Medical Image Segmentation Application

Deepika . B
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai,
India

Vishnupriya . J
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai,
India

***Abstract*—There is a significant increase in the number ofcases related to different spine pathologies in the recent years. For an accurate computer aided diagnostic system, an unambiguous and reliable segmentation of the human vertebral column is crucial. It aids doctors in detecting different spine pathologies with improved visualizations. Magnetic Resonance Imaging (MRI) is the best method available to visualize spine. MR images of spine are clearer and more detailed than images obtained from other imaging methods. Correct segmentation of vertebrae from MR images is a challenging task because of the high variations in shape topology, low contrast of images and different fields of view. We propose an efficient computer aided diagnostic system for fully automatic segmentation of vertebrae. Our approachuses Mask R-CNN, a deep neural network algorithm that was initially developed to detect objects and perform instance segmentation of objects in natural images. In our paper, we demonstrate the application of Mask R-CNN in segmentation of vertebrae in Magnetic Resonance images of spine. The aim of this system is only to perform segmentation and a medical professional is required to arrive at an accurate diagnosis from thesegmentation results.**

Keywords*—*Medical Image Segmentation, Mask R-CNN, Vertebrae Segmentation, MRI Spine

I. INTRODUCTION

An important part of the human body is the spinal column. It provides the main support for the body allowing it to perform various actions like standing upright, bending etcetera in addition to protecting the spinal cord. The spine extends from the neck region to the pelvic region. It consists of 33individual bones called the vertebrae stacked on top of one another. The spinal column is divided into five regions namely the cervical, thoracic, lumbar, sacrum and coccyx regions. The cervical region is composed of seven cervical vertebrae numbered from C1 to C7. Its function is to support the weight of the head and movement of skull. The thoracic region comprises twelve thoracic vertebrae numbered from T1 to T12. Its main function is to support the rib cage and provide protection to heart and lungs. The lumbar region consists of five lumbar vertebrae from L1 to L5. Its main function is to bear the weight of the body. The vertebrae in the sacrum and coccyx region are fused. There are five sacral vertebrae and four vertebrae in the coccyx region. The function of sacral vertebrae is to connect the spine to the hip bone and the vertebrae in coccyx region provides attachment for ligaments and muscles of pelvic floor. The vertebrae in the spinal column are separated and cushioned by inter vertebral discs that prevents them from rubbing together. There are twenty-three inter vertebral discs in the spinal column. The discs also provide padding between vertebrae during weight bearing. The labelled diagram ofspine is given in Fig. 1.

There are different types of imaging techniques that is used by doctors to diagnose spinal disorders in patients. XRay is the first type of imaging test that a patient may be advised to take. X-rays are collected to specifically visualize the bony portions of the spinal column by applying radiation to penetrate into bones and tissues of different densities. It helps diagnose aging changes in the spine and intervertebral discs, instability in spine, birth defects of the vertebral column and trauma-caused fractures. It also helps detect osteoporosis and tumors.

Computed Tomography Scan(CT Scan) is obtained by using radiation source that generates 2-D and reconstructed 3D images of the spine. Multiple cross-sectional images of the spine are generated. They help to detect degenerative changes in the spine, alignment of spine, fractures, congenital spine defects, herniated discs, and narrowing areas in the spinal canal.

Magnetic Resonance Imaging (MRI) is the most efficient

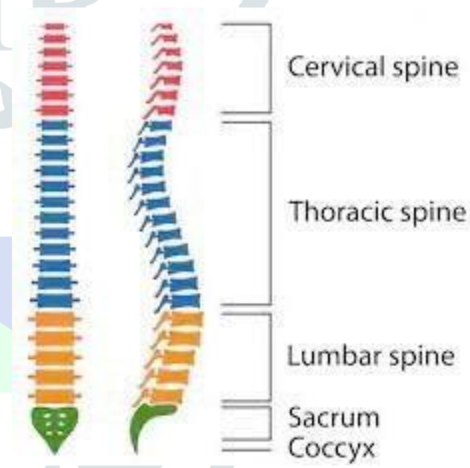

Fig. 1. Labelled Diagram of Spine

imaging technique that is used currently in diagnosing spinal disorders. A magnetic field is used to align the nuclei of hydrogen atoms present in water molecules in the body and then radio frequency pulses are provided to alter the alignment. The magnetic field changes and the energy generated by the atoms are detected and computerized by the scanner to provide detailed images of the body. It is excellent for visualizing spinal disorders. Aging changes like the herniated and bulging intervertebral discs, narrowing areas of the spinal cord, congenital defects, infections in the spine, tumors and soft tissue injuries can be diagnosed using MRI scans.

