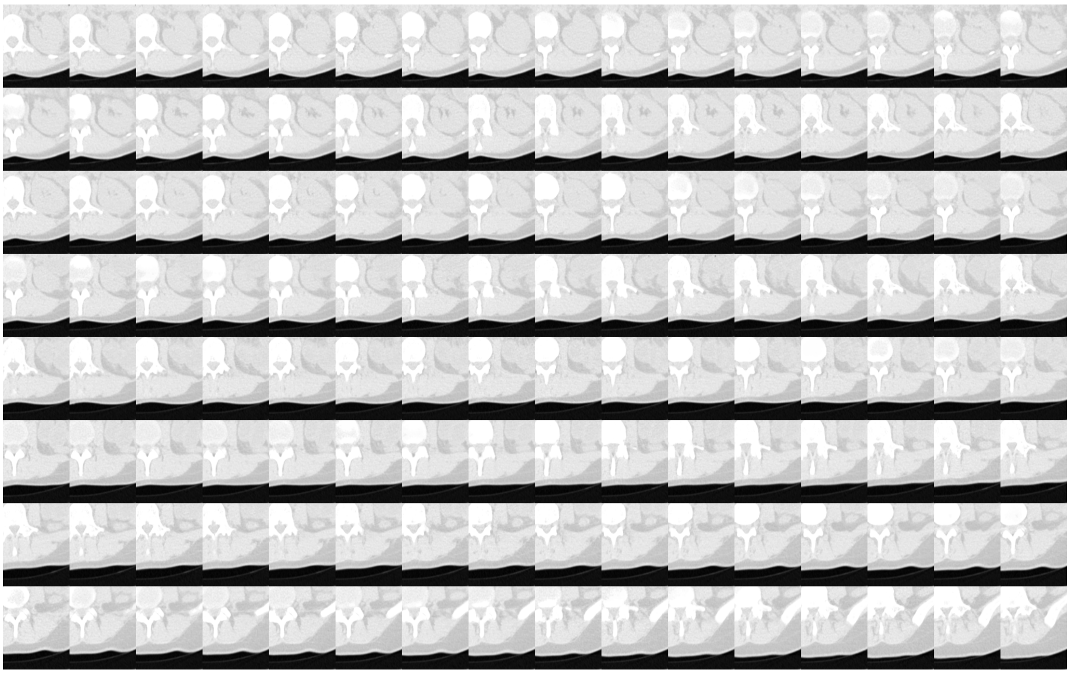

Figure 10: Source

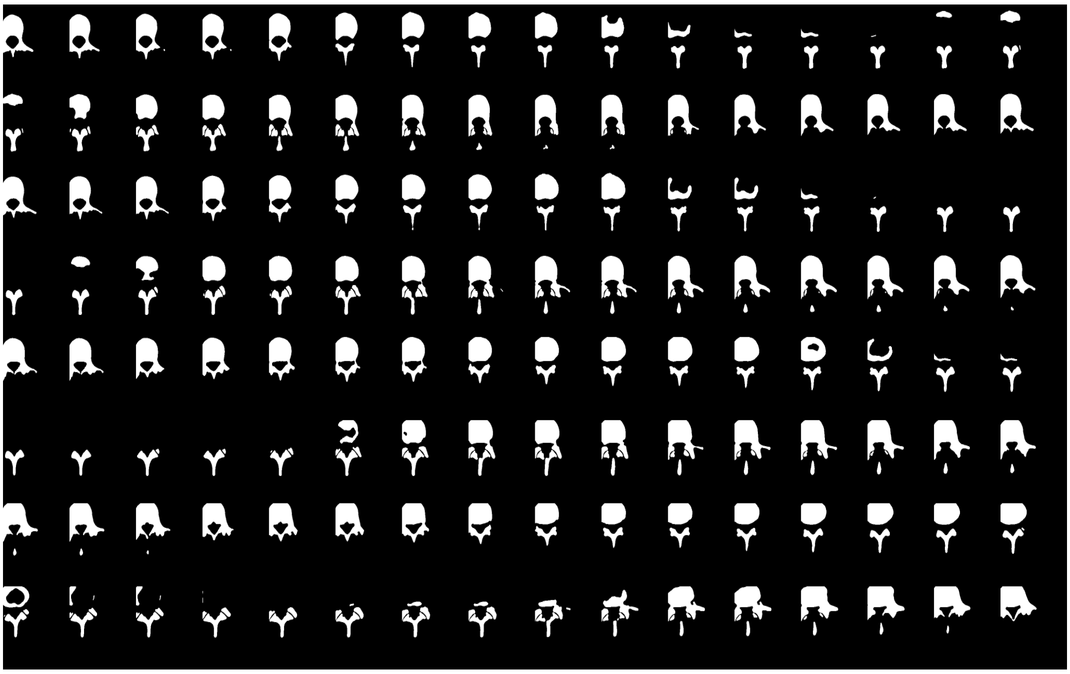

Figure 11: Ground Truth

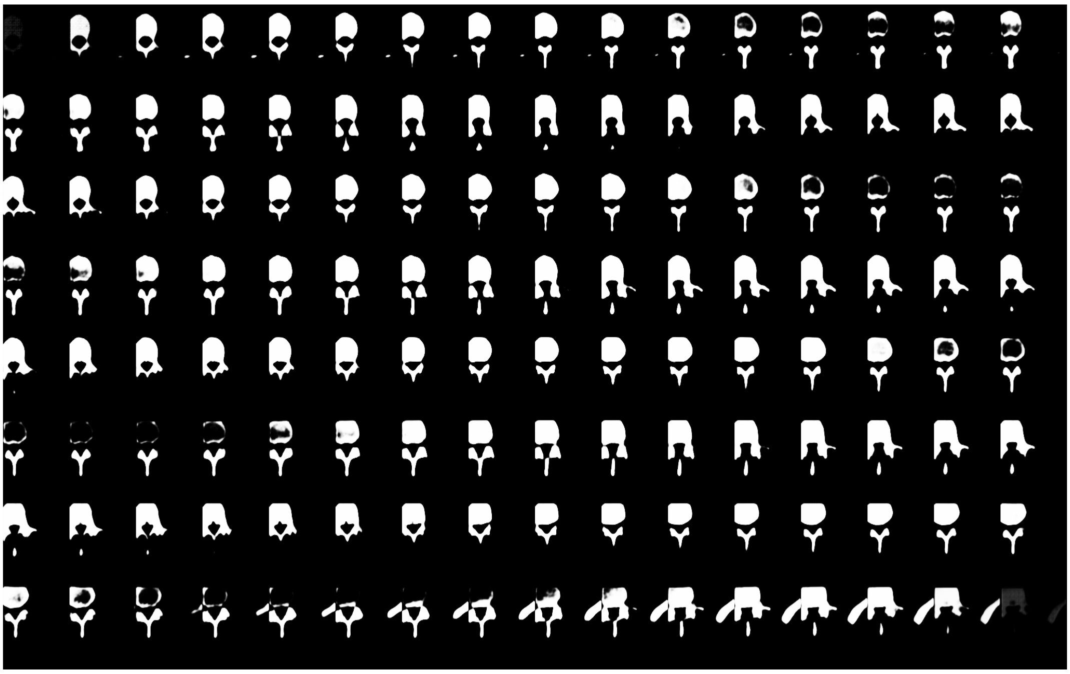

Figure 12: Predicted

Figure 10 and 11 shows the input data in the model which is a patch of shape 128*128*128. Where figure 10 is the original patch and figure 11 is its ground truth. The final predicted output of the V-Net model can be seen in figure 12.

6.2 Sensitivity and Specificity

The model is not perfect though it covers some relevant areas. Hence per pixel sensitivity and specificity is used to quantify its performance. Which can be recalled in terms of true positives, true negatives, false positives, and false negatives.

$$\text{sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{specificity} = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}}$$

```
sensitivity, specificity = compute_class_sens_spec(patch_pred[0], y, 2)

print(f"Sensitivity: {sensitivity:.4f}")
print(f"Specificity: {specificity:.4f}")
```

```
Sensitivity: 0.5857
Specificity: 0.6286
```

Figure 13: Dice Loss and Soft Dice Loss (Validation)

Table 1: Evaluation scores.

Validation Soft Dice Loss	0.5152
Validation Dice Coefficient	0.4742
Sensitivity	0.5857
Specificity	0.6286

Figure 13 shows the sensitivity and specificity of the model. The sensitivity of 0.5857 means the model correctly segmented 58.57 % of values correctly. The specificity of the model is 0.6286 which means the model correctly identified 62.86 % the negative values from the image patch.

Table 1 summarises the results obtained from the model.

