Novel Image Processing and Deep Learning Methods for Head and Neck Cancer Delineation from MRI Data

Baixiang Zhao

Department of Electronic and Electrical Engineering

University of Strathclyde

Glasgow, United Kingdom

A thesis submitted for the degree of

Doctor of Philosophy

2022

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Baixiang Zhao

2022

Acknowledgements

I would like to thank the following people for helping with my PhD research:

First and foremost, I am deeply grateful to my supervisor, Professor John Soraghan, for his continuous encouragement and guidance throughout the PhD project, also his help for identifying funding and career advice. His critical thinking, enthusiasm, and knowledge provide me with strong help for my PhD time.

I would also like to appreciate my supervisor Dr Gaetano Di-Caterina, my senior colleague Dr Trushali Doshi for their technical help and modification of my writing of research output.

I sincerely thank Dr Derek Grose, Dr Clair Paterson, Ms Lisa Hay from Beatson West of Scotland Cancer Centre, Glasgow for their help in data collection, clinical knowledge and ideas, and contributions in this research.

The acknowledge also goes to Mr Miguel Arriscado Guild for his work in the development of RTP Aid software tool in Appendix 1.

I would like to extend my sincere thanks to my friends and colleagues in Glasgow, and my family in UK and China for their company and support during my PhD life. I would like to thank my girlfriend, Zi Lin, for her understanding, patience, and encouragement. I would like to offer my special thanks to my mother for her unvaluable support and encouragement throughout the years.

This research would have been impossible without the research grants from the University of Strathclyde and Beatson West of Scotland Cancer Centre.

Abstract

Intensity modulated radiation treatment aims to achieve accurate treatment of cancer without introducing damage and side effects to organs at risk (OAR). Development of medical imaging technique enables molecular study of cancer to provide quantitative analysis and three-dimensional visualisation to oncologists and radiotherapists for better radiation treatment planning (RTP). Conventional radiation treatment process of head and neck cancer (HNC) requires the manual delineation of gross tumour volume (GTV), abnormal (cancerous) lymph nodes (ALNs), and organs at risk. While the manual delineation subjects to inter- and intra- observer variabilities. Novel image processing and deep learning method for head and neck cancer delineation from MRI data are presented. Firstly, a head and neck ALNs segmentation pipeline including pre-processing of MRI data, knowledge-based detection, and extraction of 3D volume of ALNs is presented. Secondly, a 3D HNC delineation via improving detection and segmentation modules is presented. In these methods, T1 axial MRI slices were firstly pre-processed using contrast enhancement (CE), background noise removal, and bias field correction to improve image quality. 2D slices were then interpolated vertically to reconstruct 3D MRI volume. A knowledge based ALNs detection algorithm was proposed to use throat as key spatial landmark and use fuzzy c-mean (FCM) to classify tissues in intensity, so that find the ALNs on each slice. The 2D detection results were ensembled by a proposed majority voting scheme to give the 3D location of ALN in MRI volume. The 3D volume was finally extracted by 3D level set method (LSM) starting from the detected centre of ALN. The knowledge-based detection method achieved localisation of ALNs by transferring clinical knowledge to automatic algorithm, and by ensemble of results of multiple slices to improve confidence level of detection. This method provided objective 3D segmentation, visualisation, and quantification of ALNs from MRI data, the delineated ALN volumes were comparable (70% in DSC) to conventional manual delineation but with a lower time cost.

Furthermore, a knowledge-based method for segmentation of 3D volume of HNC from MRI data was proposed. This method also has pre-processing, detection, and

segmentation steps. In pre-processing stage, the raw MRI data went through CE, bias field correction, intensity standardisation (IS), and vertical interpolation to generate 3D reconstructed MRI volume with better quality. The target HNC was found by knowledge-based detection on central slice in MRI volume. Detection used FCM to classify tissues, used throat to guide spatial searching, and used localized LSM to refine the region of 2D detection. The detection gave location of 3D volume of HNC and kept spatial information in central slice. A modified 3D LSM was started from detected volume centre to extract the 3D volume of HNC. The extracted volume was smoothed by morphological filtering. The interpolation and 3D segmentation method extracted uniform smooth 3D HNC volume from 2D T1-axial MRI slices. The modified 3D LSM improved the accuracy of volume segmentation via combining spatial information to supress the false positive (FP), i.e., overestimation in segmentation. The proposed automatic 3D segmentation method achieved comparable (70% in DSC) 3D volume of HNC with manual segmentation but lower time cost. Algorithm was further developed to window-based software as useful RTP tool.

Thirdly, a new DCNN for pixel-wise end-to-end segmentation of HNC from 2D T1 axial MRI slices is presented, this architecture was trained and tested on manual consensus outlined from clinicians. The network took similar structure from classical network U-Net, which included an encoder part to extract features and reduce resolution, then a decoder part to recover resolution, fuse features, and classify pixels. The proposed new DCNN improved feature extraction by using a two-pathway encoder with classical and dilated convolutional kernels to combine local and non-local information. To design this DCNN using limited HNC MRI data, data augmentation was used to help the training, depth-separable convolution was used to reduce number of parameters, cross-validation was used to avoid overfitting. The proposed DCNN improved accuracy of HNC segmentation from real MRI data by 5% compared to classical U-Net.

Contents

List of acronyms

2D	Two Dimensional
3D	Three Dimensional
ACM	Active Contour Model
ALN	Abnormal Lymph Nodes
ANN	Artificial Neural Network
ART	Adaptive Radiation Therapy
BoT	Base of Tongue
CE	Contrast Enhancement
CO	Clinical Oncologist
CT	Computed Tomography
CTV	Clinical Target Volume
DICOM	Digital Imaging and Communications in Medicine
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DSC	Dice Similarity Coefficient
FCM	Fuzzy C-means
FCN	Fully Convolutional Neural Network
FI	Fourier Interpolation
FFT	Fast Fourier Transform
GAN	Generative Adversarial Network
GC	Graph Cut
GPU	Graphical Processing Unit
GTV	Gross Tumour Volume
GUI	Graphical User Interface

HE	Histogram Equalisation
HNC	Head and Neck Cancers
IIH	Intensity Inhomogeneity
IS	Intensity Standardisation
IMRT	Intensity Modulation Radiation Therapy
LSM	Level Set Method
MFCM	Modified Fuzzy c-means
MHD	Modified Hausdroff Distance
MR/MRI	Magnetic Resonance/Magnetic Resonance Imaging
OAR	Organs at Risk
PET	Positron Emission Tomography
PTV	Planning Target Volume
PVE	Partial Volume Effect
ROI	Region of Interest
RT	Radiotherapy/Radiation Therapy/Radiation Oncology
RTP	Radiotherapy Treatment Planning
SRG	Seeded Region Growing
SUSAN	Smallest Univalue Segment Assimilating Nucleus
SVM	Support Vector Machine
T1 + Gd	Gadolinium Enhanced T1-Weighted
TE	Echo Time
TNM	Tumour-Neck-Metastases
TR	Repetition Time

List of figures

List of tables

List of symbols

Symbol	Meaning	Page
I	An image	23
C_{weber}	Weber contrast	23
$C_{Michelson}$	Michelson contrast	24
C_{RMS}	RMS contrast	24
$I(x, y)$	A pixel on image I at coordinate (x, y)	27
$O(x, y)$	A pixel on raw signal O at coordinate (x, y)	27
$\eta(x, y)$	A pixel on noise N at coordinate (x, y)	27
$B(x, y)$	A pixel on bias field B at coordinate (x, y)	29
c_h	Cumulative density function (CDF) of histogram	42
T_c	Transformation on CDF of histogram	42
se	Structure element for morphological filtering	46
$\oplus$	Dilation	46
$\ominus$	Erosion	46
$S(x, y)$	A pixel on Bias field corrected image at (x, y)	49
V(s)	Vector represents Snake model	59
E_{snake}	Energy to update Snake model	59
E_{external}	External energy of Snake model	59
E_{internal}	Internal energy of Snake model	59
ϕ	Level set representation of a contour	60
t	Time	60
C(t)	Contour at time t	60

F	Force for updating Level sets	60
$\nabla\phi$	Gradient of level sets	61
μ_{in}	Mean value of points inside a contour	61
μ_{out}	Mean value of points outside a contour	61
$D(C_A, C_B)$	Distance between two sets (two contours)	84
$d(c_a, C_B)$	Closest distance from a point in set A to set B	84
T	Threshold to binarize an image	96
I_f	Foreground part of an image	96
I_b	Background part of an image	96
F_{VL}	Vertical fuzzy rules for throat detection	97
μ_{VL}	Vertical membership value for throat detection	97
F_H	Horizontal fuzzy rules for throat detection	97
μ_h	Horizontal membership value for throat detection	97
F_{VLH}	Fuzzy rules for throat detection	98
$\mu_{F_{VLH}}$	Membership value for throat detection	98
B_I	Candidate binary regions in image for throat detection	98
J_T	Objective function for fuzzy c-mean	100
$\mu_{ik}^{m_f}$	Membership value of kth pixel belongs to ith cluster	100
m_f	Fuzziness of membership function	100
$d^2(v_i, I_k)$	Distance from pixel I_k to cluster centre v_i	100
W	Probability map for ALN detection	104
α, β, θ	Weights of subcomponents of W	104
$W_{location}$	Location component of probability map	105

a-h	Spatial thresholds define subarea of probability map	106
w_{size}	Size component of probability map	107
R	A region in binary image	107
(x_c, y_c)	Centre of a region	107
$W_{eccentricity}$	Eccentricity component of probability map	108
C_e	Distance from an ellipse centre to the focus along the major axis of ellipse	108
A_e	Length of major axis in an ellipse	108
Rad	Radius of a region	110
C_g	Centre of a group of regions	110
μ_{LN}	Mean intensity of detected LNs	112
ε_{LN}	Standard deviation of detected LNs	112
λ	Weighting factor between internal and external force	112
C_I	Image with binary regions refined from B_I	128
$R(B_I)$	Right half of binary image B_I	129
$L(B_I)$	Left half of binary image B_I	129
$\delta(\phi(x,y))$	Pixels around $\phi = 0$	133
$A_{\mu_{in}}$	Area of local region inside ϕ	134
$A_{\mu_{out}}$	Area of local region outside ϕ	134
$M_l(x,y)$	Local region around point (x,y)	134
μ	Lower bound of the gray-level of the target object	137
$g$rad	Gradient map	137
$\omega_L(I)$	Location weighted image	137
div	Divergence	137

σ	Standard deviation	137
μ_{seg2D}	Lower bound of segmented 2D tumour	137
V_{in}	Input vector	156
V_{out}	Output vector	156
M_{conv}	Sparse matrix represents convolutional filter	156
γ_B, β_B	Trainable parameters to scale and shift the distribution of input mini-batch	161
$\hat{x}_i$	Normalization of values in the mini batch	161
$\mathcal{B}$	A mini batch input of neural network	161
$\mu_{\mathcal{B}}$	Mean value of a mini batch	161
$\sigma_{\mathcal{B}}^2$	Standard deviation of mini batch	161
q	Predicted likelihood of a class	162
p	Classification target	162
$H(\mathrm{p}, \mathrm{q})$	Loss function between prediction and target value	162
L_{dice}	Dice loss	162
L_{ce}	Cross entropy loss	

Chapter 1 Introduction

1.1. Preface

Head and Neck Cancer (HNC) refers to several types of cancer normally located in the head and neck body regions, it includes tongue cancer, throat cancer, salivary gland cancer, laryngeal cancer and so on [1]. According to American Cancer Society [2], in 2020 there were 1,806,590 new cancer cases and 606,520 cancer deaths occurring in the United States. Oral cavity and pharynx cancer occurs more among male patients, which accounted for 4% of new cases in 2020. Furthermore the death rate of HNC has increased in the past decade in the USA [2]. Data from Cancer Research UK show that, from 2015 to 2017 there were more than 12,000 HNC cases in UK each year, leading to 3989 deaths every year. In 2017, HNC is the 8th most common cancer in the UK, accounting for 3% for all new cancer cases [1]. The incidence rates of HNC in UK increased by 20% in the past decade. Major risk factors of HNC are from alcohol and smoking [3].

Early-stage HNC is treated by surgery or radiation therapy (RT), and concurrently chemoradiotherapy will be locally applied to cancerous tumours [3]. RT can be applied to both the primary site and to the lymph nodes. As an alternative approach to surgery, RT can achieve similar oncological outcomes but conserve the organs at risk (OAR) instead of resecting it [4]. Advanced diagnostic medical imaging techniques provide essential tools for curative surgery and RT, it enables cellular levels visualisation and quantification of cancer regions. However, manually determination and segmentation of tumours from medical imaging relies on clinical expertise and available facilities, also it is time consuming and suffering from subjective variabilities. Thus, development of automatic tumour delineation methods can decrease subjective variabilities, and provide objective and consistent assistance to oncologists to reduce their work and improve RT.

This research aims to design advanced computer vision techniques to improve the understanding of magnetic resonance (MR) images for HNC analysis. The results of this research will lead to automatic systems for 3D segmentation of base of tongue

(BoT) and throat tumour from MR images. This system will provide powerful tools to clinicians for efficient cancer treatment, and finally improve patients' quality of life.

1.2. Research motivation

Along with surgery and chemotherapy, radiotherapy plays a vital role in the treatment of HNC. The radiation treatment planning (RTP) refers to the procedure of planning the precise dose of radiation on the target areas of patients. The aim of planning is to apply enough RT to cancerous area avoiding damage to organs at risk. To achieve the goal of accurate RTP, precise delineations of targets such as cancerous tumour should be provided to radiotherapist to guide the process. In the classical practice, the outlines of targets are manually delineated by clinicians. However, the manual delineation suffer from significant inter- and intra- subjectiveness [5, 6]. It also results in a large time cost and relies on expertise of clinicians [7, 8]. Thus, it would be beneficial to introduce computer algorithms to assist RTP to build a platform for the automatic delineations of cancerous tumours and abnormal lymph nodes. The motivation of work in this thesis is to develop novel image processing, machine learning, and deep learning algorithms for the automatic 3D segmentation of head and neck tumour, and abnormal lymph nodes from real MRI data. The proposed auto-segmentation algorithms will improve the RTP in several aspects including providing accurate objective contours, saving time cost, and reduce workload for clinicians.

1.3. Research objectives

The research objectives of this thesis include:

1) Pre-processing of MRI data to get well-prepared images as well as volumes for head and neck cancer analysis.
2) Detection of abnormal lymph nodes from head and neck MRI data
3) Detection of cancerous tumour from head and neck MRI data
4) Segmentation, and visualisation of 3D volume of lymph nodes from head and neck MRI data
5) Segmentation, and visualisation of 3D volume of cancerous lymph nodes from head and neck MRI data
6) Semantic segmentation of cancerous tumour area from head and neck MRI data.

1.4. Summary of original contributions

The original contributions presented in this thesis are:

1) A novel pre-processing workflow to prepare the raw 2D MRI data into 3D volume with enhanced quality is proposed. The 3D reconstruction involves Fourier interpolation. The image quality enhancements include noise removal, bias field correction, contrast enhancement for each image, also the intensity variation standardisation between images of each patient. The proposed pre-processing work adjusts the distribution range of intensities of MRI images so that improve the efficiency of intensity-based detection and segmentation work.
2) A novel knowledge-based algorithm for the detection and localization of cancerous tumours and abnormal lymph nodes is developed. The detection scheme utilizes image processing techniques of fuzzy c-mean, watershed, morphological filtering, a proposed probability map together with the anatomical knowledge of head and neck region from clinical experts, to build a detection pipeline based on intensity and spatial information of targets. The detection pipeline starts from an MRI slice as input and gives course localization of cancerous tumour as well as lymph nodes which are in binary mask form as output. With the success detection from complex environment of head and neck area on single slice, 2D interactive segmentation methods such as level set methods can be triggered and guided.
3) The third contribution combines 2D detection and a proposed majority voting process to fuse 2D detections on single slices. This process considers inter-slices spatial relationship of 2D detections so that firstly refine the 2D results to remove false positive then providing 3D rough location of cancerous tumours and abnormal lymph nodes, which indicates the centre of target and can be used as seed of interactive 3D segmentation.
4) The fourth contribution proposes a spatial-constrained 3D level set method for the segmentation of cancerous tumours. The proposed modified level set method uses spatial-constrained level set function, which locates and tracks the surface of tumour's volume automatically under the penalty of spatial information obtained from anatomical knowledge and previous detection. The proposed novel level set

method supresses the false positive in segmentation so that improve the accuracy of segmentation results.

5) The fifth contribution proposes a novel two pathway dilated deep convolutional neural network which is a modified version of classical U-Net. The proposed DCNN introduces two pathway structures and dilated convolution scheme to improve the multi-resolution feature fusion without adding heavy computation cost. The proposed DCNN achieves end-to-end 2D segmentation of cancerous tumours with improved accuracy compared to original U-Net structure.

1.5. Author's publications

Engineering conference papers:

1. B. Zhao, J. Soraghan, G. Di-Caterina, L. Petropoulakis, D. Grose, and T. Doshi, "Automatic 3D segmentation of MRI data for detection of head and neck cancerous lymph nodes," in 2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2018, pp. 298-303.

2. B. Zhao, J. Soraghan, D. Grose, T. Doshi, and G. Di-Caterina, "Automatic 3D detection and segmentation of head and neck cancer from MRI data," in 2018 7th European Workshop on Visual Information Processing (EUVIP), 2018, pp. 1-6.

3. B. Zhao, J. Soraghan, G. Di Caterina, and D. Grose, "Segmentation of head and neck tumours using modified U-Net," in 2019 27th European Signal Processing Conference (EUSIPCO), 2019, pp. 1-4.

Medical conference:

1. B. Zhao, J. Soraghan, G. Di Caterina, and D. Grose, " Automatic 3D Segmentation of Head and Neck Tumours and Cancerous Lymph Nodes," in 2019 National Cancer Research Institute (NCRI) Cancer Conference, Glasgow, UK.
2. B. Zhao, J. Soraghan, G. Di Caterina, and D. Grose, " Segmentation of Head and Neck Cancer from MRI Data using Deep Neural network," in 2019 National Cancer Research Institute (NCRI) Cancer Conference, Glasgow, UK.

1.5. Thesis organisation

The rest of this thesis is organized as following:

Chapter 2 provides a review of biomedical imaging techniques and treatment of head and neck cancer. Specifically, the principle of the magnetic resonance imaging and its application on radiation treatment planning of head and neck cancer is covered. The fundamental physics of MRI and characteristics of MRI image data are reviewed. The anatomical knowledge of head and neck region will be introduced, as well as the major cancers originated in this area. Then the regular treatment approaches for HNC is introduced through the RTP workflow. Several shortcomings that exist in current RT of HNC are specified that reveal the motivation of developing automatic software tools for contouring of head and neck cancerous tumour.

Chapter 3 presents a review of image processing techniques for enhanced RTP. These include image processing techniques which can be applied for pre-processing of medical images such as image enhancement and de-noising. This chapter introduces image processing techniques which can interpret and analyse sophisticated patterns in images, so that they can be used for detection and segmentation of targets such as tumours from medical images. This chapter also reviews both conventional image processing methods and machine learning approaches as both will be explored for the development of automatic contouring work in this Thesis.

In Chapter 4 the specific real MRI data sets that are used for development and testing of automatic algorithms are presented. Specific metrics which are used for the evaluation of automatic algorithms are shown. These will help the understanding of the work procedure and results in this thesis.

