Segmentation and Registration of OARs in HaN Cancer

Remya R[1],Julia Marylin Raphael[2], Manju Mariya[3], Maria Sibi Philip[4] ,Pinky leo[5]

[1]*Assistant Professor, Dept. of Computer Science and Engineering, Federal Institute of Science and Technology, Ernakulam, Kerala, India*
[2]*BTech student, Dept. of computer science and engineering, Federal Institute of Science and Technology , Ernakulam, Kerala, India*
[3]*BTech student, Dept. of computer science and engineering, Federal Institute of Science and Technology , Ernakulam, Kerala, India*
[4]*BTech student, Dept. of computer science and engineering, Federal Institute of Science and Technology , Ernakulam, Kerala, India*
[5]*BTech student, Dept. of computer science and engineering, Federal Institute of Science and Technology , Ernakulam, Kerala, India*

---***---

Abstract - *Head and neck cancer refers to a group of cancers that affect the head and the neck region. It is one of the most prevalent cancers and a significant factor in cancer-related deaths globally. Image guided radiation therapy (IGRT) is commonly used for the treatment of head and neck cancer due to its high accuracy and effectiveness in delivering precise doses of radiation to the affected area. Precise delineation of Organ-at-Risk (OAR) is crucial for the successful implementation of Image-Guided Radiation Therapy (IGRT). In routine clinical practice, OARs are usually manually segmented by oncolo- gists on CT images OARs are tediously, subjectively, and laboriously sub-divided by oncologists by hand and can cause inconsistencies due to inter and intra-observer variations. Automated OAR segmentation methods can significantly reduce the time cost of this process and improve its accuracy, which is desirable in the context of image guided radiation therapy (IGRT). Accurate OAR registration can also assist clinicians in making a diagnosis by comparing patients to healthy individuals using the registration results. .Our project is a three-dimensional (3D) architecture for concurrent OAR registration and segmentation to help oncologists with OAR contouring. The registration network's purpose is to localise an OAR by aligning a chosen OAR template to a fresh picture volume. The findings are passed on to two 3D U-Nets, LocNet and SegNet for segmentation. The segmentation outcomes can be further improved by incorporating shape details during the demarcation process.*

Key Words: **3D U-Nets, LocNet, SegNet, CNN, ReLU, Organ-at-risk, CT Images**

1. INTRODUCTION

Head and Neck (HaN) cancer refers to a group of cancers that affect the head and the neck region.It belongs to the most prevalent cancers and a crucial source in cancer-related death worldwide [1]. Each year, around 600,000 new instances of head and neck cancer are identified, and 40 to 50 percent of these cases result in mortality [2]. If found early, HaN cancer is frequently completely curable with single-modality therapy. Treatment for more advanced head and neck malignancies often consists of a variety of surgical, radiation, and chemotherapy procedures. Currently, one of the most efficient treatments for cancer is image-guided radiation therapy (IGRT) due to its high accuracy and effectiveness in delivering precise doses of radiation to the affected area.The most important stage in IGRT planning is to consider volumes and organ-at-risks (OARs) [4]. Oncologists manually segment OARs in everyday clinical practise, which takes time even for seasoned oncologists. A single patient may experience this process for about three hours. Additionally, due to inter- and intra-observer differences, manual segmentation generates inconsistencies. Particularly for soft tissue in a small volume, the amount of OARs is determined by the physicians experience and the particular imaging procedures. Particularly for head and neck cancer radiotherapy, which involves numerous significant OARs, including the optical nerves, pituitary, and so forth, the accuracy of OARs delineation directly effects the dose distribution in OARs. For head and neck cancer radiation, a more reliable and accurate automatic OARs segmentation is preferred clinically. It has been demonstrated that a precise registration technique can help clinicians increase the accuracy of localizing OARs, and in that situation, specialists can establish a diagnosis based on a comparison between patients and healthy individuals and data obtained from registration. Additionally, using an automated segmentation method can save the time required for OAR delineation greatly. The development of a automated, precise and effective method for OAR registration and segmentation in CT images is therefore of great importance in the field of radiation therapy. Creating such a system is difficult, largely because of the four factors listed below. (1) OARs have complicated anatomical structures with significant variance. (2) The organs such as the optical nerve and chiasm are difficult to distinguish in CT images because they have poor contrast . (3) The physical imbalance in the proportions of the organs makes it difficult to educate the model. (4) High resolution planning CT may extend the trained model's inference time. In this paper, we presented a cascade structure involving a registration network and a segmentation network for joint registration and segmentation of multiple OARs from CT image volumes in order to address these issues. The registration network sought to match the template's OARs to the fresh CT image volume [4]. The results are then passed to two U-Nets which are LocNet and Segnet [6].LocNet is used to locate the bounding box of each OAR. The bounding box is then passed through SegNet where the segmentation of OAR takes place. The segmented results are corrected using shape correction for accurate results.