6.3 Discussion

The project is primarily aimed to help medical practitioners to detect and segment vertebra which can be helpful in early diagnosis of various problems related to spine and vertebra. The V-Net model first introduced by (Milletari et al., 2016) is used to segment vertebra by changing the loss function. The loss metric used is the Dice coefficient and the loss function used is soft dice loss function. The total patches generated from the data are 18591 out of which only 4200 patches are used. The model could be trained only on 2940 patches and was tested on 630 patches and validation was done on 630 patches, because of the extensive GPU usage and the limited training time of only 12 hours offered by google collaboratory. The GPU required to train the model is at least 14GB (Tesla K80) with the batch size of 1 and the model could only be trained on a batch size of 1 because of huge memory usage.

Even though the data used for training the segmentation model was less the results are promising with soft dice loss of 0.4742 and dice coefficient of 0.5152. The sensitivity of the model is 0.5857 which means it correctly segmented positive data points with 58.57 % accuracy and the specificity of the model is 0.6286 which indicates the model correctly identified 62.86 % negative.

Table 2: Comparison With Previous Work.

Author	**Model**	**Dice Loss**
(Milletari et al., 2016)	V-Net	0.869
(Chuang et al., 2019)	3D U-Net	0.8616
(Wang et al., 2019)	3D U-Net	0.8818
Our	V-Net	0.5152

The comparison of various models can be seen in the table 2. The V-Net model used in this research showed promising results with a limited amount of resources and training data.

7 Conclusion and Future Work

The primary aim of the project is to help medical practitioners detect and segment vertebra automatically which can be helpful in early diagnosis of various problems related to spine and vertebra. The implementation, evaluation and the results discussed in section 6 has enabled a solution to the research question. The main goal was to accurately segment vertebra from CT images and the model performs well on the limited data set used. The model achieved a dice score of 0.5152 and specificity and sensitivity of the model is 0.6286 and 0.5857 respectively. The limitation of this work is that it requires extensive processing power to deal with a huge data set. It took almost 7 days to generate 3D patches on the local machine, also the GPU requirement is more hence the model could not be trained on the entire dataset.

For future work, the research can be done on various CT and MRI scans of different regions for segmentation. The model could be enhanced to lower the computation time and power by training the model using Big Data frameworks like Apache Hadoop and Spark. The preprocessing of the data can be done on Spark or Hadoop frameworks to speed up the process of generation of 3D patches.

8 Acknowledgement

This thesis could only be completed because of my supervisor, Dr. Manaz Kaleel. I would especially like to thank him for his extensive advice and encouragement. At all points of my research, he kept me motivated. Without his guidance and support, I could not have finished my thesis project. I would also like to thank all the other faculty professors, for their guidance and support throughout this masters course.

References

Akbaş, C. E. and Kozubek, M. (2020). Condensed u-net (cu-net): An improved u-net architecture for cell segmentation powered by 4x4 max-pooling layers, *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 446–450.

Ali, A. M., Aslan, M. S. and Farag, A. A. (2014). Vertebral body segmentation with prior shape constraints for accurate bmd measurements, *Computerized Medical Imaging and Graphics* **38**(7): 586 – 595. Special Issue on Computational Methods and Clinical Applications for Spine Imaging.
URL: *http://www.sciencedirect.com/science/article/pii/S0895611114000603*

Aslan, M. S., Ali, A., Farag, A. A., Rara, H., Arnold, B. and Xiang, P. (2010). 3d vertebral body segmentation using shape based graph cuts, *2010 20th International Conference on Pattern Recognition*, pp. 3951–3954.

Athertya, J. S. and Saravana Kumar, G. (2016). Automatic segmentation of vertebral contours from ct images using fuzzy corners, *Computers in Biology and Medicine* **72**: 75 – 89.
URL: *http://www.sciencedirect.com/science/article/pii/S0010482516300634*

Castro-Mateos, I., Pozo, J. M., Pereañez, M., Lekadir, K., Lazary, A. and Frangi, A. F. (2015). Statistical interspace models (sims): Application to robust 3d spine segmentation, *IEEE Transactions on Medical Imaging* **34**(8): 1663–1675.

Chu, C., Belav?, D. L., Armbrecht, G., Bansmann, M., Felsenberg, D. and Zheng, G. (2015). Fully Automatic Localization and Segmentation of 3D Vertebral Bodies from CT/MR Images via a Learning-Based Method, *PLoS ONE* **10**(11): e0143327.

Chuang, C., Lin, C., Tsai, Y., Lian, Z., Xie, H., Hsu, C. and Huang, C. (2019). Efficient triple output network for vertebral segmentation and identification, *IEEE Access* **7**: 117978–117985.

Egger, J., Kapur, T., Dukatz, T., Kolodziej, M., Zuki?, D., Freisleben, B. and Nimsky, C. (2012). Square-cut: a segmentation algorithm on the basis of a rectangle shape, *PLoS ONE* **7**(2): e31064.