Chapter 5 presents the knowledge-based 3D segmentation algorithm for abnormal lymph nodes. The proposed knowledge-based algorithms include a pre-processing, knowledge-based detection, level set method (LSM) segmentation, and post-processing pipeline. The pre-processing pipeline enhances quality of raw MRI data and constructs 3D volume from 2D slices. The detection uses anatomical knowledge to search the ALNs on each slice and fuse all results to get the 3D location, so that start the 3D LSM segmentation phase from detected ALN centre. The post-

processing step uses 3D morphological filtering to smooth the surface of obtained 3D ALN volume and remove outliers. The results of proposed algorithms on real MRI data will demonstrated and analyses in these chapters.

Chapter 6 presents the knowledge-based 3D segmentation algorithm for HNC. The proposed knowledge-based algorithms include a pre-processing, knowledge-based detection, level set method (LSM) segmentation, and post-processing pipeline. Compared to ALN detection in Chapter 5, this Chapter focusses on a central slice for the detection of HNC, and watershed transform is additionally introduced in the pipeline. The segmentation part uses a modified 3D LSM to tackle the challenge in HNC segmentation such as arbitrary shapes and complex texture. The results of proposed algorithms on real MRI data will demonstrated and analyses in these chapters.

Chapter 7 will present the deep learning-based segmentation network for HNC. The proposed network modified the classical U-Net by introducing two-pathway structure and dilated convolution to improve the segmentation accuracy. The main sub-modules of the proposed network and training settings will be introduced in this chapter. The results on real MRI data measured with cross-validation will be demonstrated and analysed.

Chapter 8 will conclude the thesis and present some potential work that can be further explored in the future.

Chapter 2 Biomedical Imaging for Head and Neck Cancer (HNC)

2.1. Introduction

The majority of malignant neoplasms of the head and neck region starts in the upper aerodigestive tract, arising from mucosa lining the oral cavity, oropharynx, hypopharynx, larynx, sinonasal tract, and nasophynx. Squamous cell carcinomas (SCCs) comprise 95% of all HNCs [4], and human papillomavirus (HPV) has been recognized as a cause of a subset of HNSCCs. Smoking and drinking alcohol take up an estimated 75% of HNC cases, but HPV-related HNC is biologically and clinically different from tobacco related HNC [9]. Symptoms and signs of head and neck cancer includes swelling anywhere in the HN or a sore that does not heal, red or white patch in the mouth, frequent nose bleeding, painful swallowing, dysphagia, loss of weight, difficulty breathing, and double vision [4, 9].

Radiotherapy, surgery and chemotherapy are general options for head and neck treatment [10]. Because of the complicated anatomy of the head and neck, conventional open surgical approaches can lead to significant functional impairment. The difficulties of HNC surgery appears in preserving cervical spine functional structure such as great vessels, nerves and muscle during neck dissection [4]. Alternatively to surgery, radiotherapy provides organ-preserving therapy rather than resecting it [11]. RT has significant advantages in laryngeal and pharyngeal tumours as the organ-preserving therapy makes it possible to retain speech and swallowing function. However, RT can result in significant toxicities to tissues surrounding the tumour, thus causing side effects such as dermatitis, myelitis, and hearing loss. Thus, the very key requirement of radiotherapy is to deliver accurate and appropriate dose to tumour and surrounding body tissues. Chemotherapy can be used as monotherapy for early-stage cancers or be concurrently applied with RT or surgery for advanced-stage cancers in HNC treatment. Monotherapy is more preferred in HNC treatment as it leads to fewer side effects. But systemic treatment is necessary for advanced-stage cancers, to eliminate tumours, and also prevent distant metastasis or recurrent disease [4].

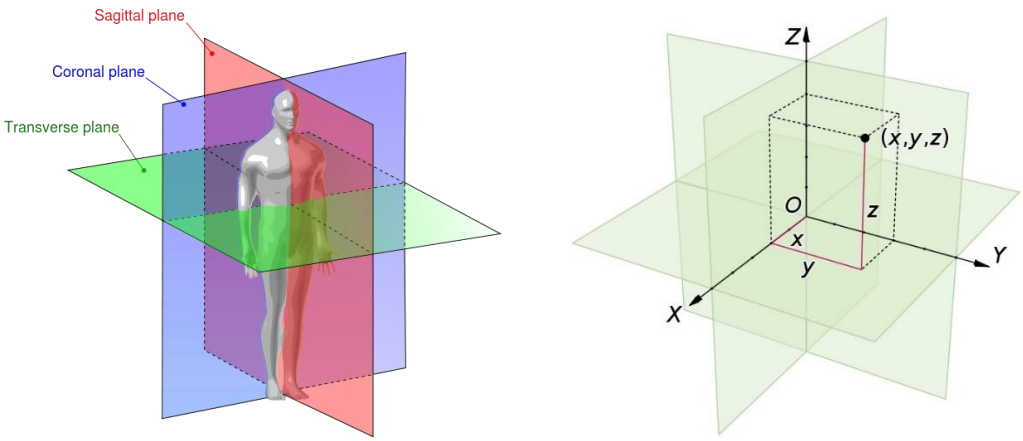
Medical imaging uses non-invasive techniques and instruments to reveal and record internal body structure of humans or animals, which cannot be viewed using visually observed or physically examined from exterior of the body [12, 13]. Medical imaging techniques create 2D/3D representations of bone, soft tissues, body fat, muscle, air, and different organs [14], it plays an import part in diagnosis, prognosis, screening, staging, treatment evaluation, and image-guided intervention [15, 16]. Analysis of medical images helps medical experts locate, quantify, and extract features of target organs, tissues or diseases [15]. However, the manual understanding of medical images introduces subjective bias. Even the images are analysed by very experienced clinicians, the interpretation of information in images varies across individuals.

In this chapter, the head and neck anatomical knowledges and HNC treatment procedure will be reviewed in section 2.2 to help explain the research target. Then a widely used medical imaging technique called MRI will be introduced in section 2.3, including its principles and how it is related to HNC treatment. Also, the difficulties of manual interpretation of MRI data in HNC treatment will be discussed in section 2.4.

2.2 Head and Neck cancer and its treatment

2.2.1 Head and neck region

The head and neck regions have several main sites, including the nose and paranasal sinuses, oral cavity, neck, larynx, pharynx, salivary glands, thyroid and parathyroid glands [17]. Before describing the HN region, the idea of anatomical planes is now described. As illustrated in Fig 2.1 there are three main orthogonal anatomical planes (also generally used as main imaging planes), axial (also known as transverse, or horizontal), sagittal (also known as median) or coronal (also known as frontal) planes.

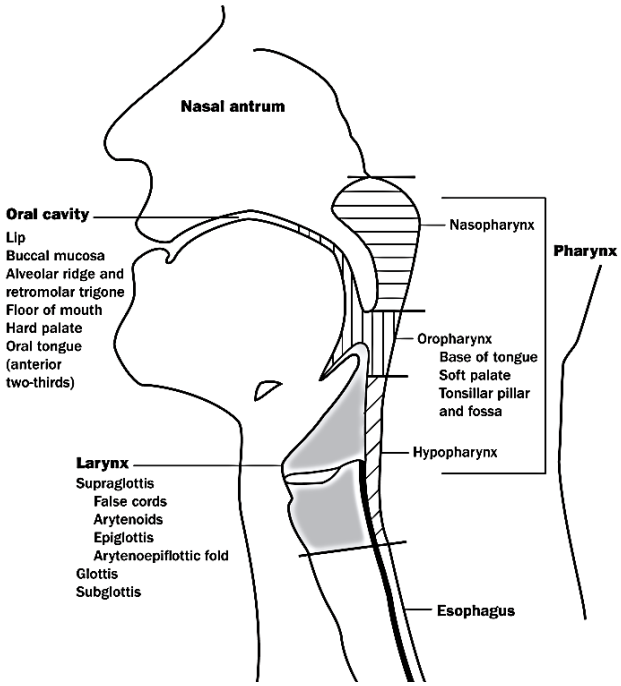


(a) (b)

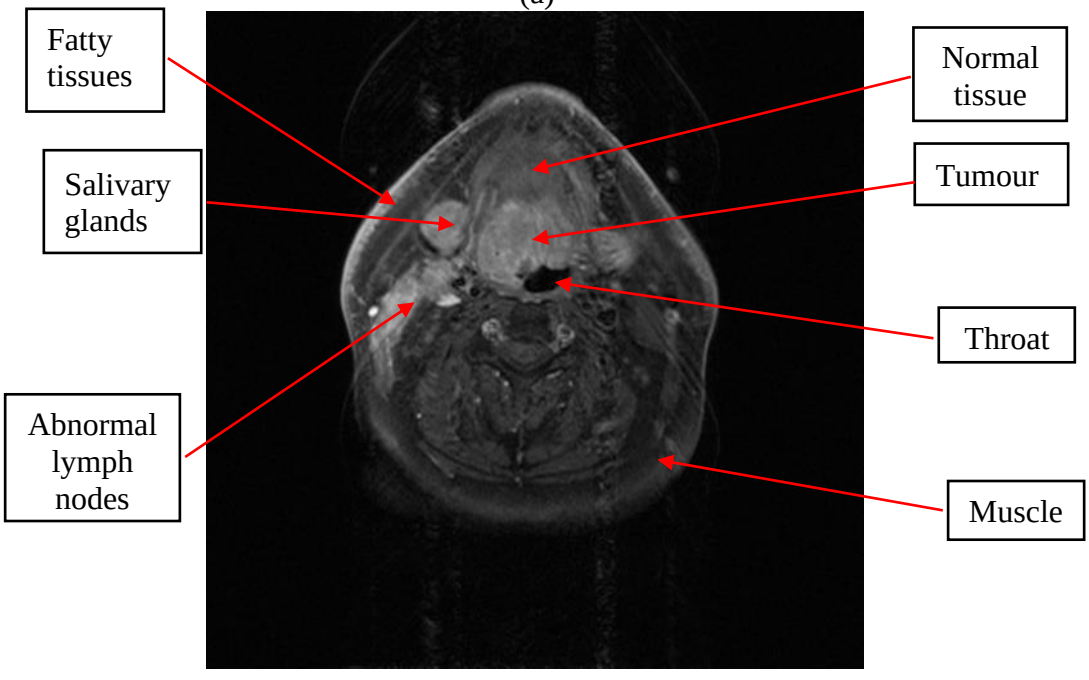
Fig. 2.1 Anatomical planes. Left, a demonstration of orthogonal planes around a human body model. Right, a 3 dimensional coordinates [18].

As shown in right of Fig. 2.1(b), the three major planes around the human body represent the directions. Top to bottom represent the axial (transverse) plane; front to back, represent the coronal (frontal) plane and side to side represent the sagittal (median) plane. It can be also represented in 3 dimensional coordinates: assuming body orients to positive x axis, then x-axis means front to back, y-axis represents side, and z-axis represents top to bottom. Thus, the axial plane is xy plane, coronal plane is yz plane, and sagittal plane is xz plane as illustrated in Fig 2.1(b)

Taking the basic plane knowledge, Fig. 2.2 shows the demonstrating sites and sub-sites of head and neck regions.



(a)



(b)

Fig. 2.2 Anatomic sites and sub-sites of the head and neck region. (a) the sagittal display of head and neck sites and sub sites [19]. (b) a real axial MRI slice with marked sites.

Fig. 2.2(a) displays the main anatomic sites of head and neck region. From top of head, the nose is the first part of the respiratory system. It has external part and internal part. The external part is covered by skin. The internal part is formed by nasal cavity [17]. The external part is horizontally between the eyes, and vertically between eyes and mouth. The internal nasal cavity locates at the start of the upper airway and is divided in the midline by the nasal septum [20].

Oral cavity is the most anterior part of digestive system. Outside the oral cavity it is enclosed by upper and lower lips. Inside the cavity, the borders are mylohyoid muscles, gingivobuccal regions, hard palate, anterior tonsillar pillar, and circumvallate papillae [20]. The tongue lies on the floor of the mouth, fill in most of space of oral cavity [17].

The pharynx is a 12-14 cm long musculomembranous tube [17]. It is divided into three parts: nasopharynx, oropharynx, and laryngopharynx [21]. The nasopharynx is behind the nasal cavity; the oropharynx is behind oral cavity; and the laryngopharynx is at the lower level of cricoid bone, behind the larynx [20].

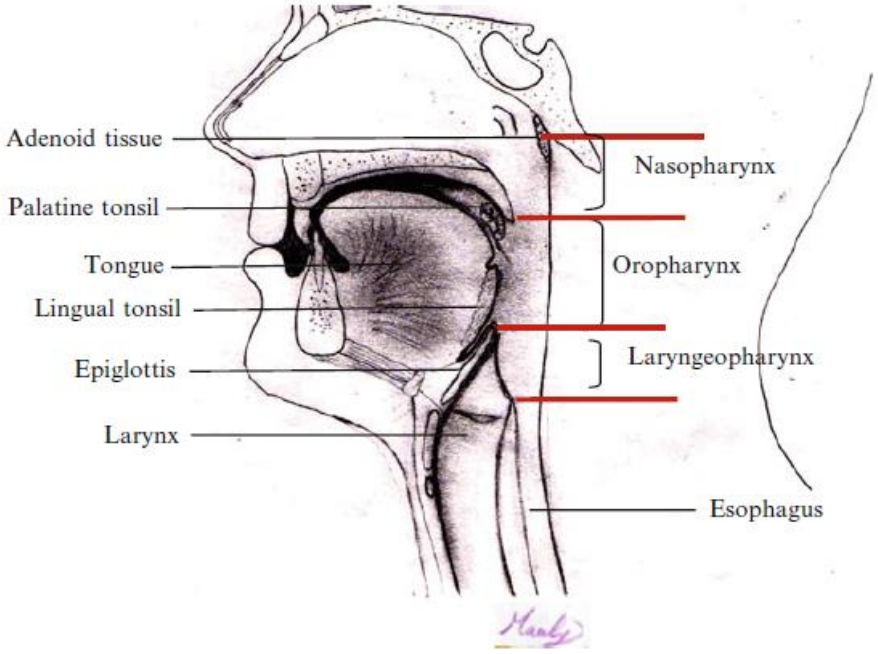


Fig. 2.3 An illustration of sagittal view of the pharynx [17].

Fig. 2.3 shows the illustration of sub-sites of pharynx. Vertically the pharynx region ends at the larynx. The larynx is continuous with the pharynx superiorly and with the trachea inferiorly [17]. The larynx is supported by laryngeal skeleton, covered by a mucosal surface. Between the surface and supporting skeleton, there is a soft tissue layer containing fat, ligaments and muscular structures [20].

The major head and neck region of interest (ROI) of this thesis are the pharynx and larynx regions. Fig. 2.2(b) shows an axial MRI slice of pharynx region (base of tongue). Some key regions are marked with text box, they are normal tongue tissues, normal muscle, tongue base tumour, cancerous lymph nodes, fatty tissues, salivary glands, and throat.

2.2.2 Head and neck cancer

This subsection will review the head and neck cancer. Among all anatomical sites of head and neck region, more than 95% of all malignant neoplasms are SCCs [17, 19, 20]. The most frequent are the epithelial malignancies of the mucous membranes of the upper aerodigestive tract, which is shorted as HNSCC (head and neck squamous cell carcinoma). The second most abundant head and neck cancers are glandular neoplasms, which are mainly arising from thyroid, and also minorly from salivary glands [20].

According to Cancer Research UK, each year there are 12,238 new HNC cases between 2015 to 2017 in UK, that lead to an average of 4077 deaths every year [1]. Table 2.1 lists statistics of HNC in UK in terms incidence, mortality rates, and survival rates over factors such as gender and ages. Based on the data acquired from 2009-2013 in England, between 19%-59% of patients diagnosed with HNC survive their disease for ten years or more; between 28%-67% of patients survive five years or more; between 61%-86% of patient survive one year or more. Among subtypes (at different sites) of HNC, the salivary glands cancer has highest survival rate, and hypopharynx has lowest survival rate. In terms of genders, HNC 10-years survival rate in England is generally similar in men and women (data also from 2009-2013). In terms of ages, the patient aged 15-49 generally has higher survival rate compared to other age groups. And the data shows that the survival of salivary glands cancer is mostly influenced by age, 90% of 15-49 age group can survive 5 years or more, while only 50% of patients

aged 70-89 survive 5 years or more. In conclusion, the HNC is the 8th most common cancer in the UK, accounting for 3% of all new cancer cases (data in 2017); the HNC is the 16th most common cause of cancer death in the UK, accounting for 2% of all cancer deaths (data in 2018) [1].

Table 2.1 UK HNC statistics [1]

	Over Gender	**Over Age**	**Additional Information**
Incidence (2016-2018)	8600, 3900 new cases per year in males, females	Highest on 70 to 74 years old	Most found at tongue. Around 2300 cases each year are linked with deprivation
Mortality (2016-2018)	2900, 1200 new deaths in 2018 in males, females	Highest in 90+, More than 36% in aged 75+ years old	Since 1970s the mortality rates are decreasing but the rates increased 17% in last decade
Survival (2009 - 2013)	Similar in males and females, 19-59% for ten years survival	Highest (34%) in 15-49 years old	Hypopharyngeal cancer has lowest 5-year survival rate at 27%

According to American Cancer Society [2], in 2020 there are 1,806,590 new cancer cases and 606,520 cancer deaths occurred in the United States. Oral cavity and pharynx cancer occurs more among male patients, which accounts for 4% of new cases in 2020. And the death rate of HNC is increasing in past decade in USA [2]. In 2021, there are 54,010 new oral cavity and pharynx cancer cases in USA that resulted in 10,850 deaths. The number of new cases and deaths of male are over 2 times of female [22]. The death rates of oral cavity and pharynx cancer in America has increased by 0.5% per year from 2009 to 2018.

From Global Cancer Statistics 2020 [23], new cases of lip and oral cavity, larynx, oropharynx, hypopharynx, and salivary glands take about 4.2% out of all cancer types, and lead to 3.9% deaths out of all cancer types. In terms of regions, the lip and oral cavity cancers are highly frequent in South Central Asia (e.g., India and Sri Lanka) as well as Melanesia. The incidence of head and neck cancer is increasing with age [19]. In terms of gender, the HNC is more frequent in men.

Based on decades of research [24], the chronic use of tobacco is the most prevalent and most correlated risk factors of HNSCC. Alcohol is another powerful risk factors for incidence of pharyngeal and laryngeal tumours. It shows that the combination usage of tobacco and alcohol induces multiplicative risks on incidence of HNSCC [19]. Another type of risk factors is poor dietary habits, the nutrition deficiencies, nitrosamines consumption, betel quid chewing, picked vegetables and very hot food and beverages are additional suspected risk factors for HNSCC [23]. Chronic sunlight or radiation exposure is also an important risk factor of HNSCC. Viral infections are also correlated to the carcinogenesis of HNSCC [20].

The prevalent symptoms of HNC include mucus problem, mouth and throat sores, taste problem, dysphagia, xerostomia, pain and fatigue [25]. The signs and symptoms also include globus sensation, hoarseness, a change of speech ability, nose bleeding (epistaxis), epiphora, haemoptysis, stuffiness of the ears, and trismus [19].

2.2.3 Head and neck cancer treatment

In the last section, many aspects of HNC were reviewed such as risk factors, incidence, and symptoms. In this section the general treatment pipeline of HNC will be reviewed. The review will include diagnosis and treatment.