2. LITERATURE SURVEY

Over the past years, researchers have been investigating the possibility of applying image segmentation and registration.

To increase the identification and delineation of key organs with more precision and efficiency, D. Guo et al [4] uses a deeply-supervised 3D P-HNN as its backbone and employs differentiable neural architecture search to find suitable convolutional blocks for each OAR category. But the computing cost and time required for the neural architecture search may limit its applicability in some circumstances.

Using two 3D U-Net model [5] a research suggests a two-stage segmentation framework for organ at risk (OAR) segmentation . The region of interest is identified in the first stage, LocNet and the OARs are segmented in the second stage, SegNet. The two stage method lowers the possibility of false positives or false negatives and enables more accurate identification of the area of interest. Additionally, the application of a 3D U-Net model makes it possible to take into account 3D spatial data, which is crucial for precise segmentation in volumetric medical images.

Another study that aimed to accurately segment organs at risk (OARs) in HaN CT images utilizes segmental linear functions (SLFs) to enhance contrast, a 3D-SepNet network[6] with hard-voxel weighting to prioritize difficult voxels, and an ensemble of models with varying loss functions and SLFs to estimate uncertainty. It tackles challenges such as low contrast, anisotropic resolution, and imbalanced organ sizes and outperforms alternative methods. The proposed framework has the potential to support informed clinical decision making.

CDED-net [8] is a deep learning method for image segmentation that can handle complex backgrounds and objects with fuzzy boundaries. It includes a boundary-emphasization data augmentation step, which involves creating a large number of object-boundary patterns from each image in order to improve the segmentation performance. But for collecting long range dependencies in the input data, CDED-Net might not be adequate.

Joint Registration and Segmentation using 3D CNN network [9] is a multi-task learning (MTL) in order to improve the performance of both tasks by leveraging shared knowledge between them and it is fast and has good generalization to independent test sets. However, the quality of the input CT scans, such as the existence of noise or distortions, may have an impact on how well the approach performs.

The U-Net [10] architecture extracts features from the input images, and generates a segmentation or registration output. The U-Net architecture is particularly well-suited for image registration tasks because it allows for the integration of spatial information from different scales. The model may not perform well on new images if the training data are slanted or constrained, and the performance of the method may depend on the selection of the hyperparameters.

Generalized Hough Transform [11] involves building a deformable model and then used to segmentation. The method can be used with big image datasets and has the potential to increase the consistency and accuracy of segmentation findings. The quality and representativeness of the training data used to build the model determines how accurate the segmentation results will be. The model might not function properly on images if the training set of data is limited.

Hansen and Heinrich [12] proposes GraphRegNet, which is a registration method based on generating displacement vectors at a sparse set of keypoints and uses graph regularization. It can withstand noise and artifacts in the input data. However, a substantial amount of labelled training data is needed. Such data collection and annotation can be time-consuming and expensive.

The FCN-UTA [13] architecture enables the feature maps to transferred across layers. By capturing both local and global context information, this feature map transfer can increase the precision of segmentation. The segmentation output can be captured with the help of the FCN-UTA architecture. Understanding the reasoning behind the network's choice of segmentation strategies may be difficult with such methods due to their interpretability issues. This can be a problem in some situations where transparency and understandability are crucial.