Freburger, J. K., Holmes, G. M., Agans, R. P., Jackman, A. M., Darter, J. D., Wallace, A. S., Castel, L. D., Kalsbeek, W. D. and Carey, T. S. (2009). The rising prevalence of chronic low back pain, *Arch. Intern. Med.* **169**(3): 251–258.

Huang, S., Chu, Y., Lai, S. and Novak, C. L. (2009). Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri, *IEEE Transactions on Medical Imaging* **28**(10): 1595–1605.

Janssens, R., Zeng, G. and Zheng, G. (2018). Fully automatic segmentation of lumbar vertebrae from ct images using cascaded 3d fully convolutional networks, *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 893–897.

Korez, R., Ibragimov, B., Likar, B., Pernuš, F. and Vrtovec, T. (2015). A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation, *IEEE Transactions on Medical Imaging* **34**(8): 1649–1662.

Kumar (2014). Analysis of medical image processing and its applications in healthcare industry, *International Journal of Computer Technology and Applications* **5**(3): 851–860.

Lessmann, N., van Ginneken, B., de Jong, P. A. and Išgum, I. (2019). Iterative fully convolutional neural networks for automatic vertebra segmentation and identification, *Medical Image Analysis* **53**: 142 – 155.
URL: *http://www.sciencedirect.com/science/article/pii/S1361841518305905*

Lim, P. H., Bagci, U. and Bai, L. (2014). A robust segmentation framework for spine trauma diagnosis, *in* J. Yao, T. Klinder and S. Li (eds), *Computational Methods and Clinical Applications for Spine Imaging*, Springer International Publishing, Cham, pp. 25–33.

Milletari, F., Navab, N. and Ahmadi, S. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation, *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571.

Schwarzenberg, R., Freisleben, B., Nimsky, C. and Egger, J. (2014). Cube-cut: vertebral body segmentation in MRI-data through cubic-shaped divergences, *PLoS ONE* **9**(4): e93389.

Shoham, M., Lieberman, I. H., Benzel, E. C., Togawa, D., Zehavi, E., Zilberstein, B., Roffman, M., Bruskin, A., Fridlander, A., Joskowicz, L., Brink-Danan, S. and Knoller, N. (2007). Robotic assisted spinal surgery–from concept to clinical practice, *Comput. Aided Surg.* **12**(2): 105–115.

Suzani, A., Rasoulian, A., Seitel, A., Fels, S., Rohling, R. N. and Abolmaesumi, P. (2015). Deep learning for automatic localization, identification, and segmentation of vertebral bodies in volumetric MR images, *in* R. J. W. III and Z. R. Yaniv (eds), *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, Vol. 9415, International Society for Optics and Photonics, SPIE, pp. 269 – 275.
URL: *https://doi.org/10.1117/12.2081542*

Wang, L., Xie, C. and Zeng, N. (2019). Rp-net: A 3d convolutional neural network for brain segmentation from magnetic resonance imaging, *IEEE Access* **7**: 39670–39679.

Wang, Y., Yao, J., Roth, H. R., Burns, J. E. and Summers, R. M. (2016). Multi-atlas segmentation with joint label fusion of osteoporotic vertebral compression fractures on ct, *in* T. Vrtovec, J. Yao, B. Glocker, T. Klinder, A. Frangi, G. Zheng and S. Li (eds), *Computational Methods and Clinical Applications for Spine Imaging*, Springer International Publishing, Cham, pp. 74–84.

Yang, D., Xiong, T., Xu, D., Huang, Q., Liu, D., Zhou, S. K., Xu, Z., Park, J., Chen, M., Tran, T. D., Chin, S. P., Metaxas, D. and Comaniciu, D. (2017). Automatic vertebra labeling in large-scale 3d ct using deep image-to-image network with message passing and sparsity regularization, *in* M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap and D. Shen (eds), *Information Processing in Medical Imaging*, Springer International Publishing, Cham, pp. 633–644.

Yen, C., Su, H. and Lai, S. (2013). Reconstruction of 3d vertebrae and spinal cord models from ct and stir-mri images, *2013 2nd IAPR Asian Conference on Pattern Recognition*, pp. 150–154.

Zukić, D., Vlasák, A., Egger, J., Hořínek, D., Nimsky, C. and Kolb, A. (2014). Robust detection and segmentation for diagnosis of vertebral diseases using routine mr images, *Computer Graphics Forum* **33**(6): 190–204.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12343*