There are multiple approaches for screening and examination of HNC. The physical examination is the best way for detecting lesion of the upper aerodigestive tract. The examination should follow a systematic approach, and here some examples are concluded in the following Table 2.2 [19],

Table 2.2 The systematic physical examination of HNC [19]

Site	**Examination**
Skin/scalp	Search of ulcers, nodules, and pigmented or other suspicious lesions
Cranial nerves	Assessing eye motion, testing face sensation, examining functionalities of facial muscle, testing hearing and so on
Eyes/ears/nose	Evaluation of sign of mass effect, abnormal drainage, bleeding, or effusion
Oral cavity	Inspection of teeth, gingivae, and entire mucosal surface and lymphoid tissue of tonsillar pillars. Evaluation of Halitosis and tongue mobility.
Neck	Systematic examination of any mass

The physical examination can be conducted by Endoscopy. The indirect laryngoscopy (mirrors) is used for vocal cords check so that inspecting mobility and asymmetry, which may indicate a hidden tumour [19]. The direct laryngoscopy (nasopharyngoscopy) enables the thorough inspection of upper aerodigestive. Other endoscopies include esophagoscopy and bronchoscopy are also widely used with laryngoscopy to inspect hidden or known primary HN tumours.

Another category of examination is biopsy, such as punch or cup forceps biopsy, fine-needle aspiration (FNA), core biopsy, and open biopsy. These approaches can be used for examination of mucosal lesions, distinguishing a metastatic SCC from other malignant histologist, inspection of neck mass.

Medical imaging techniques are also widely used in the diagnosis of cancer. The imaging techniques use non-invasive methods visualise interior body structures and tumour sites. Medical imaging provide important clues to diagnosis when biopsies are difficult to obtain in deeply located lesions or when located in close relation to vital structures [17]. Also, medical imaging is capable of accurate localization of malignant tumour and metastatic lymph nodes, this makes medical imaging play important roles

not only in diagnosis but also treatment planning as well as prognosis. Much more details of medical imaging techniques will be introduced in the next section.

The pathological analysis is also important for diagnosis of HNSCC, which includes analysing the histologic grade and morphologic growth patterns of lesions. Based on knowledge of spectrum of malignancies seen in the head and neck, the pathologists classify specimens from head and neck into mundane benign specimens and SCC [17]. Pathologists also guide treatment decisions by providing specific data elements such as virally driven malignancy, margin status, and the precise depth to which tumour invades [26]. Analysis and usage of histopathology features requires nuanced interpretation by well-informed pathologists, and the collaboration of multidisciplinary treatment team.

As mentioned previously, the HNC can start from many sites of head and neck region (Fig. 2.2). It is essential to be aware of the cancer can grow and spread from primary sites to adjacent lymph nodes in head and neck region, which is called lymph node metastases, and it plays import role for prognostic [19]. Here the TNM staging system should be introduced.

The TNM is short for tumour-node-metastasis staging system, which is firstly reported in 1940s [27]. Currently, the TNM staging system of the American Joint Committee on Cancer (AJCC) and International Union Against Cancer (UICC) is widely used for staging of head and neck tumours. The TNM staging system evaluate the extent of cancer through three aspects: size of main tumour (T), involvement of nearby lymph node (N), and extent of distant metastases (M) [28]. The stages of the extent of cancer can be represented by T, N, M separately as shown in Table 2.3,

Table 2.3 Staging of cancer through T, N, M aspects separately

T	N	M
TX: Primary tumour cannot be measured	NX: Regional lymph nodes cannot be assessed.	M0: No distant metastases
T0: No evidence of main tumour	N0: No metastases in regional lymph nodes.	M1: Evidence of distant metastases
T1, T2, T3, T4: The higher the number of T, the larger the tumour is, and more significant local extension is.	N1, N2, N3: The higher the number of N, the higher number of regional lymph nodes have cancer	

TNM measurements can be combined to give five brief stages, which is more generally used by doctors for describing the extent of cancer. The five stages are in following Table 2.4,

Table 2.4 Staging based on combination of all T, N, M aspects

Stage	Interpretation
Stage 0	Abnormal cells are present but have not spread to neighbouring tissue. These cells could become cancer but not yet.
Stage I, Stage II, and Stage III	Cancer is present. The number of the stages stands for the extent of the cancer spread into neigh tissues, the higher the more spread.
Stage IV	Metastases present to distant (from primary site) parts of body

Radiotherapy, surgery, and chemotherapy are general options for head and neck treatment. As previously mentioned, each option has its limitation. In practical treatment planning of HNC, the early stage (I, II) cancer can be effectively finished by single modality such as surgery or radiotherapy, while for stage III or Stage IV caner, multiple treatment options are combined [17].

The surgery may be definitive as the standard care of salivary glands, thyroid, or oral cavity cancers [17, 26]. But in many cases, the surgical and nonsurgical therapies have equivalent effectiveness, such as some oropharyngeal and laryngeal cancers. As open surgical resection these tumours may result in functional impairment, such as dysphagia and aspiration. Thus, nonsurgical therapies such as radiotherapy are more preferred when they can perform as well as surgery. Recently, many advanced surgical techniques are introduced to modify traditional surgical approaches. For example, minimally invasive surgical techniques such as transoral laser microsurgery (LTM) [29], transoral robotic surgery (TORS) [30] are developed which can access to tumour site and avoid disruption of the surrounding anatomical structures [4].

Like surgery, radiotherapy can also be used as single modality for treatment of primary cancer site as well cancerous lymph nodes. Because radiotherapy is nonsurgical, it can preserve the organs compared to surgery. However radiotherapy can cause toxicities to tissues nearby the target tumour, and lead to acute side effects such as mucositis and dermatitis [4]. The radiotherapy techniques have been used for decades, and steadily keeps improving through years. Nowadays, the intensity-modulated radiation therapy (IMRT) is the standard radiation treatment technique for HNC. IMRT can deliver varying radiation distribution, so that minimize the radiation doses to neighbour normal tissues [26]. The radiation treatment planning will be more specifically described in the following sections.

Different from surgery and radiotherapy, chemotherapy is generally not solely used in HNC treatment, instead it is combined with other treatment modalities to give systematic treatments. The chemotherapy can be used to enhance local effects of radiation treatment, or as preoperative therapy before surgery or radiation, or as adjuvant concurrently used with radiotherapy. Chemotherapy is only singly used in cases with distant metastasis or recurrent disease, which are not suitable for surgical

resection nor radiotherapy [4]. However, the systematic therapies with chemotherapy and radiotherapy can also cause side effects, such as mucositis, acneiform rashes, and hydro electrolytic disorders.

2.3. Medical imaging and Magnetic resonance imaging (MRI) data

2.3.1 Medical imaging techniques

Medical imaging is a set of techniques used to visualise the internal structures of humans and animals. The purpose of the medical imaging is to non-invasively record the interior anatomical structure and functional state of the body, which cannot be completed through exterior visual inspection or physical examination [12, 13]. In [31], it is suggested that this medical imaging and biomedical informatics can be regarded as two revolutions which have changed the nature of medicine and related research. In modern medicine, medical imaging is crucially and widely used for diagnosis, prognosis, and treatment assessment of diseases including but not limited to cancers. This section reviews some commonly used imaging techniques such as Computed Tomography (CT) [15], Positron Emission Tomography (PET) [32], X-ray [32, 33], and Ultrasound [34]. Some developing approaches, such as infrared thermography (IRT) are under development [12, 35]. As the Magnetic Resonance Imaging (MRI) is the main topic of the thesis, it will be covered in detail in the following sections.

2.3.2 MRI Fundamentals

The research of medical imaging analysis in this thesis is based on the imaging techniques called Magnetic Resonance Imaging (MRI). MRI can provide high-quality images comparable to CT without using harmful radiation [13], but the cost is relatively high.

During MRI scanning, patients are placed in strong uniform and static magnetic field B_0. Due to the vertical magnetic field B_0, hydrogen atoms in patients' body are aligned uniformly by the strong magnetic field. The strength of B_0 is mostly 1.5 T or 3.0 T, the higher strength the better SNR of images can be acquired in theory, at a higher cost. The recovery of the atoms state is governed by two major types of relaxation time.

The first is the spin-lattice (T1) relaxation time, which is the time from the process of recovery of atoms to vertical magnetization, when the protons emit energy to the environment. On the other hand, the decrease of transverse magnetization is called free induction decay (FID). is affected by the additional (compared to vertical) mechanisms called T2 and $T2^*$ decay. T2 is the 'true' decay time for tissue, while $T2^*$ is the observed value. Due to influence of inhomogeneity in B_0, $T2^*$ is generally shorter than T2 The different body tissues will have different T1, T2 and protons density, thus this physical information will be recorded by the scanner and generates the contrast of MRI image [36]. For example, fat has shorter T1 and T2 than water, and both fat and water have relatively long T1, T2 (compared to $T2^*$), and $T2^*$ decay occurs very quickly [37].

The repetition time (TR) and echo time (TE) can be adjusted to emphasize a particular type of contrast, there are multiple modalities of MRI data, such as T1-weighting and T2-weighting [37]. Table 2.5 explains the differences between the two modalities,

Table 2.5 Difference between T1- and T2-weighting MR images

	T1-weighting	**T2-weighting**
Tissue intensity-Bright	Fat, blood products, slow-flowing blood, radiation change, paramagnetic contrast agents	Fat, proteinaceous tissue, fatty bone marrow, high free water tissue, blood products
Tissue intensity-Intermediate	Proteinaceous tissue	Fat, fatty bone marrow, high bound water tissues
Tissue intensity-Low	High free water tissue, high bound water tissue, collagenous tissue	high bound water tissues, collagenous tissue, bone islands
Tissue intensity-Dark	Air, mineral rich tissue, fast-flowing blood	Air, mineral rich tissue, fast-flowing blood
TR	Short	Long
TE	Short	Long

As shown in Table 2.5, the different MRI modalities (T1-weighting and T2-weighting) provide different contrast for tissues and body structures in images. Some tissues are subtle to view in one modality but significant to view in another modality, thus combination usage of available MRI modalities can help the medical image analysis, such as in brain tumour segmentation [38]. Further fundamentals of MRI are described in [14].

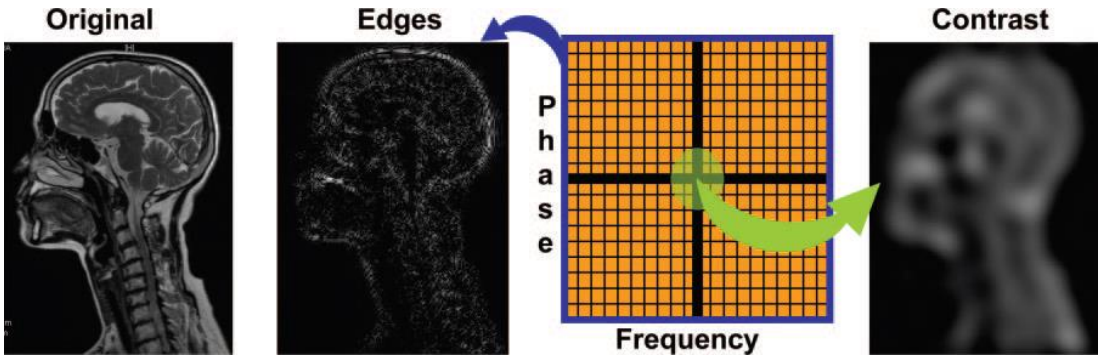


Fig. 2.4 Schematic and corresponding MR images show the characteristics determined by data at the periphery of k-space (ie, spatial resolution, or definition of edges) and those determined by data at the center of k-space (ie, gross form and image contrast) [37].

The encoded frequency and phase information are stored in k-space. K-space is a matrix of voxels whose x-axis usually corresponds to the frequency, and y-axis usually corresponds to the phase. As shown in Fig.2.4, the centre of k-space records contrast information of tissues, and the periphery of k-space contains the details and fine structures of scanned body region. The information in k-space is then decoded through Fourier transformation so that final image is acquired [37].

There are several factors which affect the SNR of obtained MRI image, such as value of B_o, values of the sequence TE and TR, and sensitivity of the receiving RF coil [37]. The spatial resolution is decided by k-space. The CNR (contrast-to-noise ratio) is primarily affected by the difference in relaxation times and the proton density and the values of the TR and TE parameters used in the imaging sequence [32]. MRI contrast agents can be used to increase the CNR between healthy and pathological tissue. The most commonly used clinical paramagnetic contrast agent is gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA). The contrast agents are introduced into patients' body during scanning, and accumulate in certain tissue, T_1 of the tissue will be shorten, thus the contrast will be more significant between tissues.

The MRI technique has multiplane capacity (also available in multiline CT scanners), but it introduces zero ionizing radiation compared to CT imaging. The first disadvantage of MRI scanner is the higher cost. MRI scanner takes longer (compare to Ultrasound Imaging for example) processing time to generate images. Another

problem is that it cannot scan patients who have ferromagnetic material such as shrapnel [14]. The MRI techniques have been used in brain, liver and the reticuloendothelial system, musculoskeletal system, cardiac system, and digestive system [32].

2.3.3 MRI artefacts

MRI artefacts can be from MRI system hardware alone or through the interaction of the patient with the hardware [39]. The hardware related artefacts can arise from RF pulse discrepancies, fluctuating power supply, and inhomogeneity in the magnetic field. The patient related artefacts can arise from patients' moving, breathing of body, blood flow, tissues with different magnetic susceptibilities, and fat/water interface in tissues. Table 2.6 summarizes the main MRI artefacts. Several artefacts will be introduced in this section, which can deteriorate medical image analysis if not be tackled.

Table 2.6 Summary of MRI artifacts

Hardware-related Artifacts	**Patient-related Artifacts**
Zipper artifacts	Motion Artifacts
Shading artifacts	Metal artifacts
Spike artifact	Aliasing artifacts (Wrap around)
External magnetic field (B0) inhomogeneity	Partial volume effect (PVE)
RF (B1) inhomogeneity	Chemical shift artifact
…	…

2.3.3.1 Poor contrast

In both single channel images (grayscale) and multi-channel images (HSV, RGB, etc.), the pixels are numerically quantized with discrete values (scalar or vector) to show their luminance (or colour). The higher luminance stands for higher power, which

should give more significant visual acuity. While in practice, an object's saliency more relies on the relationship between its own luminance and the luminance of adjacent background, instead of its absolute luminance [40]. The human vision system is more sensitive to changes of luminance.

Therefore, the concept of Luminance Contrast is introduced, which describe the degree of how distinguishable regions with different luminance (or colour) in an image are. And statistically a classic formula of Luminance Contrast called Weber Contrast is given as,

$$C_{weber} = \frac{I - I_b}{I_b} \tag{2.1}$$

where I represents the luminance of the features (edge, region, or object), and I_b stands for the background. The Weber Contrast is generally used in the circumstance that features are small then background are large and uniform, such as letter stimuli [41].

Other common formulas Michelson contrast and RMS contrast are

$$C_{Michelson} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{2.2}$$

$$C_{RMS} = = \sqrt{\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(I_{ij} - \bar{I})^2}$$

In the Michelson Contrast, I_{max} and I_{min} represent maximum and minimum luminance. Michelson contrast is preferred for gratings [41]. The RMS contrast is for natural stimuli and efficiency calculations [42, 43]. In RMS contrast, $\bar{I}$ is the mean intensity value in image, and I_{ij} is the intensity of pixel at i^{th} row and j^{th} column. The calculation of RMS contrast is same as the standard deviations of pixels in image.

In this thesis, when the term 'Contrast' is mentioned, visually it refers to the level of distinguishing different regions, and statistically the standard deviations of pixels (narrowly or widely distributed). Thus, here several examples are given to show the good contrast and poor contrast in visually and statistically way,

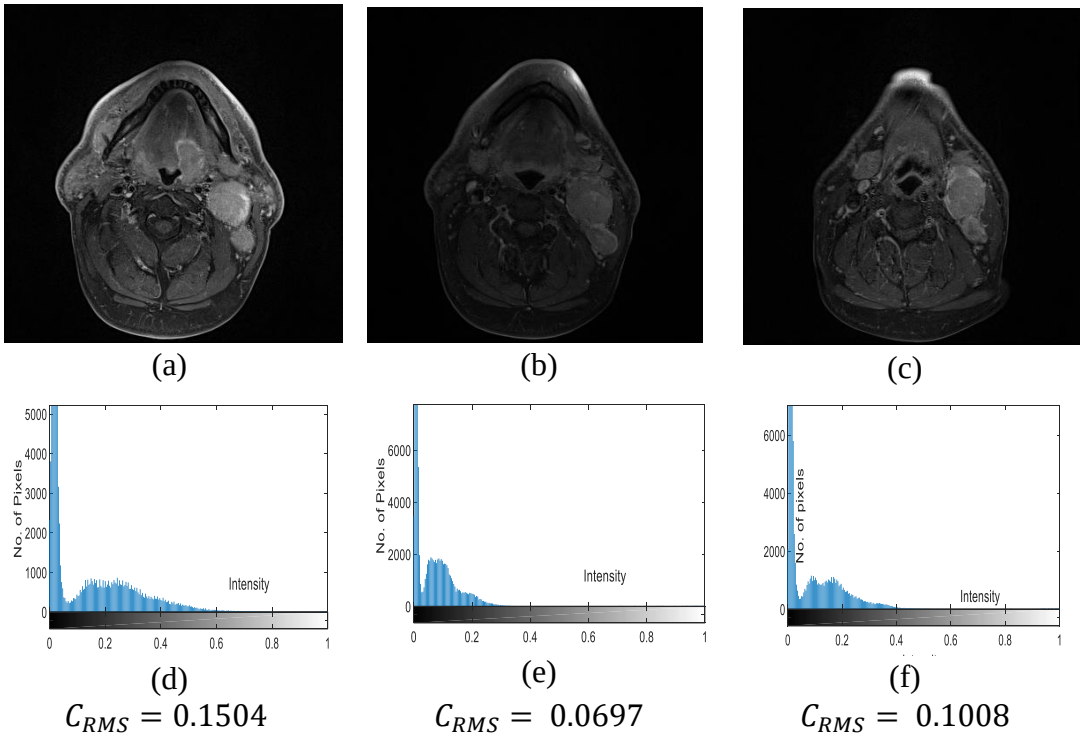


Fig. 2.5 (a)(b)(c): three MRI slices with different contrast. (d)(e)(f), their histogram of pixel intensities. Last row, their RMS contrast values.

As shown in Fig. 2.5 (b), the tissues in MRI data will be harder to distinguish when the slice has lower C_{RMS}. And in this case, the histogram of this slice will be more crowded distributed. This is the artefacts called poor contrast, and a slice with relatively better contrast should be as shown Fig. 2.5 (a)(c). The poor contrast will cause problem for clinicians to find target tissues, and it will also misleading computer-based algorithms to generate more errors.

2.3.3.2 Noise

The main noise in MRI data is generated from motion artefacts and the MR imaging system [44]. The motion artefacts can lead to diffuse image noise. The motion can be caused by patients' breathing and other non-periodic movement (an example Fig. 2.). The MR system can cause noise in many ways, such as eddy currents produced from changing magnetic fields, stray RF signals, bad electronic, loose connections to surface coils, or any RF pollution from Faraday cage etc.