Unlike previous works, we presented a cascade structure involving a registration network and a segmentation network for registration and segmentation of multiple OARs from CT image volumes in order to address these issues. The registration network sought to match the template's OARs to the fresh CT image volume . The results are then passed to two U-Nets which are LocNet and Segnet. The segmented results are corrected using shape correction for accurate results.

3.COMPARISON

The two 3-D U-Net framework achieves promising results in accurately segmenting organs at risk, which are crucial for radiotherapy planning in head and neck cancer treatment. It makes the segmentation process more efficient and automated. This framework can be adapted to other medical image segmentation tasks, making the proposed method generalizable to other medical imaging applications. Its also ensures consistent and reliable results. LocNet is computationally efficient, requiring fewer parameters than other object detection architectures, making it faster and more cost-effective to train and run. SegNet uses convolutional layers to capture both low-level and high-level features, making it effective in accurately segmenting objects of different sizes and shapes in complex images. SegNet's architecture is computationally efficient, which allows for faster training times compared to other CNN architectures for segmentation.

Table -1: Comparison table

Comparison table			
Paper	**Methodology**	**Advantages**	**Disadvantages**
[4]	3D P-HNN	Manage the highly complex segmentation space of OARs.	Effective organ stratification has not been studied for complex segments..
[5]	Gradient boosting random forest support vector machine	Gradient boosting with feature agglomeration outperforms other methods	Cannot classify ECE and non-ECE based on detected lymph node regions
[6]	Two 3D U-Net LocNet and Segnet	Decomposition allows the two tasks to be completed more accurately and quickly	For some of the output,there were several isolated regions that did not belong to the target structure
[7]	3D-SepNet	Multiple OARs have a good visibility at the same time.	The average DSC is 0.14 percent lower than that of the second best method
[8]	CDED-net	Can capture contextual information of multiple input	Takes considerable amount of time
[9]	3D-CNN network	Can dynamically learn to share the feature maps	They had trouble with larger deformations
[10]	U-NET	Robust to capture the intensity invariance in diverse regions.	Shows instabilities in regions of low invariance in regions of high intensities.
[11]	3D generalized GHT	Increases the degree of freedom of the allowed deformations	Suffers from high memory demand.
[12]	GraphReg - Net	Low number of model parameters and fast training data	Distribute information on a high resolution irregular grid
[13]	FCN-UTA	Accurate description of vitiligo severity	Add overhead to the system

Unlike previous works, for the registration and segmentation of several OARs from CT image volumes, we provided a cascade structure that involved a registration network and a segmentation network for an accurate outcome.

Table -2: Performance Analysis

Method	**Performance measure**
3D P-HNN	56.3 DSC
Gradient Boosting	66.3 MSE
Segnet	86.5 DSC
3D-SepNet	47.6 DSC
CDED-net	64.1 MSE
3D-CNN Network	65.8 DSC
U-NET	76.15 DSC
3D generalized GHT	55.58 MSE
GraphRegNet	72.0 MSE
FCN-UTA	71.2

Table -3:Details of Datasets

Dataset	**No of CT scans**
The Head-Neck Cetuximab dataset	80
The HNSCC-DB	106
The Multi-Parametric Radiomics dataset	60
The University of Michigan Head-Neck dataset	50
The MDA Head and Neck Cancer dataset	63
The MICCAI 2015 dataset	33

4.METHODOLOGY

Our proposed system contains both registration and segmentation of CT scan images to segment required OARs.The CT image of one of the patient was used as a template .The procedure for choosing a template involves aligning the template image with each image in the data set. Dice similarity coefficient (DSC) and mean square error (MSE) are two measures that are used to assess the alignment's quality. A difference index (DI), which calculates the difference between the registered template and other pictures, is created by combining these measurements. The ideal template is chosen as the one with the lowest DI. Three rotational angles, three translations in various dimensions, and one scaling factor make up the seven parameters of the 3D similarity transformation used in the alignment process. The MSE serves as the optimization's loss function, while a stepwise gradient descent method serves as the optimizer. While the OAR gold standard employs nearest neighbour interpolation, CT image volumes are interpolated using the trilinear approach.