Computer Aided Diagnostic Systems present a fundamental tool for medical practitioners to visualize the radiological scans better. It enhances the general process of clinical diagnosis and conjointly prevents from medical error. There is a need for definite and reliable computer aided diagnostic system since it has a proportional impact on the decision made by the doctors. For the diagnosis of spine pathology, there is a requirement for an efficient and robust spine segmentation algorithm that can be used to build an efficient system.

There are numerous algorithms proposed for vertebral segmentation but no system to aid the doctors have been built. Doctors and radiologists perform manual segmentation that makes the overall process of clinical diagnosis less efficient as well as time-consuming. This also could lead to false diagnosis. To overcome this, we propose to develop an application that can automatically segment vertebrae in MR images of spine to help doctors in diagnosis.

II. LITERATURE REVIEW

Image Segmentation is the activity of splitting a digital image into a number of segments. The objective of segmentation is to make it easy and/or alter the way the image is represented to make it more simple and meaningful for analysis. Detection of objects of interest and their boundaries in a 2D or 3D image automatically or semi-automatically is called medical image segmentation. The high variability of images is a major complication in the process of segmenting medical images. The result obtained from segmentation can be used to obtain further diagnostic insights.

A framework integrating U-Net and parametric level sets was proposed to perform segmentation of bones and disks accurately [1]. Extraction of features from the image that is given as input is done by U-Net and the output obtained in this step is fed to the level set for further processing. Their system predicts an output of uniform sized segmented patch from a 128*128 sized input.

A method to perform automatic segmentation of lumbar vertebrae using cascaded 3D Fully Convolutional Networks (FCN) comprising localization and segmentation FCNs was presented [2]. The bounding box of the vertebrae in the lumbar region is found by training the Localization FCN. A Segmentation FCN which after being trained, executes a multiclass segmentation for each pixel that could map lumbar region data to its respective labels.

A method that uses four pixel- wise networks for segmentation was proposed [3]. MR images are segmented at different scales by each network. Output obtained from one network in the series is given as input to the subsequent network. Each subsequent network gives an increasingly better segmented output.

A learning based approach was proposed by Tom van Sonsbeek et al. to localize and identify vertebrae from 3D MR spine images that were weakly labelled [4]. The approachcomprised two cascading networks that localize and identify vertebrae simultaneously. A slice-based level detection is performed by the first network on 3D sagittal volumes using adaptive loss function. Center slices of each vertebra are predicted and provided as output. Sub volumes are obtained from the sagittal slice to be given as input to the consequent network. The second network classifies and localizes the vertebrae.

A FCN that performs iterative instance segmentation for segmenting and labelling the vertebrae was proposed [5]. The images are analyzed in patches and a single vertebra is segmented in the patch.

III. PROPOSED WORK

In this paper, we propose a novel application which is a computer aided diagnostic system for medical image segmentation that would be used to automatically segment vertebrae in MR images of spine. This application uses the Mask RCNN [6] network to perform segmentation of vertebrae. The final output would be a segmented 2D image of the spine. The Mask R-CNN model is deployed as a web application using Streamlit.The activity diagram of the proposed application is presented in Fig. 2.

IV. TECHNICAL APPROACH

A. Data Collection

The data that is used in this paper was obtained from a competition on Kaggle called " Segmentation in 3D MR Spine". The dataset contains 215 3-D MR images of spine.The experimental data were T2-weighted MR image sequences of 215 patients. It was classified into Train and Test images. There were 200 training images and 15 images for testing. The images were in Neuroimaging Informatics Technology Initiative (NIfTI) file format given by the extension nii.gz .