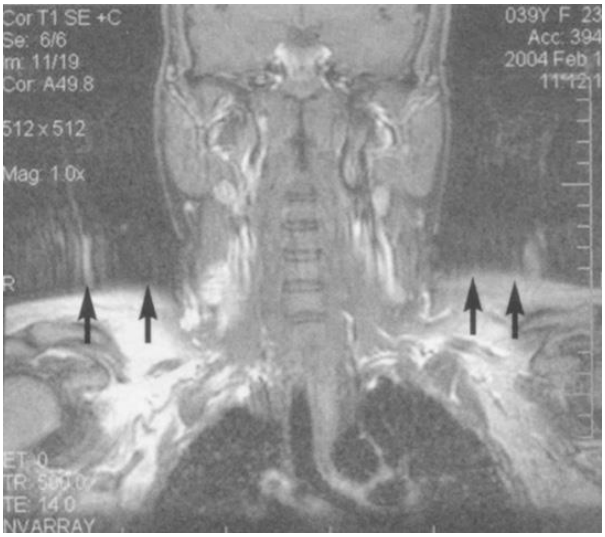
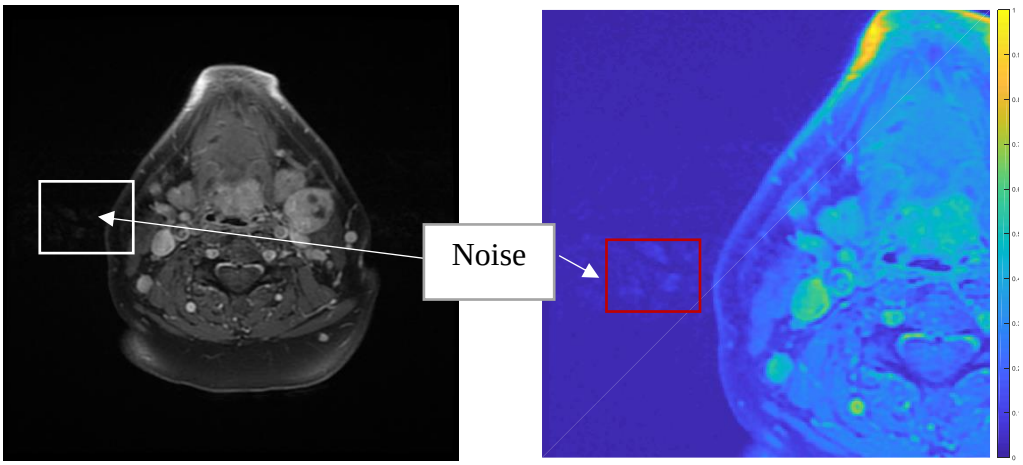


Fig. 2.6 Motion artefacts (marked with black arrows) in T_1 coronal study of lumbar vertebrae [44].

Compared to the errors contributed from the system, the major noise is thermal noise caused from thermal motion of patients' body as shown in Fig. 2.6.



(a) (b)

Fig. 2.7 An MRI slice with noise (marked with arrow and text box). (a), a slice, (b), zoomed region of left slice, and converted to parula colour space to make it more visible.

Fig. 2.7 shows an example of noise in MRI data. The arrow-pointed regions should be dark visually and has zero values because they are background and nothing is scanned, while there are non-zero values in these regions, which is noise may be caused by thermal motion. Thermal noise can be represented in additive way as,

$$I(x,y) = O(x,y) + \eta(x,y) \tag{2.3}$$

where I represents an obtained and reconstructed raw MRI image, O stands for the original signal, and N stands for the noise parts in I. The methods of reducing noise will be introduced in next chapter.

2.3.3.3 Partial volume effect

The partial volume effect (PVE) is caused by limitation of spatial resolutions of medical imaging modalities [44, 45]. The definition of PVE is that, in medical image one voxel contains multiple tissue types, so that the voxel will not be correctly reconstructed.

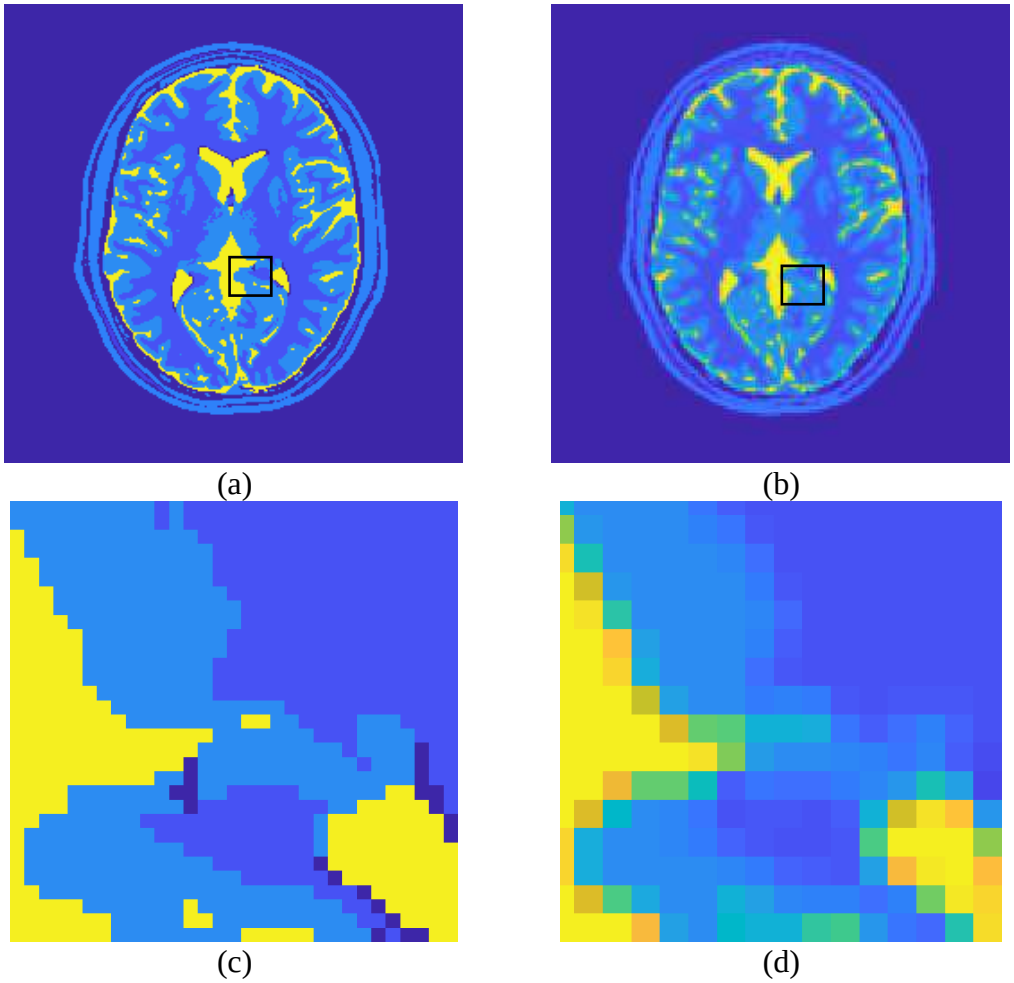
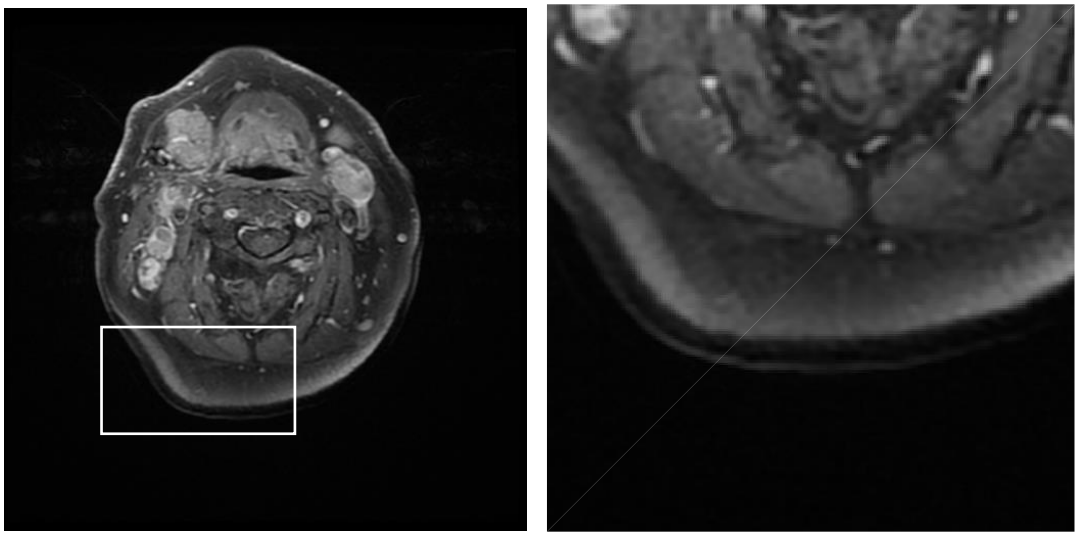


Fig 2.8 Partial volume effect in a T_1 parameter map. (a) a brain simulation with standard T_1 spatial resolution. (c) the zoomed region in black box of (a) in top. (b) image same as (a) but with lower resolution. (d), zoomed region of lower resolution [45].

Fig. 2.8 shows an example of PVE. The left column is a stimulated brain T_1 parameter map. And right column has four times bigger voxel size, where the partial volume voxels cannot be correctly reconstructed. If the PVE existed, quantitative analysis of target boundaries or volume would be severely affected. As shown in Fig. 2.8 (d), the boundaries are blurred, and the intensity difference between different tissues are decreased, which will cause errors for classification.

2.3.3.4 Gibbs ringing artefact

Gibbs ringing artefacts refers to the phenomenon that, multiple fine parallel lines occur adjacent to high-contrast interfaces. This artefact arises from the MRI reconstruction process, and the ringing is generated by inverse Fourier Transform (FT). Due to limitation of Signal to Noise Ratio (SNR) and scan time, the outer parts of the k-space that contain the high frequency information of the image, are generally not recorded [46].



(a) (b)

Fig. 2.9 The Gibb's ring effect. (a) an MRI slice. (b) zoomed region of left slice (a) (marked with box in left image).

As shown in Fig. 2.9 (b), the edges of foreground, i.e., high-frequency information, are repeated as parallel lines (ringing) nearby.

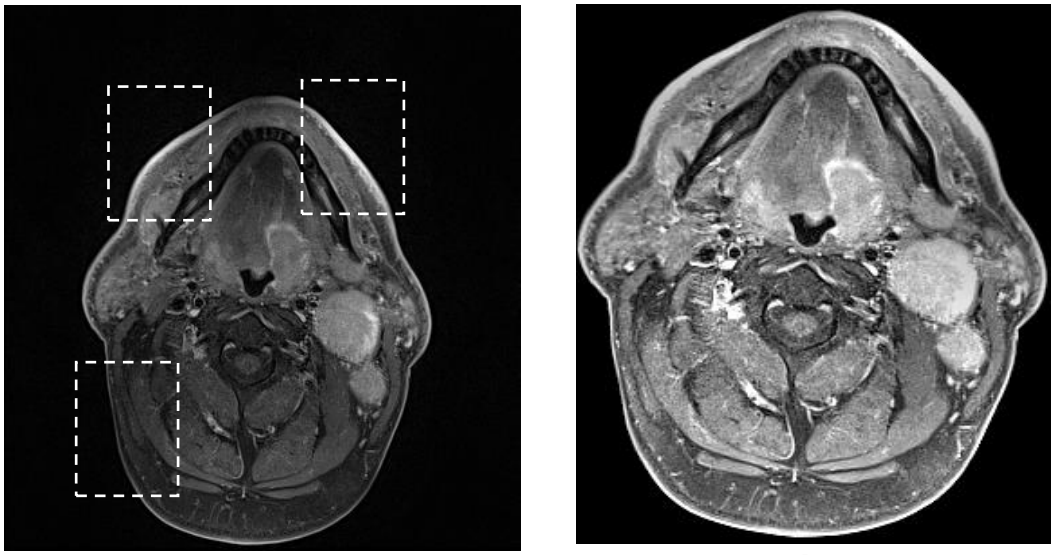
2.3.3.4 Intensity variations inside single slice: Bias field

Another artefact called intensity homogeneity (IIH, or bias field) can badly deteriorate the computer-aided diagnosis. The bias field can possibly arise from non-uniform RF-

field, irregular main magnetic field, susceptibility effects of normal tissue and receiver coil sensitivity profile [47]. The bias field causes the intensity distortions [44]. Normally intensity of a tissue class is assumed constant over the range of the MRI image, the distortion caused by inhomogeneity introduces a shading artefact, so that intensities of tissues are mis-mapped [48]. Previously, the addition of noise on original signal is shown in Eq. 2.3, the influence of intensity inhomogeneity can be modelled in multiplication with original signal, the so the Eq. 2.3 can be updated as,

$$I(x,y) = O(x,y) * B(x,y) + N(x,y) \tag{2.4}$$

where the O(x,y) is original signal, B(x,y) is bias field existed at (x,y), and N(x,y) is noise, I(x,y) represents the actually obtained and reconstructed MR image. An example is shown in Fig. 2.14.



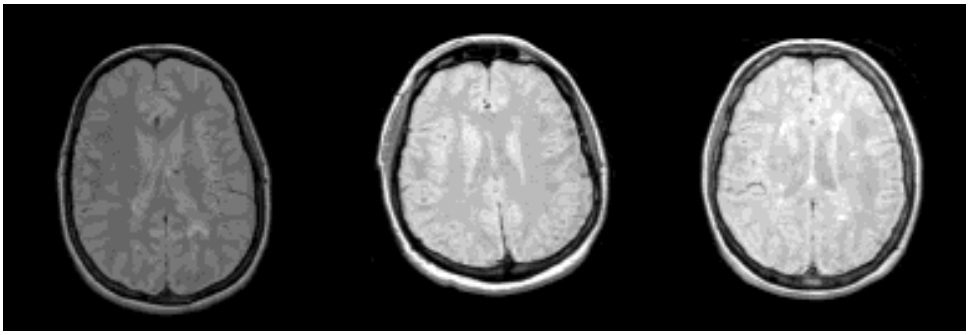
(a) (b)

Fig. 2.10 An example of intensity inhomogeneity. Left, slice with inhomogeneity (some shown in white dash boxes), Right, slice after bias field correction.

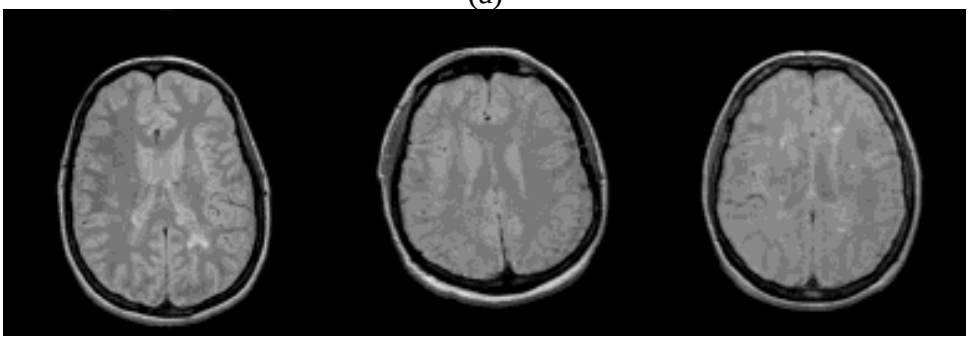
The example in Fig. 2.10 (a) shows that slice with bias field will have non-uniform appearance (shown in dash boxes) on similar type of tissues, which needs to be corrected first before applying intensity-based analysis. Fig.2.10 (b) shows the bias field corrected version of Fig.2.10 (a), where the intensity distributions of same types of tissues are more uniform. The methods of correction will be described Chapter 3.

2.3.3.5 Inter Intensity variations among MRI images

Apart from the intensity inhomogeneity (bias field) in single MRI image, there are also intensity variations between different MRI slices. These refers to the artefacts that the tissues should have similar intensity range in different MRI slices but in practical distribution of intensity range of same tissue type varies across different MRI slice. There are variations of intensity profile for same tissue type in inter-scanner, intra-scanner, inter-patient, intra-patient circumstances [49]. The possible reasons of the intensity variations include different scanner technologies, different acquisition settings, different illuminations, and different post-processing algorithms [50, 51]. If the intensity variations are significant, it will be hard to set good parameters for intensity-based diagnosis algorithm, and this will be even worse for 3D work. An example of the existence of inter intensity variations is shown as Fig. 2.11,



(a)



(b)

Fig. 2.11 Original slices from three studies acquired as per the same FSE Pd protocol before standardization displayed at a fixed window that was actually set up for the first image (first row), and after standardization displayed at a fixed 'standard' window (second row) [52].

As shown in Fig. 2.11 (a), before the standardization of intensity, three slices from the same scanning protocol have different intensity ranges from the same type of tissues

(e.g., cortex). And after the standardization (shown in Fig. 2.11 (b)), the intensity ranges are uniform for same tissues across three slices, which is helpful for potential intensity-based analysis.

2.4 Radiation therapy

2.4.1 Radiation treatment planning (RTP)

Apart from the diagnosis, the medical imaging has also been widely applied in radiation treatment planning (RTP) of cancer. Radiation therapy (RT) has been used for treatment of HNC for decades [17], and it is arguably the most challenging and sophisticated radiation treatment among all RT delivered to any anatomical subsite [26]. The advances in computer technologies have significantly improved the ability of accurate RTP, which maximize tumour dose and minimizing the dose to adjacent normal tissues and structures (also mentioned as organ at risk, OAR). Nowadays, the intensity-modulated radiation therapy (IMRT, an example is shown in Fig. 2.17) is regarded as the standard RT technique for HNC. In IMRT, multi leaf collimator (MLC) is used for sculpting the radiation beam to conforms tightly to the shape of target. And outside the target, a steep dose gradient is created to protect the adjacent normal tissues [17]. The IMRT outperforms conventional radiotherapy techniques, improves the dose distribution, and introduces fewer side effect. On the other hand, IMRT does require slightly longer daily treatment times and additional planning, quality assurance, and safety checks before the patient can start the treatment [17].

2.4.2 Definition of GTV, CTV, PTV in radiotherapy treatment planning

RTP aims to maximize the dose to target and minimize dose to adjacent healthy tissues, thus a good design of RT relies on specialists with expertise in the management of HNC [53], because they have abilities to delineate highly acceptable targets for RTP.

To design an appropriate RTP, the first step is to form a three-dimensional map of the patient's anatomy in the exact position for everyday treatment. The map normally using CT scans. Then, on the map, the radiation oncologist defines the targets and critical normal structures.

In the definition process, the first step is to define gross tumour volume (GTV) at both the primary site and any regional lymph nodes. The GTV is defined by drawing the

primary tumour itself based on single or multiple imaging studies registered to the CT map, such as MRI, and PET [26].

The definition of targets and critical normal structures is implemented by delineation of GTV, CTV, and PTV, as illustrated in Fig.2.12,

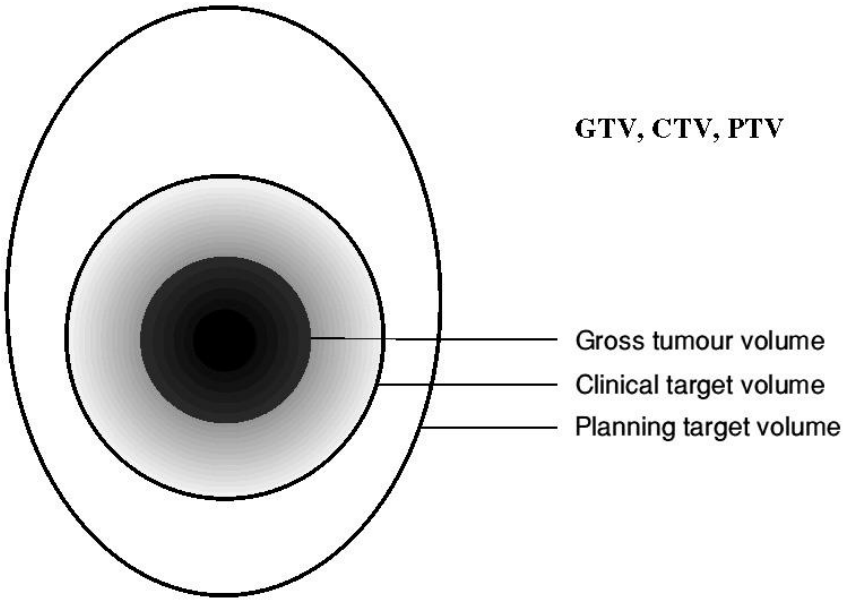


Fig. 2.12 Diagram to illustrate the main radiotherapy planning volumes [54].