The network has a convolutional neural architecture. Three encoding blocks and three decoding blocks make up the CNN blocks, as well as a layer for 3D spatial distortion. It's encoded blocks include a squeeze-out block, two 3D convolutional blocks, and a 1x1x1 convolutional layer, and a squeeze-a 1x1x1 convolutional layer, and a squeeze-3x3x3 convolutional layers, leaky normalisation layers and switchable normalisation layers linear rectified units (leaky ReLUs) are present in the convolutional blocks. The blocks for decoding are applied to expand the feature maps size to speed up the encoding process and to shrink the size of the feature diagrams The network's input information consists of the volume of the template CT image, the target CT image, and the OAR mask for the segmentation output and the template. The output of the network is a deformation field with three channels indicating the x, y, and z offset coordinates. The spatial transformation layer then uses this deformation field to determine the new position of each voxel.

The system includes two 3D U-Nets, LocNet and SegNet. Loc- Net, the first U-Net, is utilized to estimate the position of the target structure by generating a rough bounding box. LocNet takes in a reduced version of the images and masks, and outputs a binary 0-1 classification for each voxel, indicating its presence within the bounding box. A subsequent processing step converts this output into a bounding box applied to the original image, resulting in a smaller volume containing only the target structure. SegNet, the second U-Net, performs segmentation of the target structure from the target volume. SegNet outputs a binary mask volume where each voxel is either 0 (background) or 1 (target). Both LocNet and SegNet are trained independently, with one network for each of the nine structures. The training and testing data for the U-Nets are generated through preprocessing steps. The method enhances the accuracy of segmentation outcomes.

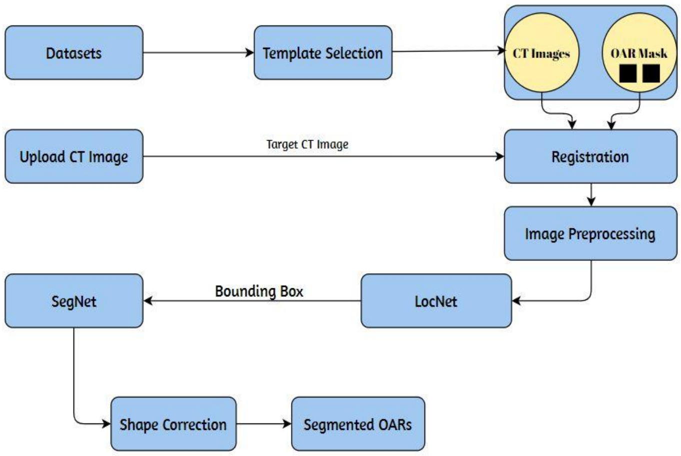

Fig -1: The Proposed Framework

The process begins with projecting the template and the segmentation result onto 2D images. The alignment is done through a conventional registration method, using mean squared error as the cost function and gradient descent as the optimizer. Next, the shape similarity and overlap between the 2D segmentation result and the registered 2D template are evaluated using the dice similarity coefficient (DSC) and Hu moments. Based on these measures, the segmentation result is either corrected or replaced. If both DSC and Hu moments are high, the segmentation result's shape is adjusted by multiplying it by the template. On the other hand, if the Hu moments are high but the DSC is low, the segmentation result is completely replaced with the registered template.