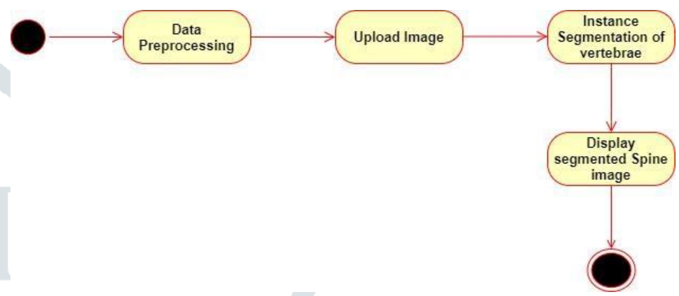

Fig. 2. Activity Diagram

B. Data Preparation

The first step in data preparation was to extract 2-D slices from the 3-D MR images in the dataset. The images in the training directory were split into train and validation sets : 75% (150 images) and 25% (50 images) respectively. The images in both the train and validation sets were resized to 1024 * 1024 pixels and a bit depth of 32 pixels. An example of the resized training image is given in Fig. 3.

C. Generation of ground truth

The ground truth was generated using the VIA(VGG Image Annotator Tool) [7]. It is an HTML file which could be downloaded and used in a web browser without the requirement of installation of any software. The images were annotated by creating bounding polygons around the vertebrae. The annotations are saved as a JSON file in which each mask is given by a set of polygon points. Annotations were exported as two JSON files ; each for training and validation data. An example of usage of this tool can be seen in Fig. 4.

D. Algorithm Used

Mask R-CNN is a deep neural network that is used to perform instance segmentation. Instance Segmentation is identifying each object instance for every known object within an image. Mask R-CNN extends Faster R-CNN. There are two parts in Mask R-CNN model. The first part is a Region Proposal Network (RPN) that outputs class object bounding boxes. The second part includes a mask classifier that creates a mask for every class. The architecture of Mask R-CNN is presented inFig. 5.

Fig. 3. A resized training image

Fig. 4. Annotation generation using VGG Image Annotator tool

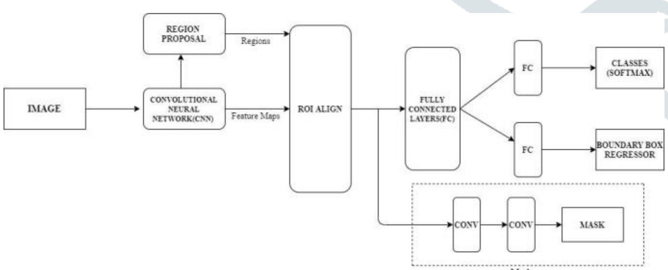

Fig. 5. Mask R-CNN architecture

The first step in Mask R-CNN model is the generation of feature maps from the input image using Convolution Neural Networks(CNN). Multiple Region Of Interest(ROI) is then generated by the Region Proposal Network (RPN) using CNN and a binary classifier. This is done using 9 anchor boxes on the image. Anchor boxes are predefined boundary boxes of certain dimensions that are used to capture the aspect ratio and scale of object classes that are to be detected. The classifier generates IoU(Intersection over Union) which computes intersection of ground truth bounding box and predicted bounding box for a given object. When the predicted and the groundtruth bounding boxes overlap perfectly, the IoU is 1. Anchor boxes that belong to the background are removed and the remaining boxes are filtered based on their IoU to perform the final object detection. Non-Max Suppression (NMS) removes bounding boxes with IoU less than 0.5 and the boxes with the greatest confidence scores are selected. An example of Region Of Interest generated by the Region Proposal Network before refinement is presented in Fig. 6.

Non-Max Suppression (NMS) removes bounding boxeswith IoU less than 0.5 and the boxes with the greatest confidence scores are selected. Multiple bounding boxes are generated by the ROI Align Network and are warped to a fixed dimension. An example of output generated after application of Non-Max Suppression is presented in Fig. 7.

The features that are warped are then fed to a Mask classifier that generates a binary mask for each ROI and also are fed to Fully Connected Layer(FCL) to perform classification and predict boundary boxes for each object. An example of prediction generated by the Mask R-CNN network is presented in Fig. 8.