In the definition process, the first step is to define gross tumour volume (GTV) at both the primary site and any regional lymph nodes. The GTV is defined by drawing the primary tumour itself based on single or multiple imaging studies registered to the CT simulation, such as MRI, and PET [26].

Then, the clinical target volume (CTV) is defined as GTV plus margin for subclinical microscopic spread and affected lymph nodes. Finally, planning target volume (PTV) is defined as an isotropic margin around the CTVs to account for small changes in patient positioning each day [7, 26]. This margin is generally 3-5 mm, and the PTV will be used for ultimate radiation dose prescribing. The amount (dose) of radiation applied is measured by unit called Gray (Gy), and an example of definition of RTP is shown in Fig 2.13.

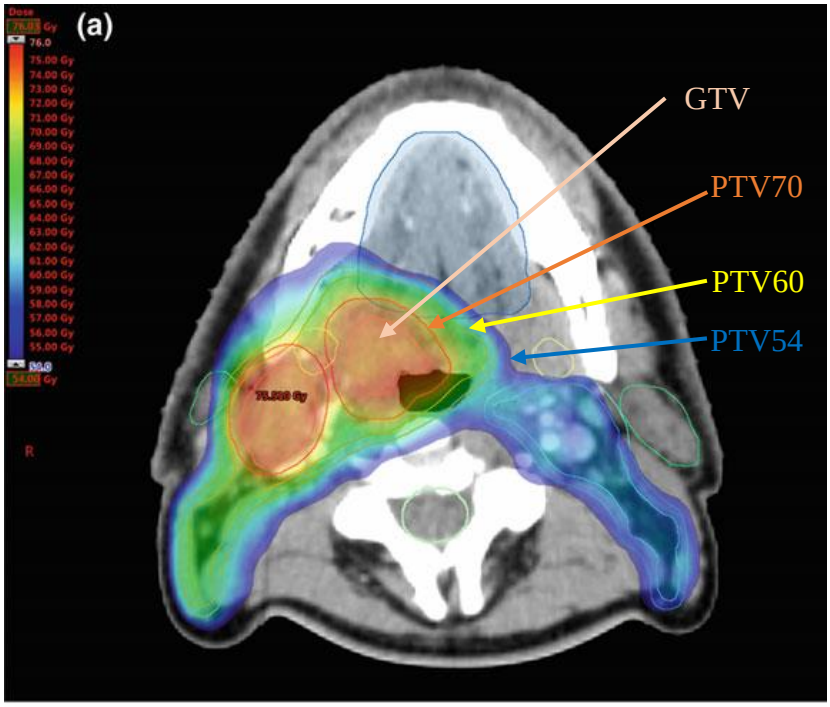


Fig. 2.13 An example of IMRT. The GTV is shown in pink. The PTV70 (70 Gy) is marked with red outline. Inner and outer orange lines represent CTV60 and PTV60, respectively, and the inner and outer cyan lines represent the CTV54 and PTV54, respectively. All coloured regions will be delivered radiation, and the dose is represented by the colour bar on left top side [26].

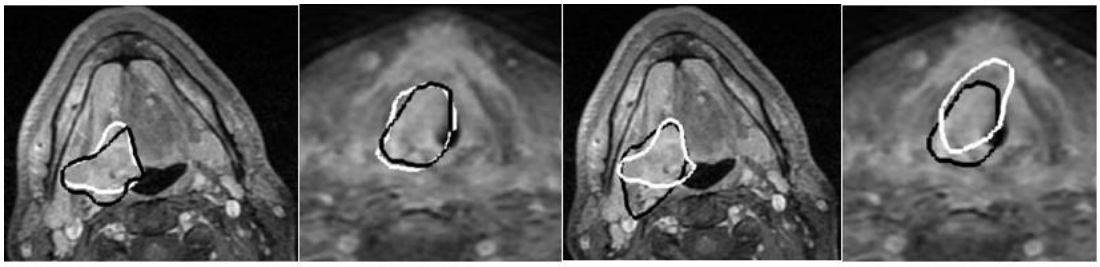
2.4.3 Inter- and intra- variabilities

The high quality RTP relies on the accurate definition of PTV. The definition of PTV is a still a challenging and time-consuming problem. In a typical RTP process, firstly the definition of GTV and CTV relies on available facilities and radiation oncologists' expertise. Then the CTVs are given to dosimetrist or physicist specifications for target covering and normal tissue constrains. Following dosimetrist or physicist determines the optimal profile such as orientation, shape, and intensity of each radiation beam to match the PTV. The definition and delivery of RTP can take weeks and numerous iterations, including the adjustments and confirmation between clinicians and planners to decide the final plan [26].

Furthermore, the manual delineations (by radiation oncologists) of tumour primary site and adjacent healthy can take from 45 minutes to 3 hours per patient [7]. Also,

based on the aim of better RT, the adaptive radiation therapy (ART) is introduced. The Adaptive radiation therapy is a closed-loop radiation treatment process where the treatment plan can be modified using a systematic feedback of measurements [8]. To improve RT, in ART clinicians monitor the variations of patients during the treatment and incorporate this information to re-optimal the RTP. Thus, ART requires radiation oncologists define GTV and CTV at each time of the re-optimal RTP process, which involves much more time cost.

Apart from the time consumption problem, other challenges of RTP are inter- and intra- variabilities during the delineation of GTV and CTV. The presence of tumour often changes the adjacent anatomy, the shape of HNC tumour could be arbitrary, the appearance of tumour is generally complex, and the uncertainty in mucosal spread is existed, all these factors make the delineation task more sophisticated. The challenges of delineations lead to inter- and intra- variabilities in manual contouring, the inter-variabilities mean that given a same scan to two clinicians, the drawn GTVs can be different; the intra- variabilities mean that given a same scan to one same clinician before and after a duration, the opinion about the GTV can also be different.



(a) (b)

Fig. 2.14 (a) Delineation by two clinicians and the intra- variability (b)Delineation by one clinician before and after a period of time and the inter-variabilities (white and black contours) in GTV of base of tongue (BoT) and larynx tumour from MRI (provided by Beatson Cancer Centre) [55].

In Fig. 2.14, the intra- and inter-variabilities in manual delineations of GTV are displayed. The manual contours are drawn by two clinical oncologists (COs) from Beatson West Scotland Cancer Centre, Glasgow; and the delineation process is following the guidelines in [56, 57]. As claimed in [6], there were averagely 13% intra-

observer error of 13% for a pharyngeal cancer. From [5], there were 12.9% inter-variation for an oropharyngeal tumours and 13.8% for a laryngeal tumours respectively in manual contouring of GTV from MRI. In recent trials [58], there were mean 6% inter-variation for delineations of GTV of OAR at oral cavity, mean 27% at glottic larynx, and mean 13.7% at supraglottic larynx. From the current clinical practice (in Beatson West Scotland Centre, Glasgow), the gold standard delineations are determined by weekly meeting of clinical oncologists, radiologists, surgeons, and radiation oncologists, where consensus manual outline is decided and considered as gold standard for RTP. In this thesis, when the computer algorithms-based auto-segmentation results are measured by ground truth, it also refers to the consensus gold standards.

2.4.4 Auto-contouring software

As manual segmentation of tumours is challenging due to both time cost and subjective variations, computer vision techniques were introduced into medical imaging processing to develop auto-contouring algorithms and software, so that improve the efficiency of tumour volume contouring. The auto-contouring algorithms have advantages in many aspects, including time saving, objective results, acceptable accuracy, and abilities of visualisation as well as quantification of 3D structure.

In Table 2.7, several image processing algorithms for auto-contouring of HNC from medical data are summarized. The algorithms for auto-contouring include classical approaches such as deformable model [59], machine learning method such as SVM [60], and also deep learning methods such as convolutional neural network [61-63]. The computer-based auto-segmentation algorithm can achieve Dice similarity score up to 76%. Methods in Table 2.7 have not provided their time cost. In [64], ten studies of auto-segmentation methods of OAR in head and neck were reviewed, six of them saved time compared to manual contouring, and the largest time saving was 59%; three of them were almost the same as manual; and the last one took 15.7 % longer than manual delineation. In [58], the automatic delineation was 33% shorter than manual contouring (34 minutes).

Table 2.7 Auto-segmentation algorithm of HNC from medical data

Auto-segmentation Algorithm	**Data Used**	**Accuracy**
Convolutional neural network [65]	7channel multiparametric MRI, 72 patients with HNC	0.65 Dice Similarity Score
Support Vector Machine [60]	120 Dynamic Contrast Enhanced MRI datasets with HNC	0.76 Area Overlap Measure
Convolutional neural network and deformable model [66]	254 patients of CT PET scans with HNC	0.75 Dice similarity score
Deformable model (Level set method) [59]	CT scans of 23 patients with HNC	0.78 Intersection Ratio
Deep learning models [61-63]	HECKTOR (HEad and eeCK TumOR) Challenge, 254 PET/CT data from patients with HNC	Best among 18 teams has 0.76 Dice similarity score

In Table 2.8, some auto-contouring software are shown. Some of them [67] [68] [69] are commercially used. From the experiments [70], these commercial software tools have saved time cost around 30% to 40%. There are also some open-sourced free software [71] [72] [73] [74] for auto-contouring, review in [75] shows that they can

save time compared to manual segmentation. The auto-contouring software have functions in 3D segmentation, 3D visualisation, registration, and other image processing methods.

Table 2.8 Software for auto-contouring of tumours from medical images

Software	**Platform**	**Expense**	**Methods**
SPICE9.8 [67]	Philips	Commercial	Atlas-based model-based
Smart Segmentation [68]	Varian	Commercial	Atlas-based
ABAS2.01 [69]	Elekta	Commercial	Atlas-based model-based
3D slicer [71]	Open-sourced	Free	Model-based
ImageJ [72]	Open-sourced	Free	Model-based Machine learning
MIPAV [76]	Open-sourced	Free	Model-based
ITK-SNAP [73]	Open-sourced	Free	Model-based
DeepImageJ [74]	Open-sourced	Free	Deep learning

As shown in Table 2.8, most segmentation methods used in these software are based on interactive segmentation such as level set method which requires manual initialization, some software also use atlas-based methos to involve abundant pre-marked atlas images and apply registration between the atlas and input images, [74]

introduces deep learning into auto-segmentation software . In current clinical practice, the auto-contouring results need to be reviewed and further modified by clinical experts before used in RTP.

2.5 Conclusion

In this chapter, the background of head and neck cancer and its treatment was reviewed. The medical imaging techniques specially MRI were described and the procedure and particular challenges of RTP were explained.

It can be concluded that, advances of medical imaging techniques enabled high quality diagnosis and treatment of cancers. However, challenges such as subjectiveness and high time consumption still exist in current RTP. These challenges as well as the development of computer vision techniques motivate the research and development of computer-based auto-segmentation algorithms and software, which aims to minimize the inter-observer variations and reduce clinician's workload, so that improve the treatment of cancers and life qualities of patients. In the next chapter, algorithms for auto-segmentation of head and neck cancer will be reviewed, which will help explain the applications of image processing techniques for improvement of HNC contouring.

Chapter 3 Image Processing Solutions for HNC based on MRI Data

3.1 Introduction

Developing medical imaging technologies not only provide powerful tools to medical experts, but also enabled research of computer algorithm aided medical image processing. Computer-aided diagnosis (CAD) evolved from imaging techniques and computer vision algorithms, that aimed at modifying accuracy and consistency of medical image understanding, and provides radiologist with computer driven results as a 'second opinion', which helps radiologists make overall decision [77, 78].

A broad range of technologies and research topics have applications in medical imaging, such as picture archiving and communications system (PACS), low-level image processing, and high-level interpretation of medical images [79]. The applications of medical imaging informatics include but not limited to cloud-based medical image archiving [80], medical images retrieval [81-83], lung nodules detection and classification [84-88], brain glioma segmentation [89-93], detection of interval change of bone [94, 95]. Different types of imaging data are used in different scenes. Photographs of retina are used for retina vessel segmentation to diagnose disease such as diabetes and hypertension [96, 97]. Optical coherence tomography (OCT) images are also used in classification of retinal diseases, typically in macular edema [98-100]. CT data has wide applications in CAD, such as lung nodules detection [101-103], liver tumour segmentation and classification [104-106], and segmentation of organ at risk in HNC region [107, 108]. Also, CT images play important roles in COVID-19 diagnosis and management [109-111].As introduced in previous chapter, MRI has advantages in application of most brain and spline disorders [13], thus enables lots of research in brain tumour classifications and segmentations [89, 91, 112, 113]. MRI data is also utilised in ventricular tracking, classification and volume segmentation [114-118], and pelvic tissue classification [119, 120].

This chapter will focus on reviewing image processing techniques applied on medical images especially on the segmentation of HNC from MRI data. The general

segmentation workflows for medical images can be categorized as three types, which are interactive segmentation, machine learning, and deep learning methods. In Fig. 3.1 these three types of workflows are shown.

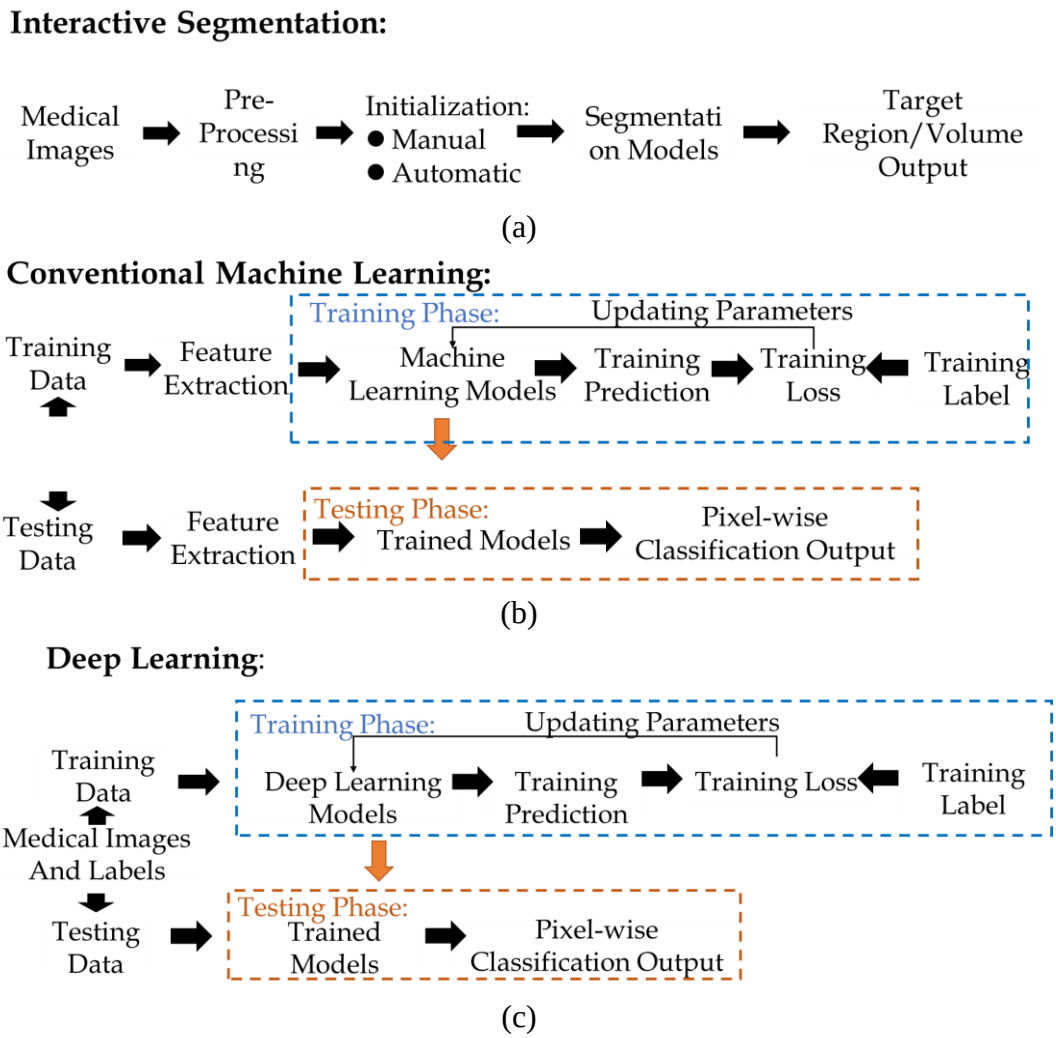


Fig. 3.1 Graphs of medical image segmentation workflow. Workflow for (a) Interactive segmentation, (b) Conventional machine learning segmentation, (c) Deep learning segmentation.

The general process of interactive segmentation methods is shown in Fig. 3.1 (a). With medical images as input, the methods firstly improve image quality and minimize artefacts through pre-processing techniques. The 'interactive' segmentation means that the algorithms require manual interaction to initial the position of foreground of target (tumours/lesions/organs), this operation can be replaced by automatic detection algorithms. Following, the segmentation models will evolve start from manual or

automatic initialisations and finally converge at boundary of targe objects. The machine learning (Fig. 3.1 (b)) and deep learning (Fig. 3.1 (c)) methods combines detection and segmentation into pixel-wise classification of clinical targets in medical images. Both machine learning and deep learning relies on labels drawn by experts. The labels are used in testing phase to train the models, a trained model will then be used in testing phase to evaluate the performance of model.

The remainder of this chapter will review classic methodologies for medical image processing, many of which are also used in the contribution of this thesis. Also, classical and state-of-the-art algorithm for HNC tumour segmentation will be reviewed. The review will be organised and categorized by interactive segmentation (section 3.2), conventional machine learning segmentation (section 3.3), and deep learning segmentation methods (section 3.4) for medical images.

3.2 Interactive methods for medical image segmentation

This section will introduce interactive segmentation methods including the general steps and their applications on medical image segmentation. The common workflow of interactive segmentation is shown in Fig. 3.1 (a). In section 3.2.1, pre-processing techniques for enhancing medical images and reducing artefacts will be introduced. In section 3.2.2, manual and automatic localization of target's foreground will be reviewed. In 3.2.3, three widely used interactive segmentation methods will be introduced, they are Region Growing [121], GrabCut [122], and Deformable Models [123]. In 3.2.4, the level set method will be reviewed with more details because this technique will link to the work in Chapter 5 and Chapter 6.

3.2.1 Pre-processing of medical image

Due to the artefacts occurred during medical imaging signal acquisition and reconstruction, there are noise and low-quality problems in some medical data, thus image processing task including de-noising and contrast enhancement can help improve data quality to make it acceptable for further usage. This section will cover the solution for several artefacts linked to segmentation work in Chapter 5 and 6, a summary figure is given as following Fig. 3.2.