5. CONCLUSION

Medical image segmentation is a field of research that involves developing methods for identifying and separating different structures within medical images. Many different methods have been developed for segmenting various structures from medical images . Some of these methods have been applied to the segmentation of organs at risk in the context of image-guided radiation therapy (IGRT). Research in the area of medical image segmentation focuses on creating techniques for recognising and classifying various components within medical pictures. For the purpose of segmenting diverse structures from medical pictures, numerous approaches have been devised. Our proposed framework uses multiple OAR registration and segmentation techniques, utilizing multiview CT image information. It decomposes the original segmentation into two subtasks: bounding box detection and targeted structure segmentation, by training a separate 3D U-Net for each subtask. The framework will improve the accuracy and speed of the overall segmentation process. SegNet is a great alternative for image segmentation tasks due to its accurate segmentation, precision, and adaptability, whereas LocNet excels at object detection and localization due to its high object detection accuracy, tolerance to partial occlusion, and low computational cost. These architectures have numerous uses in a variety of industries, including autonomous driving, robotics, surveillance, and medical imaging. We may anticipate increasingly more complex designs that will substantially improve the capabilities of computer vision systems as deep learning continues to develop. The segmentation outcomes can be further improved by incorporating shape details during the demarcation process.

REFERENCES

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," CA, Cancer J. Clinicians, vol. 61, no. 2, pp. 69–90, 2011.

[2] Wenbing Lv, Saeed Ashrafinia , Jianhua Ma , Lijun Lu , and Arman Rahmim, "Multi-Level Multi-Modality Fusion Radiomics:Application to PET and CT Imaging for Prognostication of Head and Neck Cancer", IEEE Journal of biomedical and health informatics, vol. 24, NO. 8, August 2020

[3] Bin Huang , Yufeng Ye, Ziyue Xu , Xin Chen,"3D Lightweight Network for Simultaneous Registration and Segmentation of Organs-at-Risk in CT Images of Head and Neck Cancer", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 41, NO. 4, APRIL 2022 .

[4] D. Guo et al., "Organ at risk segmentation for head and neck cancer using stratified learning and neural architecture search," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 4223–4232.

[5] Yibin Wang; Christian Zamiela; Toms V. Thomas; William N. Duggar; "3D Texture Feature-Based Lymph Node Automated Detection in Head and Neck Cancer Analysis" , IEEE International Conference on Bioinformatics and Biomedicine (BIBM),2020,DOI 10.1109/BIBM49941.2020.9313482 .

[6] Yueyue Wang , Liang Zhao , Manning Wang , ans Zhijian song, "Organ at Risk Segmentation in Head and Neck CT Images Using a Two-Stage Segmentation Framework Based on 3D U-Net", IEEE Access, Vol 7, 2019 .

[7] Wenhui Lei, Haochen Mei, Zhengwentai Sun, Shan Ye, "Automatic Segmentation of Organs-at-Risk from Head-and-Neck CT using Separable Convolutional Neural Network with Hard-Region-Weighted Loss", elseveir,Feb 2021

[8] Ngoc-Quang Nguyen and Sang-Woong Lee,”Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network”,IEEE Access , vol 7, 2019 .

[9] Mohamed S. Elmahdy, Laurens Beljaards , Sahar Yousefi, Hessam Sokooti ,”Joint Registration and Segmentation via Multi-Task Learning for Adaptive Radiotherapy of Prostate Cancer”, IEEE Access, VOLUME 9, 2021

[10] Hassan Mahmood, Asim Iqbal, Syed Mohammed Shamsul Islam, ”Exploring Intensity Invariance in Deep Neural Networks for Brain Image Registration”, arxiv ,eess, Sep 2020 .

[11] Jochen Peters, Hauke Schramm, Cristian Lorenz, Jens von Berg,”Automatic Model-Based Segmentation of the Heart in CT Images”, IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 27, NO. 9, SEPTEMBER 2008.

[12] Lasse Hansen, and Mattias P. Heinrich,”GraphRegNet: Deep Graph Regularisation Networks on Sparse Keypoints for Dense Registration of 3D Lung CTs”, IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. XX, NO. XX, XXXX 2021.

[13] Yanling LI , Adams Wai-Kin Kong , and Steven Thng,”Segmenting Vitiligo on Clinical Face Images Using CNN Trained on Synthetic and Internet Images”, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 25, NO. 8, AUGUST 2021