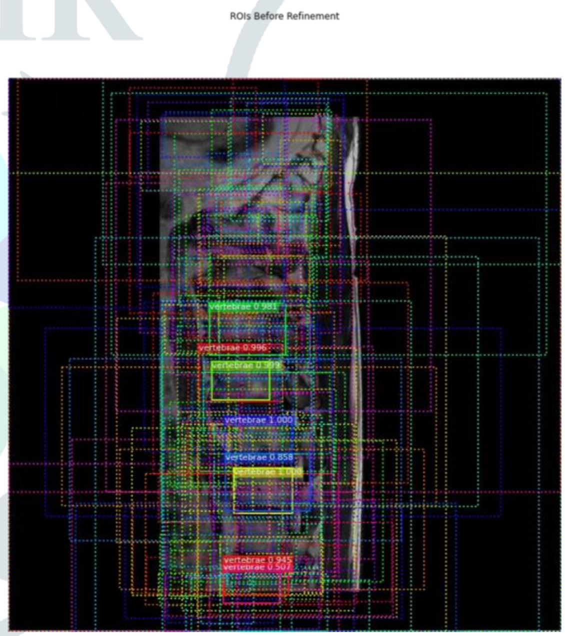

Fig. 6. Region Of Interest generated by Region Proposal Network before refinement

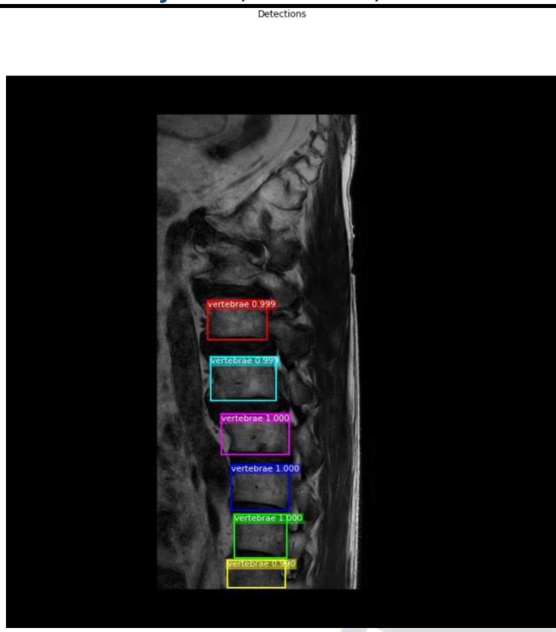

Fig. 7. Output generated after application of Non-Max Suppression

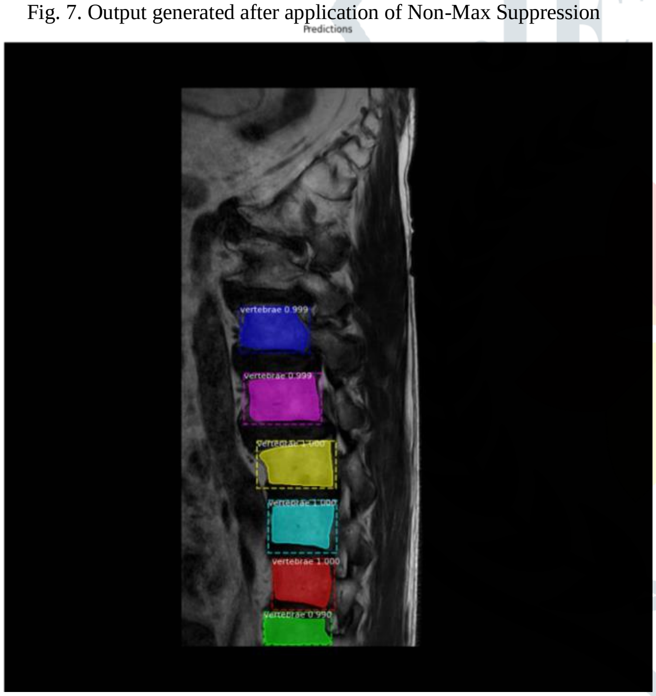

Fig. 8. Prediction by Mask R-CNN network

E. Training

The version of Mask R-CNN that is used in this paper is based on Matterport Inc's [8] implementation of the algorithm which was released under MIT License. The implementation is based on Keras and Tensorflow. We used the ResNet-101 architecture for feature extraction from training images in our implementation. The model was trained on 150 training images and 50 validation images. The model was trained using the weights that were got from training on the MSCOCO dataset [9] rather than from scratch. The training constituted of a total of 40 epochs with 100 steps per epoch, learning rate of 0.001 and a minimum detection confidence of 0.9. The network heads were trained for 20 epochs and all the layers of the network were trained for an additional 20 epochs. The training was carried out on a 12GB NVIDIA Tesla K80 GPU on Google Colaboratory.