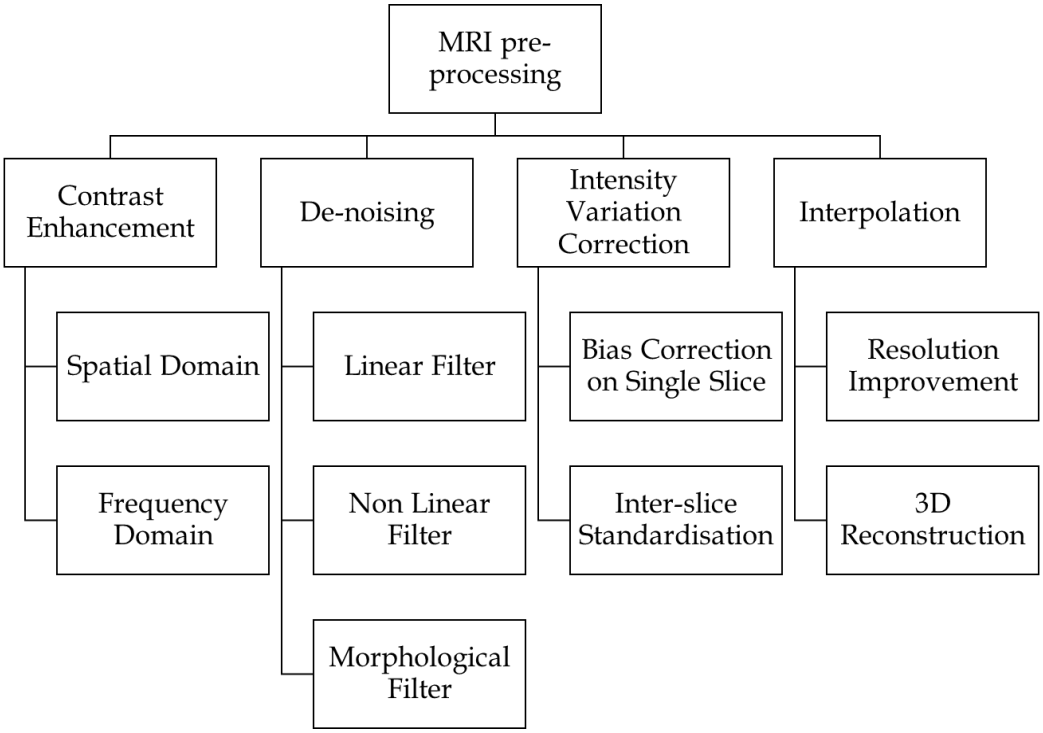


Fig. 3.2 Pre-processing techniques for medical images

As shown in Fig. 3.2, four types of pre-processing will be described namely: 1) contrast enhancement through spatial domain and frequency domain; 2) de-nosing using linear filter, nonlinear filter, and morphological filter; 3) Correct the intensity variation including the variations on single slice and across multiple slices; 4) Interpolation of medical images to improve their spatial resolutions.

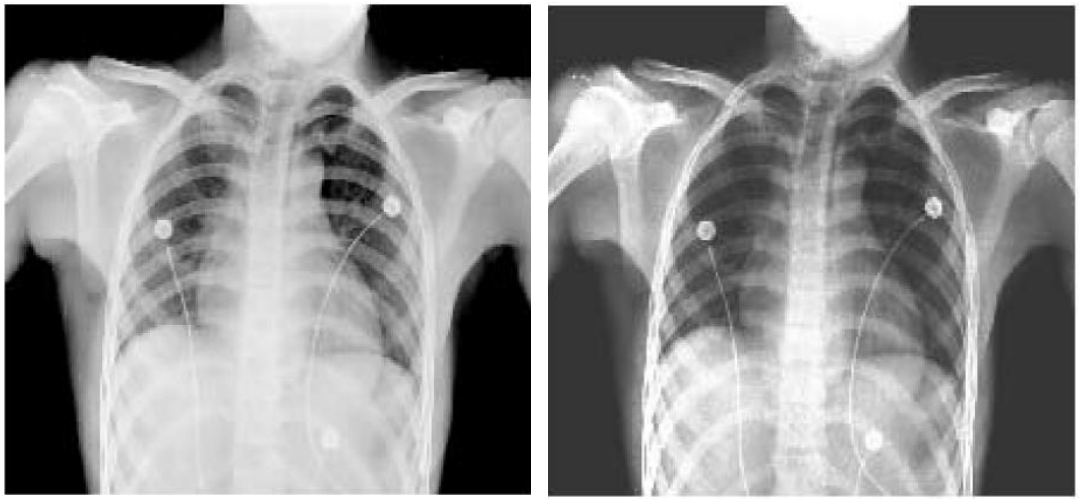
3.2.1.1 Contrast Enhancement

As the majority of medical data are grayscale images, thus the intensity values are the most essential factor for medical image analysis. While poor contrast medical images will cause difficulties for clinicians to distinguish different tissues and recognise important targets in images.

In medical image processing, the CE techniques improve the quality or clarity of images to increase the interpretability in images. The most common contrast enhancement method is histogram equalization [124]. Call the cumulative density function (CDF) of histogram of input image I_{in} as $c_{h0}(I_{in})$, CDF of output (contrast enhanced) image y as $c_{h1}(I_{out})$. The HE aims find a spatial transformation $T_c(I_{in})$ to map input image I_{in} to output image I_{out}, so that minimize the $|c_{h1}(T_c(I_{in})) -$

$c_{h0}(I_{in})|$. And there are constrains to form of T_c as introduced in [52]. The obtained transformation T_c is then applied on input image to enhance the contrast. It should be noted that HE changes the brightness of an image, limits its direct application in consumer electronics [125].

There are advanced CE techniques proposed after HE. These techniques can be categorised as global and local methods [126]. Global techniques take only global histogram information over the whole image; thus, it is fast and simple. Global methods limit the contrast ratio in some parts of image such as background and other small regions. Examples of Global techniques include Brightness Preserving Bi-Histogram Equalization (BBHE) [125] and Dualistic Subimage Equalization (DSIHE) [127]. Local methods use small window that slide through each pixel in image and enhance the pixels only inside the window each time. Local methods make good use of local information, but it can lead to over-enhancement of some parts of the image and enhance noise effect in image. Also, local methods have relatively high computational costs compared to global methods. Examples of local methods include Adaptive Histogram Equalization (AHE) [128] and Contrast Limited Adaptive Histogram Equalization (CLAHE) [129]. Recently many research work on reducing over-enhancing problems in CE, improving image sharpening and peak signal to noise ratio (PSNR). These research introduce gamma correction [130], homographic filters [131], genetic algorithms [132] into classical CE methods. Additional, in [133] the generative adversarial network (GAN) is proposed as a new tool for medical image enhancement. Majority of CE methods demonstrated their applications on medical images across CT, MRI, and colour images such as fundus images. An example of CE on medical images are shown in Fig. 3.3.



(a) (b)

Fig. 3.3 Examples of contrast enhancement on medical images. (a) A chest X-ray-Vandy image. (b) After contrast enhance [131].

After the CE, different components in medical images are more distinguishable. It can be seen in Fig. 3.3 (a) that before CE intensities of two sides of lungs are whiten, make them hard to separate from white backgrounds such as bones. In Fig. 3.3 (b) the chest image after CE provides better contrast between lungs and backgrounds so that distinguish them better.

3.2.1.2 De-noising for medical images

Apart from poor contrast, another general challenge in medical images is noise. The noise in medical images is reduced by image filtering techniques. The filters can be designed and applied in the spatial or frequency domain. Image filtering techniques can be classified into two categories: linear and non-linear. Linear filters are those filters whose output values are linear combinations of the pixels in the original image [134], such as an Average Filter [135]. The non-linear are those filters whose output is not a linear function of its input. Some non-linear filters work directly on spatial domain, such as Median Filter [136], Non-local Mean Filter[137] and Anisotropic Diffusion Filter [138]. There are also non-linear filters work on transformed domain, such as Wavelet Image De-noising Filter [139]. Block-matching and 3D filtering (BM3D) combines non-local filter on spatial domain and 3D Wiener filtering on transformed (wavelet) domain, and achieved big success in de-nosing tasks [140], although it takes more complex computation.

The major noise in MRI image is from thermal motion (body motion, cardiac pulsation, or respiratory motion.), and most commonly under *Rician* distribution. As discussed in [134], the application of different image filters on MRI image has pros and cons. Linear filters including Average Filter and are generally poor in maintaining the edges after smoothing the image. Median Filters can effectively remove salt and pepper noise and preserving the edges, while it produces blur images. The Anisotropic Diffusion Filter outperformed isotropic filters (such as Gaussian Filter) in preserving edges and lines in image. The Non-local Mean filter can work well for *Rician* noise removal and preserve edges well, while it requires expensive computation. The Wavelet De-Noising Filter performs better to additive noise but sometimes it introduces characteristic artifacts. Nowadays, many methods also use deep learning for de-nosing of MRI [141, 142], although it takes much more computation power. An example of noise removal in medical images is shown as Fig. 3.4.

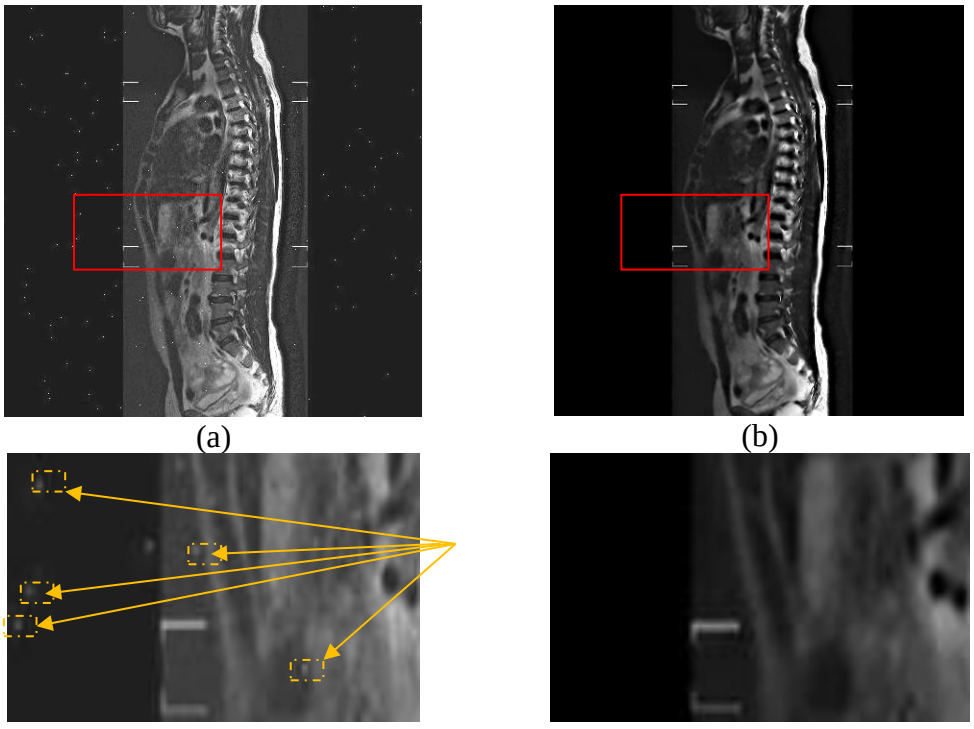


Fig. 3.4 (a) Noisy MRI image. (b) After noisy removal [143]. (c) Cropped noisy image (red region in (a)). (d) Cropped de-noised image after (red region in (b)). The noisy points are marked with dashed yellow boxes in (c).

Fig. 3.4 shows the removal of salt and pepper noise in MRI image of human spline using the proposed median filter approaches in [143]. As seen in Fig. 3.4 (a)(c), this MRI image has salt and pepper noise marked with yellow arrows and dashed boxes. After the de-noising, there are no significant pepper noise in Fig. 3.4 (b)(d). This will help further intensity-based processing such as thresholding and segmentation.

3.2.1.3 Morphological filtering for medical image processing

Mathematical morphological operations can be used in several low-level image processing tasks such as background noise removal, object separation, and remove imperfections introduced during image segmentation. The basics of mathematical morphology is set theory [144]. Given two or more sets, there are several well-known elementary operations among sets such as subset, union, intersection, and complement.

Dilation and Erosion are basic morphological processing operations, which are defined by multiple elementary set operations. The dilation is given as $I_{out} = I \oplus \text{se}$, where I represents image, and s is a structuring element. Dilation produces output image $I_{out}(x, y) = 1$ if at (x, y) the origin of s hits I (i.e., the centre of s intersects I). Dilation can be used for repairing breaks inside target [144]. The erosion is given as $I_{out} = I \ominus \text{se}$, erosion produces output image $I_{out}(x, y) = 1$ if at (x, y) the origin of s fits I (i.e., se is subset of I). Erosion can be used for splitting apart joined objects and can strip away extrusions [144]. Dilation and erosion are dual operations in that they have opposite effects.

Combination use of dilation and erosion are termed as Opening and Closing. The opening is given as $I_{out} = I \circ \text{se} = (I \ominus \text{se}) \oplus \text{se}$, which is an erosion on I followed by a dilation using same s. On the contrary, the closing is given as $I \cdot \text{se} = (I \oplus \text{se}_{\text{rot}}) \ominus \text{se}_{\text{rot}}$, where se_{rot} is se rotated by 180 degree, while it is same as se if se is symmetrical with respect to rotation [145]. The opening operation smooths the contour object, breaks narrow isthmuses, and eliminates thin protrusions. Closing operation generally fuses narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour.

There are more applications of morphological operations. The hit-or-miss transformation is a basic tool for shape detection. Convex hull combines segments

together. The morphological filter can also be used in edge detection of MRI image [146], image fusion of MRI image [147], contrast enhancement of X-ray image [148]. More mathematical morphological operations' applications on image processing can be found in [131].

(a) (b)

Fig. 3.5 Morphological filtering for X-ray hand image enhancement. (a) Original image. (b) After enhancement.

Fig. 3.5 gives an example of medical image enhancement by using morphological operation. The input image (Fig. 3.5 (a)) is a hand image taken by X-ray, this image is processed by an edge-detect morphological filter method from [148]. The enhanced image (Fig. 3.5 (b)) has better contrast between bones and background, as well as sharpened edge of bones.

3.2.1.4 Bias field correction for medical images

The bias field signal is a low-frequency, smooth, and undesirable signal that corrupts MRI images. The bias field is produced by the inhomogeneities in the magnetic fields of the MRI machine [149]. This artefact can be characterized as a smooth variation of intensities across the image [150]. An example is shown in Fig. 3.6.

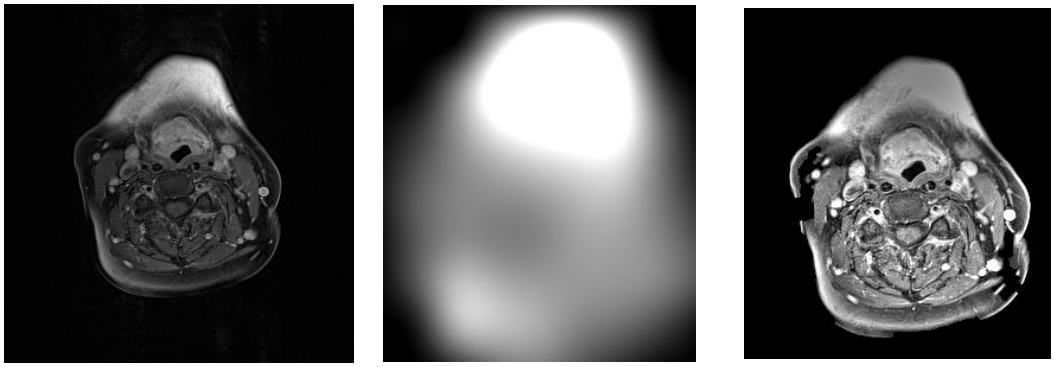


Fig. 3.6 An example of bias field artefacts. (a) a HN MRI slice corrupted with bias field. (b) Estimated bias field in (a). (c) Uncorrupted (corrected) slice.

The MRI slice in Fig. 3.6 (a) is corrupted by bias field. This bias field is visualized in Fig. .3.6 (b) where the top part of slice exists a background signal with higher intensity than other parts of the slice, this varying background intensity signal is so called bias field. Fig. 3.6 (c) shows the slice whose bias field has been corrected. This low-level intensity variation will not badly influent the clinical diagnosis, while it degrades the image processing algorithm including classification, segmentation, and any other intensity information-based image processing algorithms.

Many research work on minimizing the adverse influence of bias field in medical image processing, a summary is shown in Fig. 3.7.

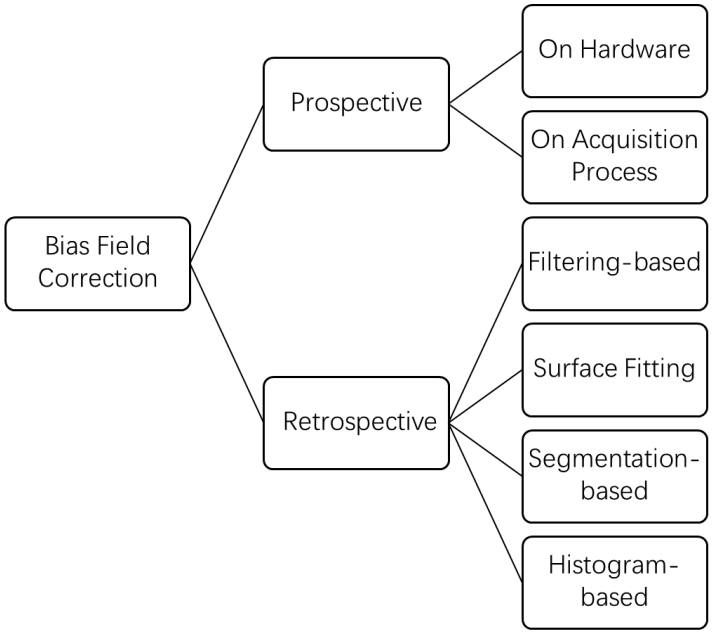


Fig. 3.7 Summary of bias field correction methods.

As, shown in Fig.3.7, the correction methods of bias fields in medical images can be categorized into prospective and retrospective approaches [133]. Prospective approaches aim at reducing the bias field caused by imperfect acquisition process, such as adjustments of hardware and acquisition devices. Retrospective approaches correct the bias field generated by some specific properties of the image object, such as the shape, position, orientation of the object. These methods are based only on image intensities and prior knowledge about image objects. This section will only review image processing level bias correction.

The bias field is commonly modelled as low-frequency artifact and smoothly varying function changes the image intensities [151]. The main correction methods are based on the following three different models [151]:

Model 1: $$S(x,y) = I(x,y)B(x,y) + \eta(x,y) \tag{3.3}$$

Model 2: $$S(x,y) = (I(x,y) + \eta(x,y))B(x,y) \tag{3.4}$$

Model 3: $$\log S(x,y) = \log I(x,y) + \log B(x,y) + \eta(x,y) \tag{3.5}$$

where S is the bias corrupted image, I is bias-free image, B is the bias field, η is the noise, (x, y) is index of pixels in image. Based on these assumptions, a range of algorithms were proposed to reduce the bias field artefacts.

As shown in Fig. 3.7, the retrospective approaches can be further classified into filtering-based methods, surface fitting based, segmentation based, and histogram-based methods. The brief introduction of these sub-categories is in Table 3.1.

Table 3.1 Retrospective Method for bias field correction of medical images

Retrospective Method	**Description**	**Examples**
Filtering Based	Use low-pass filter (LPF) to remove low-frequency artefacts from high-frequency components of obtained images	homomorphic filtering [152] homomorphic unsharp masking [153]
Surface Fitting Based	Model the bias field as a parametric surface which is usually a polynomial or spline function.	Intensity based [154] Gradient based [150]
Segmentation Based	Introduce bias field related factors into segmentation methods, accurate segmentation makes the bias correction insignificant.	Bias field considered Expectation Maximization [155] Adaptive Fuzzy C-mean (FCM) [156]
Histogram Based	Exploit the image histogram (intensity or gradient) to automatically correct for the bias field in medical images	nonparametric nonuniform normalization (N3) [157] local entropy minimization with a bicubic spline model (LEMS) [158]

3.2.1.5 Intensity Standardisation

As mentioned in Fig.3.2 and section 3.2.1.4, there are intensity variations contained in the same type of tissues inside single MRI slice, i.e., bias field, the intensity variations also exist between different MRI slices, which refers to that same types of tissues have different range of intensities in different MRI slices. This kind of intensity variations will also influence intensity-based segmentation of medical images, especially 3D

segmentation. To solve this, the histogram of a single slice can be adjusted, and subsequently the histograms between slices can also be standardized to make them similarly distributed. The common intensity standardisation (IS) procedure is to use one MRI slice or one series of MRI slice with same intensity distribution as reference images, to train a target distribution then mapping new input images to the target intensity distribution. The HE introduced in 3.2.1.1 can be also used for the intensity standardisation, however more research on this is shown in Table 3.2.

Table 3.2 Review of different types of intensity standardisation methods

Papers	**Methods**	**Description**
[159-163]	Histogram Matching	Use histogram landmark (percentile widely used) points from training set to build transformation (linearly or use spline) function between images
[164-166]	Joint Histogram Registration	Build joint histograms (in 2D image format) of reference slices and current slices, then apply nonrigid registration between them
[167]	Mixture Mapping	Use Gaussian Mixture Model to estimate mean intensities of major tissues in reference image and source image, and mapping them according to the tissues' mean
[168]	Generative Adversarial Network (GAN)	Use a modified CycleGAN [169] to generate another version of source image with different intensity distribution but same in other features

Table 3.2 summaries the different types of methods for intensity standardisation, a figure illustration of the influence of intensity standardisation is shown in following Fig. 3.8.

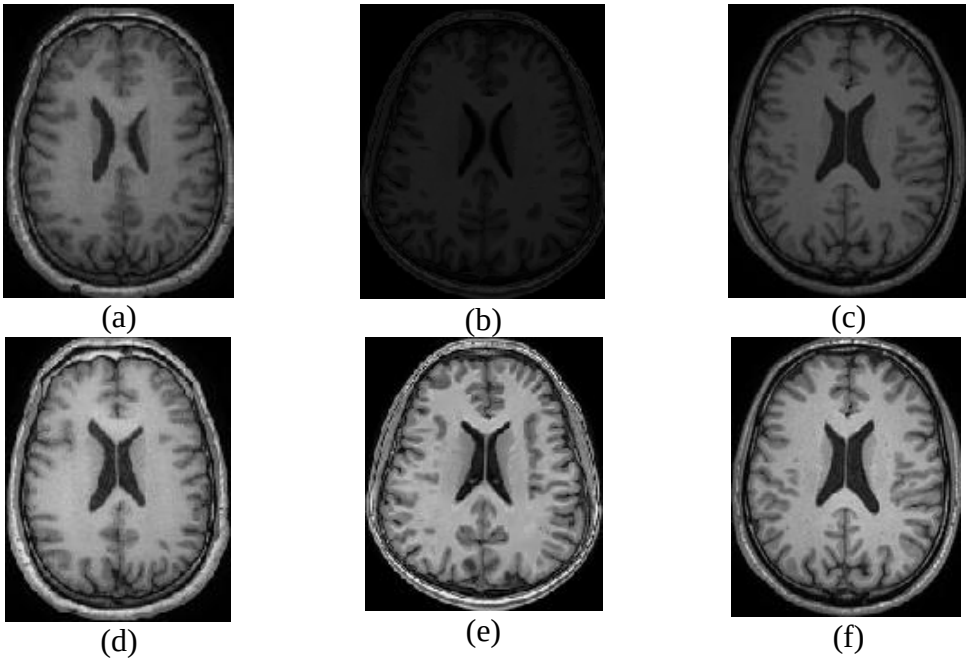


Fig. 3.8 Sample images from GE, Siemens and Phillips scanners (left to right) before (a-c) and after (d-f) intensity standardisation [49].

Before IS, the MRI slices from different scanners (shown in Fig. 3.8 (a-c)) are all taken from brain but have serious intensity variations among them. Same tissue (such as white matter) in Fig. 3.8 (b) is significantly darker than (a) and (b), this will influence intensity-based image analysis unless discard this series of data. After IS, slices (d-f) have similar intensity distribution among same type of tissues (cerebrospinal fluid, white matter, gray matter) in MRI slices from different scanners. In this way following intensity-based processing will be improved and data from all scanners can be utilized.

3.2.1.6 Interpolation for 3D construction of data

There is known that different MRI facilities/protocols will generate MRI data with different slice thickness and distances. Each MRI slice has its thickness, thus 'pixels' in MRI image are actually 3D voxels with 3 dimensions of size. If the slice distances are small enough compared to voxel size, the obtained MRI image sequence can be processed as 3D data volume. Some scanners can only produce MRI slices with limited spatial resolutions and slice distances, thus leading to need for the slices to be processed as 2D images rather than 3D volume. For the sake of better quantization and visualisation of tumours, 3D segmentations are generally preferred. Thus, a pre-

processing technique called interpolation should be introduced, which can reconstruct 3D MRI volume from separate 2D slices.

Interpolation means that resample the original images horizontally or vertically so that change the resolution of images, the key is to define the way of calculating values of pixels/voxels in resampled images based on known values from original images. The following Table 3.3 provides a review of some commonly used spatial interpolation algorithms and recent updates in this topic.

Table3.3 Review of image interpolation methods

Methods	**Description**
Filter interpolation(Mean Filter, Gaussian filter [170], Lanczos filtering[171], Bilinear interpolation [172])	Use spatial filter approach (sliding windows/kernels) to calculate the values of interpolated points
Non-filter interpolation (Nearest neighbour [173], B-spline, Cubic spline [174])	Deduce the values of interpolated points based on weighted sum of previous and subsequent points, the weights are from nearest neighbour or spline
Fast Fourier Transform (FFT) interpolation [175]	Interpolate image in spatial domain via appending zeros in FFT domain and recover to spatial domain
Deep Learning interpolation [176, 177]	Resample images use traditional way such as bilinear interpolation, then use features learned by DL to refine the resampled images

The filter methods and non-filter interpolation methods reviewed in Table 3.3 are commonly used in medical image processing as introduced in [171, 178, 179]. Nearest neighbour interpolation leads to jagged artefacts, spline techniques involve inherent blurring and distorted edges [180]. The deep learning methods take expensive computation power and are majorly used in super resolution applications rather than

3D reconstruction. In the previous work of our group [181], the FFT interpolation of 2D MRI slices to reconstruct 3D HN MRI volume has been proposed. The reconstructed volume has been used for further image analysis task, which is 3D segmentation of throat. An example of FFT interpolation of HNC slices is given in Fig. 3.9.

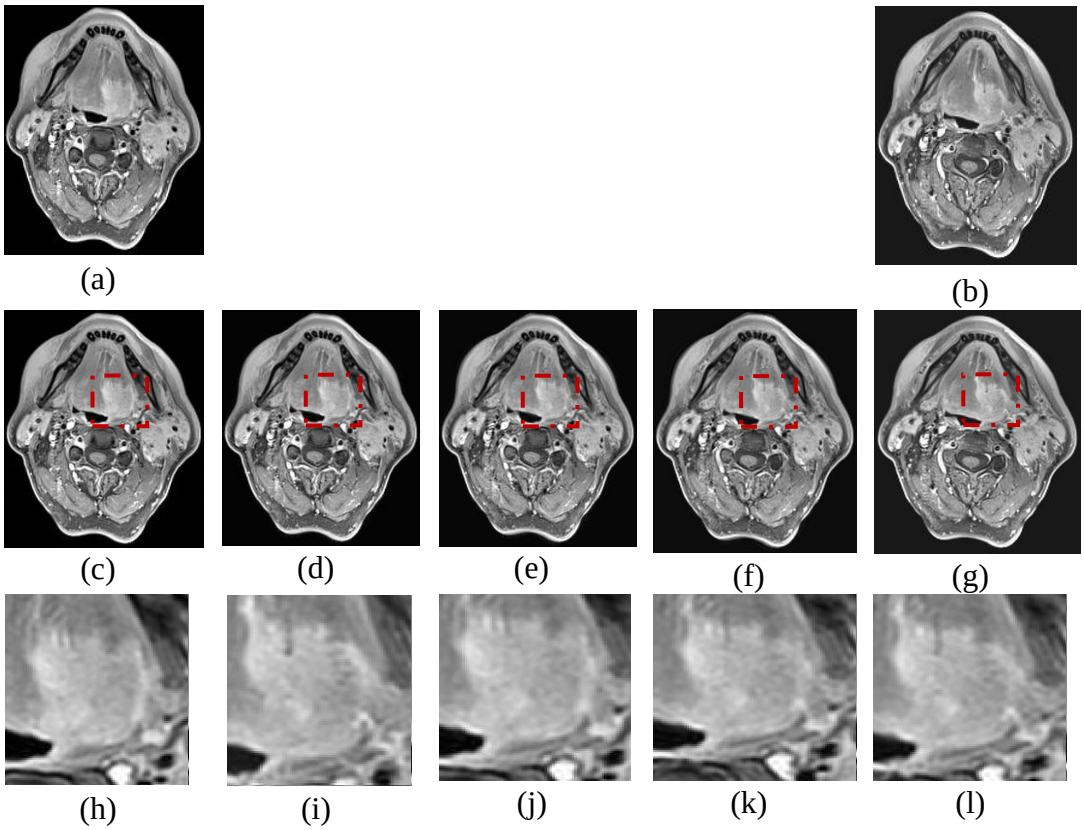


Fig. 3.9 (a)(b) two consecutive MRI slice before interpolation. (c-g) are two slices after interpolation and three synthetic slices generated in interpolation. Red dash boxes mark the rough tumour position. (h-l) The zoomed tumour region after interpolation.

Fig. 3.9 gives an example of interpolating MRI slices ((a) and (b)) so that generate synthetic interpolated slices ((c-g)) between consecutive MRI slices, from the zoomed view (h-l) it can be seen the information of HNC between two original slices changes smoothly among the interpolated slices, so that a 3D MRI volume is constructed for further 3D operation without losing or changing structural information in original images.

3.2.2 Initialization of interactive segmentation

As shown in Fig. 3.1 (a), the interactive medical image segmentation workflow solves the segmentation tasks by diving them into two phases. First phase is detection to locate the Region of Interest (ROI) of this segmentation, i.e., initialization. Second phase is to evolve the area of foreground (ROI) to fit it with the clinical target in medical images, i.e., interactive segmentation. Conventional way for detection (initialization) is manually drawing the initial point, line, or contour on the target, this can be replaced by automatic algorithms in some applications. Different types of initialization methods are summarized in Table 3.4.

Table 3.4 Summary of initialization methods for interactive segmentation

Method	**Description**
Seed-based Manually [182-184]	Manually define accurate seed points around the target of segmentation
Region Seed-based Manually [185, 186]	Use seed points roughly mark foreground and background in segmentation
ROI-based Manually [122, 187]	Draw ROI around the target object, the foreground and background will be decided from pixels inside and outside ROI
Machine Learning [188-192]	Use unsupervised/weakly supervised machine learning or deep learning methods to detect the target object
Knowledge Based [193-196]	Transfer expert knowledge into rules in computer algorithm to detect target object

Table 3.4 shows different categories of initialization methods for interactive segmentation include manually and automatically. The automatic methods [191, 192]

use weak supervision (points/lines weak label rather than accurate pixel-wise label) to train deep neural network for segmentation, which can be called as deep interactive segmentation. The knowledge-based methods [193-196] utilize prior domain knowledge from expert to design detection algorithm for domain-specific task. There are many unsupervised image segmentation methods [188-190], which can be solely used for end-to-end segmentation, while they can also be used as initialisation methods to draw an initial segmentation for following interactive segmentation methods. There are different ways of delineation of manual initialization as shown in Table 3.4, the figure illustration of manual initialization is shown in Fig. 3.10.

(a) (b) (c)

Fig. 3.10 Different ways of manual initialization of interactive segmentation. (a) Seeds-based, contour of target is drawn by red circles. (b) Region seeds based, white lines mark foreground, red lines mark background. (c) ROI based [122], red dashed box mark the ROI.

Fig. 3.10 demonstrates the manual initialization of interactive segmentation. The seeds-based methods (Fig. 3.10 (a)) must draw closely to the boundary of target, which is expensive in time and highly subjective to human interaction, so they are not commonly used nowadays. The region seeds (in (b)) based methods provide a more convenient way than seeds based (in (a)), but still depend much on good manual initialization (see the several short white lines on giraffe head in (b)) to guide interactive segmentation well, they are used in many weakly supervised segmentation also [191, 192]. The ROI based method (in (c)) gives easiest way to users for initialisation, the major requirement is to conclude the target in the drawn ROI. The ROI based methods are efficient in time, and less affected by subjectiveness from

human, thus they are used as common initialisation ways for interactive segmentation nowadays.

3.2.3 Interactive segmentation algorithms for medical image segmentation

With the course location of the target in an image, there are different interactive segmentation algorithms that can start iterating from the initial detection to converge at the condition that the segmentation best fit the target.

Table 3.5 summarized three common types of interactive segmentation methods, including graph-cut based methods, seeded region growing, and deformable models.

Graph-based approaches treat each pixel in an image as a node in a graph. By defining and optimizing the cost function among nodes, the graph-cut can generate superpixel segmentations [197-200] for image, which is similar to clustering methods [200]. if prior-knowledge (manual or automatic initialization) is given, many graph-cut based segmentation methods [122, 186, 201, 202] can extract regions/volumes for target object in image. The applications of graph-cut based segmentation on medical images have also been reviewed in Table 3.5, across different imaging modalities and body parts.

The Seed Region Growing (SRG) was firstly proposed in [203], which uses user-initialized seeds to achieve segmentation of target object. There are three main aspects need to care for SRG also for many other interactive segmentation methods, which are, 1) initialisation, 2) Growing rule, 3) Stop criterion. The initialisation can be from manual or automatic. The growing rule is decided by similarity measure such as the difference between intensity of pixels to of seeds. The process will stop at no neighbour pixels fit the similarity measure. The applications of SRG and its variants [204] on medical images are reviewed in Table 3.5, across different imaging modalities and body parts.

Table 3.5 Summary of common interactive segmentation algorithms

Method	Description	Applications (target/data)
Graph Cut (GC) based [122, 186, 201, 202]	Solve image as a graph, where the pixels in image are nodes (vertices) in graph, the nodes are connected by edges. Segmentation is done by disconnecting edges between nodes at lowest energy cost, where energy is defined by image features.	Lung Cancer, Liver, Bones, HNC / CT [205-208] Cardiac, Brain tumour, Prostate / MRI [209-211] Liver, Lymph nodes/ Ultrasound [212, 213]
Seeded Region growing and variants (SRG)[203, 204]	Given initial seed region or points, the SRG search pixels around seeds to grow (enlarge) the initial seed region via absorbing pixels who fit the similarity measure, the similarity is from the comparison between features of pixels in seed region and pixels around.	Abdomen, Teeth, Cardiac, lungs /CT [214-217] Brain, Abdomen, breast /MRI [218, 219] Breast, liver tumour /Ultrasound [220] [221]
Deformable model [182, 222-224]	Given initial seed region or seed contour, the contour evolves under the force of deformable model. The contour will converge at high gradient place and maintain the regular shape under the constrain of curvature.	Tooth, Liver tumour, Vertebra /CT [225-227] Brain tumour, heart, HNC / MRI [228-231] Breast, prostate/Ultrasound [232, 233]

3.2.3.1 Deformable models family

Finally, another class of interactive segmentation methods reviewed in Table 3.5 is deformable model. Given a 2D image or 3D tensor image, deformable models use 2D contours or 3D surfaces to separate image into sub-images by minimizing pre-defined

energy, so that segment objects and targets in given image. The deformable models can be categorized as parametric (explicit contours) and geometric (implicit contours) methods. Snake [182] model is a representative parametric method, Level Set Method (LSM) and Chan-Vese (CV) model are representative geometric methods.

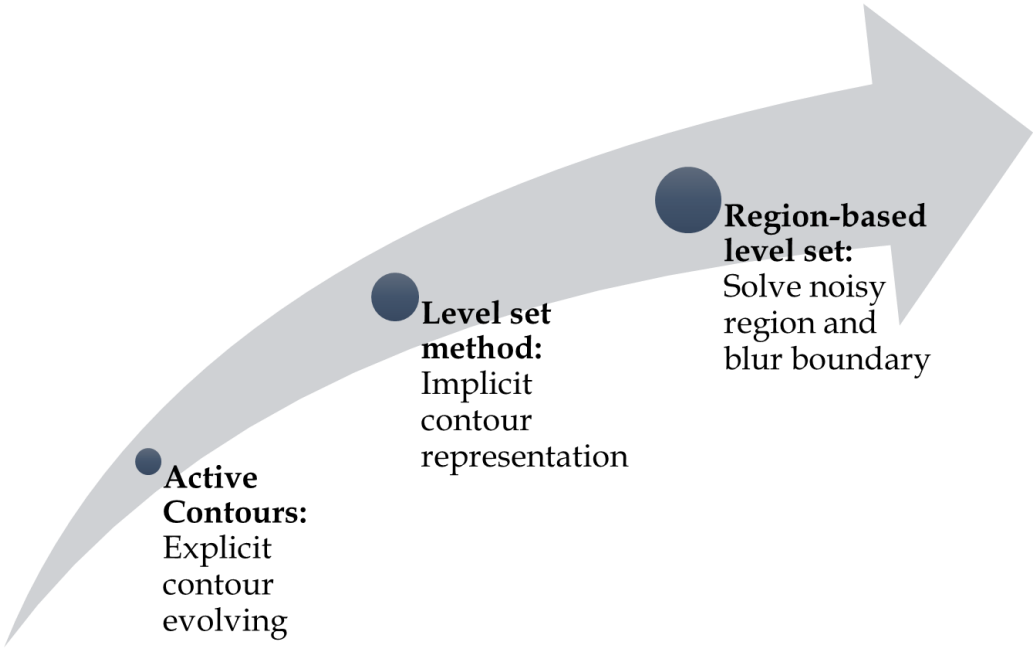


Fig. 3.11 A brief history of development of deformable models

Fig. 3.11 demonstrates three important developments of deformable models. The first stage of deformable model history is the invention of Active Contours (AC). In 1988 [182], Snakes (also known as active contour model) was introduced as a framework for interpreting high-level knowledge of images. The active contour model partitions image into sub images by using energy-minimizing spline. The position of a snake is parametrically represented as $\mathrm{V}(\mathrm{s}) = (x(s), y(s))$. The spline in snake model can be parametrically initialised by user interface or automatic pre-process. Then spline will be updated along the minimization of following functional Eq. 3.12 and Eq. 3.13,

$$E_{snake} = \int_0^1 E_{snake}(v(s))ds \tag{3.12}$$

$$E_{\text{snake}} = E_{\text{internal}} + E_{\text{external}} \tag{3.13}$$

The initialized contour evolves based on Eq. (3.12) and Eq. (3.13) then stops until pre-defined number of iterations or the Snake energy is minimized.

The classical Snake method requires initialization sufficiently close to edges, because it only uses local information along the contour [234]. Some researchers have made efforts to the modification of parametric deformable models. Such as methods in [235, 236]. A well-known Gradient Vector Flow (GVF) method was proposed in [237]. This method introduced a novel gradient vector flow as an external force E_{external} to involve region-based features, which significantly increases the capture range.

While the snake model is parametrically constructed by series of markers, which means during the evolution certain regriding mechanisms are required to avoid overlap of control points [238]. Also, using fixed number of points to represent a deformable model is inconvenient for handling topological changes during the curve/surface evolution, such as merging and splitting.

To solve limitations in parametric deformable models, geometric level set methods (LSM) are introduced to move contours implicitly as a particular level of a function [234]. This refers to the second phase of deformable model development in Fig. 3.11. The concept of implicit contours is to define a high dimension surface with a constructed function as Eq. 3.14.

$$z_{\phi} = \phi(x, y, t = 0) \tag{3.14}$$

The point with coordinate (x, y, z) lies on the surface, this surface evolves with time(iteration) t. In each time step, the $\phi = 0$ represents the evolving contour, $\phi(x, y, t) > 0$ means point is outside contour, and $\phi(x, y, t) < 0$ means inside contour. In part of the open literature $\phi(x, y, t) > 0$ means point is inside contour and vice versa.