V. USER INTERFACE

The trained Mask R-CNN was deployed as a web application. The model was deployed as a web application using Streamlit. Streamlit is an open source python app framework that lets developers in data science and machine learning areas to deploy their projects as interactive and easy to use web applications. The web page allows the user to upload a MRI scan of the spine by specifying the location of the file in the text box present in the user window. The image uploaded by the user is given as input to the Mask R-CNN model that runs in the server. The uploaded file which is in NIfTI format is converted to JPEG file format and resized as per the requirements of the network and then further processing is done to perform instance segmentation of vertebrae. The segmented image is rendered and displayed to the user in the same web page. The application's user interface to upload an image is presented in Fig. 9. The application window that displays the segmented image of the spine is shown in Fig. 10.

Fig. 9. User Interface to upload image

Fig. 10. Application Window displaying the segmented image.

VI. FUTURE WORK

We have implemented an application of the Mask R-CNN algorithm in our system to successfully perform segmentation of the vertebrae in the spinal column. Our work without a doubt would be a great help to the doctors and radiologists in quicker and efficient diagnosis unlike the conventional manual segmentation methods that are performed by doctors. Our system saves time by performing segmentation quickly and with more accuracy than manual segmentation.

In the future, this work can be extended to support segmentation of various other parts of the human body like the brain, heart, kidney etcetera. The network could be trained to produce 3D output which could also take real time interactions from the user to further improve performance. Finally, we can also extend the system to provide a sample diagnosis based on the segmentation performed.

VII. CONCLUSION

With the rapidly growing number of cases in spine pathology, this novel system to perform automatic segmentation of vertebrae will be a boon to the doctors and radiologists in diagnosing and treating patients quicker and more efficiently. The expectation from the system is that it helps doctors in better visualization of the scan images thereby reducing the amount of time consumed by conventional manual segmentation.

ACKNOWLEDGMENT

We are thankful to our guide, Mrs. K. Karthikayani for the continued guidance and support which helped us towards the completion of the project. We wish to extend our special thanks to Mr.Keerthi Ram of HTIC for his guidance throughout the project.

REFERENCES

[1] F. Rehman, S. I. A. Shah, N. Riaz, and S. O. Gilani, "A Robust Scheme of Vertebrae Segmentation for Medical Diagnosis," *IEEE Access*, vol. 7, pp. 120 387–120 398, 2019. [Online]. Available: 10.1109/ACCESS.2019.2936492

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

[3] W. Whitehead, S. Moran, B. Gaonkar, L. Macyszyn, and S. Iyer, "A deep learning approach to spine segmentation using a feed-forward chain of pixel-wise convolutional networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 868–871.

[4] T. van Sonsbeek, P. Danaei, D. Behnami, M. H. Jafari, P. Asgharzadeh, R. Rohling, and P.Abolmaesumi, "End-To-End Vertebra Localization and Level Detection in Weakly Labelled 3D Spinal Mr using CascadedNeural Networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1178–1182.

[5] N. Lessmann, B. [van Ginneken], P.A. [de Jong], and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical Image Analysis*, vol. 53, pp. 142–155, 2019. [Online]. Available:https://doi.org/10.1016/j.media.2019.02.005

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[7] A. Dutta and A. Zisserman, "The VIA Annotation Software for Images, Audio and Video," in *Proceedings of the 27th ACM International Conference on Multimedia*. New York, NY, USA: ACM, 2019. [Online]. Available:10.1145/3343031.3350535

[8] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," 2017.

[9] T.-Y.Lin, M. Maire, S. Belongie, J. Hays, P.Perona, D. Ramanan, P.Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision - ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.