Then the curve or surface is propagated by evolving a function $\emptyset(\mathrm{x}, \mathrm{t})$ according to pre-defined speed *F* along the curve/surface normal. The variable *x* is time-dependent, and the curve/surface *C(t)* is always defined as zero level of level sets as Eq. 3.15.

$$\emptyset(\mathrm{C}(\mathrm{t}), \mathrm{t}) = 0 \tag{3.15}$$

where $\emptyset(t)$ is level set surface at time t, and C(t) is the contour at time t, each time the contour is at zero level of $\emptyset$. This definition releases the limitation of fix number of contour points in parametric methods. Now the problem is how to update implicit surface with time, the level set function $\emptyset$(x, t) is iteratively updated with time as,

$$\frac{\partial \emptyset}{\partial t} = -\nabla\emptyset \frac{\nabla\emptyset}{|\nabla\emptyset|} F \tag{3.16}$$

$$\frac{\partial \emptyset}{\partial t} = -|\nabla\emptyset| F \tag{3.17}$$

where F is speed function to guide the evolution, $n = \frac{\nabla\emptyset}{|\nabla\emptyset|}$ is the normal of the surface show the direction of evolving, $\nabla\emptyset$ is the gradient of $\emptyset$ at position $\emptyset(x, y)$.

Based on the Eq. (3.14) - (3.17), the level set $\emptyset$ updates with time. Compared to AC, the geometric based methods also have advantages to handle topological changes during interactive segmentation, so that naturally find and delineate multiple objects. The geometric based methods significantly improved the efficiency of deformable model, and solved many limitations existed in parametric methods. Nowadays, the majority of deformable models use level sets to represent contours.

Apart from the way of representation of contours, there have been many other modifications proposed on deformable models. In [223, 239, 240], modifications were proposed to reduce the computation cost of LSM and speed up the converge. In [238], approaches which combine statistical information with level set segmentation are reviewed, such as integrating colour, texture, and shape prior. Another important development (in Fig. 3.11) to level set is introducing region-based fitting energy to level set in [224], the fitting energy F is modified as,

$$F_1(C) + F_2(C) = \sum_{(x,y)\in \text{inside(C)}} |\emptyset(x, y) - \mu_{in}|^2 + \sum_{(x,y)\in \text{inside(C)}} |\emptyset(x, y) - \mu_{out}|^2 \tag{3.18}$$

where C is the evolving curve (contour), F_1, F_2 represents energy inside and outside current curve. The $\emptyset(x, y)$ is intensity of a pixel at (x, y), μ_{in} is mean intensity inside curve, μ_{out} is mean intensity outside curve. The values of F_1, F_2

depend on the relative position between C and target object, the fitting energy will be minimized when the segmenting curve is right at boundaries of target object. The region-based fitting energy gives many advantages to LSM segmentation. Firstly, it allows more inaccurate initialization of starting contour, the initialization does not have to be close to target's edge. Then, it is not sensitive to noise and no need for smooth of initial image before segmentation. In addition, the region-based methods can segment object with no sharp boundaries compared to edge-based methods.

Deformable model can keep the smoothness of curve and use it as a regulation in image segmentation. This introduces benefits to medical image segmentation. The deformable models are widely used in many areas include medical applications[241-244]. In Table 3.5, applications of deformable models on medical image segmentation across different body sites and imaging modality are reviewed.

The impact of computer-based interactive segmentation methods on RTP are presented and reviewed in many research work. In [245], the work of automatic segmentation aiding RTP of cervical cancer are reviewed, it shows the segmentation of structures such as bladder, uterus, and cervix in MRI and PET/CT data help the RTP of cervical cancer. In [246], the automatic segmentation methods of prostate cancer from MR images are validated, the results show that automatic segmentation methods most likely reduce the overall contouring time and introduce minor deviation compares to clinicians and radiotherapy oncologists. Many other work [33, 247, 248] also present the benefits of interactive segmentation methods on RTP across GC-based, SRG, and deformable models. In section 2.4.4, many auto-contouring software are introduced, majority of them integrate the interactive segmentation methods.

The methods reviewed so far can perform well with limited levels of features and requires relative low computation cost. While these methods cannot automatically match semantic labels to segmentations. Also, the performance of interactive segmentation methods relies on the initializations, this involves additional human effort, also introduces subjectiveness. In following sections, techniques of automatically detecting and segmenting targets from medical data will be reviewed.

3.4 Conventional Machine learning for medical image processing

Apart from interactive medical image segmentation, another category of solution is using machine learning to tackle both detection and segmentation challenges in medical images. As shown in Fig.3.1 (b)(c), the general workflow of supervised image processing includes training and testing phases. In training steps, the trainset data and its labels are used in pairs to optimize a discriminator. In the test phase, the well-trained discriminator is used to generate predictions for test data, such as predict classes and output segmentations. As shown in Fig. 3.1 (b), in both unsupervised and supervised methods, the feature extraction is a vital step. Although some methods only rely on intensity and gradient features, many high-level patterns require more sophisticated feature extraction methods to obtain, so that to complete complex visual processing tasks. In the rest of this section, the commonly used feature extraction methods for medical images will be introduced in 3.4.1, the general conventional classifiers for medical image segmentation will be reviewed in 3.4.2.

3.4.1 Feature extraction

In original image domain, the hidden patterns of information inside medical images are hard to obtain by solely analysing simple features such as intensities. Also, it is hard to distinguish regions or objects inside an image straightforward in image domain. Thus, it is necessary to introduce feature mining to find hidden patterns of an image thus projecting data into feature space, where target foreground objects/ regions can be discriminated from background. A summary of common feature extraction methods of medical image processing is in Table 3.6.

As shown in Table 3.6, there are various types of features can be used in medical image processing. Some methods such as SCC [249, 250] capture the low-level features such as edges, similar approaches include Canny [251], SUSAN (Smallest Univalue Segment Assimilating Nucleus) [252], and Sobel Detector [121]. Methods such as LBP and GLCM extract local and global texture features to find more complex patterns inside images. Other methods such as HOG and SIFT finds high-level features to represent objects in classification and segmentation of clinical targets. The PCA

[253] can select more relevant features to achieve tasks if there are much redundancy in images or features.

Table 3.6 Summary of feature extraction methods of medical image processing

Method	**Description**	**Application**
Principle Component Analysis (PCA) [253] [254] [255]	The PCA projects high-dimensional data onto reduced number of dimensions to capture the important pattern and ignoring redundant information	Medical image fusion and registration [256, 257] Feature extraction for medical image segmentation [258-261]
Slope Chain Code (SCC) [249, 250]	Use changes of slope to represent boundaries in image, such as tortuosity	Diagnosis of eye disease from retinal images [249, 250]
Local Binary Pattern (LBP) [262]	Encode local texture feature around a pixel using binary digits, e.g., higher than central pixel is TRUE, lower is FALSE	Blood vessel segmentation [263] Uterus segmentation [264] Breast Cancer Detection [265]
Gray Level Co-occurrence Matrix (GLCM) [266]	Encode texture feature through summarize the occurrence of intensity patterns of adjacent pixels	Lung cancer segmentation [267], breast cancer detection [268], pathological area [269]
Histogram of Oriented Gradients (HOG) [265]	Extract features in image via calculating gradients of pixels and summarize then by histogram	Ventricle landmark detection[270], liver cancer detection [271]
Scale-Invariant Feature Transform (SIFT) [272]	Find key feature points of images which are invariant to scale, noise, illumination, and rotation.	3D liver segmentation, prostate segmentation [273, 274]

3.4.2 Classifier

As shown in Fig. 3.1. (b), the acquired features can be used in classifiers/discriminators to achieve high-level pattern recognition tasks, such as medical image segmentation. This section will majorly review the classifiers which can be used for medical image segmentation, a summary is shown in Table 3.7.

Table 3.7 Review of common classifiers for medical image segmentation

Method	Description	Application
Decision Tree (DT) [275]	Use hierarchy tree model to build a discriminator. Nodes are output classes of input, which are decided by selected features at branches.	Brain Segmentation [276, 277]
Random Forests (RF) [278]	Ensembles multiple binary decision trees, introduce internal feature selection and majority voting.	Liver, Kidney, Bone Segmentation [279], Cardiac [280]
Support Vector Machine (SVM) [281]	Learning a hyperplane from training samples to best separate input into classes	Lung segmentation[282] Brain tumour segmentation [283]
Artificial Neural Network (ANN) [284]	An artificial uses the multiple input unit and weight between connections to simulate the signal transmission from dendrites to neuron nucleus, so that classify input	Bladder, Chest, Brain Segmentation [285, 286]

As shown in Table 3.7, There are different types of classifiers can utilize extracted features (described in Table 3.6) to discriminate different regions and objects in medical images so that achieve segmentation. The DT [275] algorithm can select attribute (features) to split the data correctly, using techniques such as Information

Gain and Gini Index [287]. On the other hand, another way to improve the robustness and classification accuracy of DT is RF [278]. The RF starts from initial multiple binary DTs with random selected features from entire feature pool, the output of RF takes majority voting of DT's' output. The RF is an ensemble learning method based on DT, this ensemble learning idea is also used in other machine learning approaches such as ensemble of multiple neural networks. The SVM [281] is looking for a hyperplane in high-dimensional features spaces to best divide data samples from different classes. The ANN [284] uses large numbers of neuron-like units to theoretically learn any nonlinear transforming function, so that map input data into different classes. The medical image segmentation applications are also given in Table 3.7, across different body sites and image modalities.

3.5 Deep convolutional neural network for medical image segmentation

The last section introduced the classical machine learning algorithms and some of their applications in medical image segmentation. As shown Fig. 3.1 (c), the development of deep learning (DL) has replaced many traditional machine learning approaches through end-to-end learning, which combines feature extraction and classification in an entire structure and is trained together. Also, the DL will save the subjectiveness of hand-crafted features from conventional feature extraction methods.

3.5.1 History from ANN to DCNN for medical image segmentation

The DL has a very broad subtypes, such as deep recurrent neural network, deep convolutional neural network, and autoencoder. This section will only review how the deep convolutional neural network (DCNN) has been developed to perfectly match the demanding of solve sophisticated pattern recognition problems in computer vision field, especially the medical image processing. In Fig. 3.12, a brief history of the ANN technique developing to DCNN so that capable of medical image segmentation task is shown.

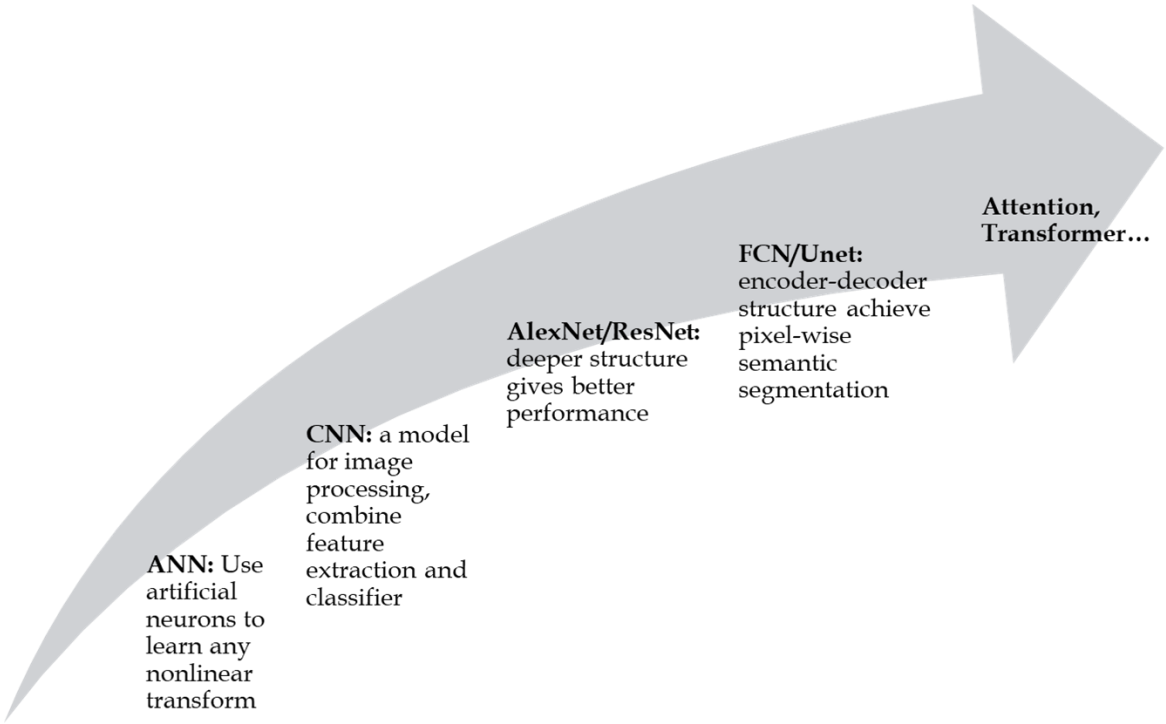


Fig. 3.12 Brief history of DCNN development on computer vision and medical image processing

As shown Fig. 3.12, at first stage, ANN [284] was proposed to simulate the highly parallel computing structure and imprecise information-processing capability of real neural networks, so that develop the artificial intelligence methods. ANN are constructed by multiple layers of fully connected neurons, thus also called as fully connected layers. The ANN can be used in regression and classification tasks, but it does not act as feature extractor, neither works much on images. While there are some applications on medical image segmentations such as in [285, 286]

The second stage of Fig. 3.12 is from [288], a convolutional neural network (CNN) architecture is proposed for hand-writing recognition. In this neural network, the feature maps of input images are extracted by convolutional layers. The convolutional layers use windows-like filters rather than fully connected filters to extract features so that keep the spatial information of images. The obtained features are classified by cascade of fully connected layers to get the correct recognition of hand-writing

characters. Compared to conventional machine learning methods, the CNN combines feature extraction and classification together into an end-to-end algorithm.

The third stage in Fig. 3.12 began in 2010s, based on several technical breakthroughs on DL [289], many DCNN models with large numbers of parameters significantly improved accuracy for many image classification tasks of large datasets, such as works from VGG [290], Alex-Net [291], and [292]. These models used multiple or dozens of convolutional modules in [288] to extract abundant deep features from images for better analysis. While these methods were only on classification of entire images. There were some DCNN based image segmentation method which transfer the segmentation into the classification of a central pixel inside an enclosing object or region, which was called patch-wise segmentation. The patch-wise training involves additional efforts to define and prepare the patches for segmentation, it could be equal windows, super-pixels, or region proposals. The patch-wise solutions lacked efficiency of fully convolutional training. There were some patch-wise DCNN used in medical image segmentation such as in brain segmentation [89, 113, 293] and bone structure segmentation [294].

To solve the challenge of DCNN working on image segmentation, the third stage comes in [295], a fully convolutional network (FCN) is proposed for end-to-end pixel-wise prediction of natural images. The FCN can be globally trained and generate dense prediction of each pixel's class, this means assigning all pixels in image with semantic label, which is a pixelwise classification but from image level the segmentation is achieved. This kind of work is called semantic segmentation. Here are many researches inspired by FCN to further improve the semantic segmentation, such as U-Net [296], and SegNet [297]. The U-Net was proposed for the segmentation of cells from microscopy images, and nowadays it has been widely applied in many types of medical image segmentation. Due to U-Net is highly related to this thesis, more introduction of it will be arranged in section 3.5.2.

Recently, there are other updates on DCNN to improve computer vision as well as medical image segmentation performance. The development of visual attention scheme [298] and vision Transformer [299] can suppress irrelevant information in DCNN and focus on salient features for the specific task to further improve the

performance in many DCNN for medical image segmentation tasks, such as pancreas segmentation [300], liver segmentation [301], and skin cancer segmentation [302].

3.5.2 U-Net: an encoder-decoder structure DCNN for semantic medical image segmentation

As introduced section 3.5.1, the invention of FCN [295] and U-Net [296] developed image segmentation problem from patch-wise classification into pixel-wise classification, which is so called as semantic image segmentation.

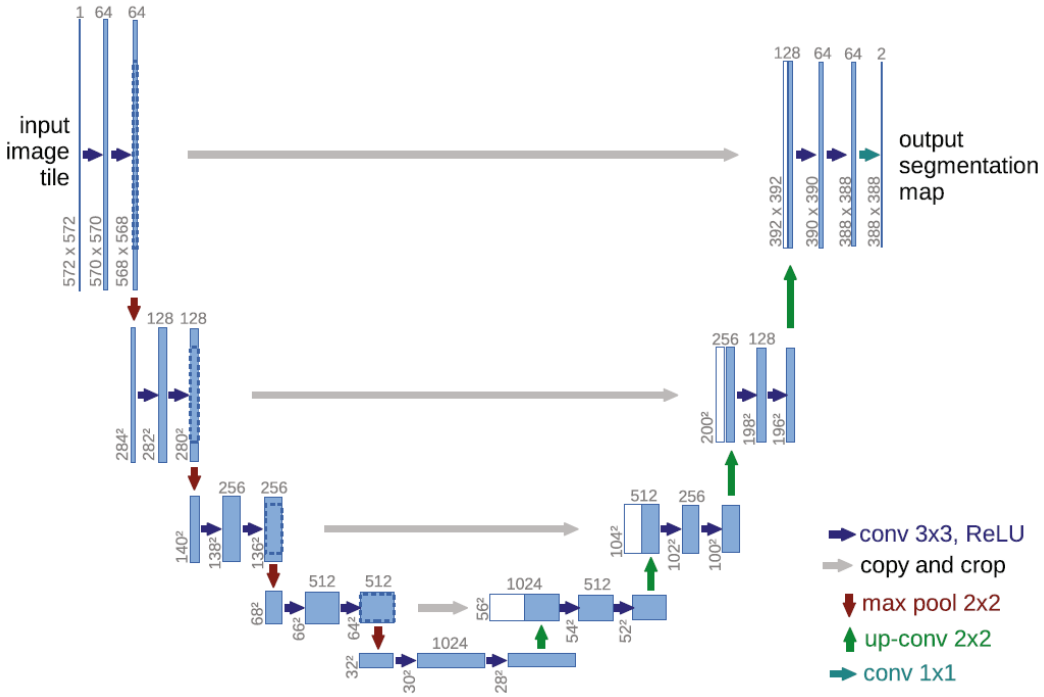


Fig. 3.13 U-Net architecture [296]

Like FCN, the U-Net has two major parts, downsampling stage and upsampling stage. As shown Fig. 3.13, the downsampling stage is cascade of convolutional layers, which downsamples the size of input image, but increase the depth of feature maps after each convolution, which is like typical architecture of DCNN. To achieve the pixel-wise dense prediction, the size of feature maps will be recovered in the upsampling stage, and depth of feature maps will be decreased. In the end of upsampling stage, a 1 x 1 convolutional layer is used, which uses 1 x 1 size filters, the number of filters is the number of possible output classes, so that plays a role of classifier.

The combination of downsampling feature extraction and upsampling dimension reconstruction can produce dense pixel-to-pixel classification results, but the predictions are quite coarse. Thus, another important novelty from FCN is combining 'what and where' to obtain good segmentation. This is achieved by the skip connection scheme shown with grey arrow in Fig. 3.13. Feature maps from downsampling stage are skip connected to upsampling stage and fused by concatenation with feature maps same size there, the fused feature maps are then upsampled and fuse another skipped feature maps from downsampling stage. The fuse of feature maps sufficiently used the low layer fine information which contains the details of image, also the deep layer coarse information which contains the spatial information of image, so that improve the detection and segmentation accuracy simultaneously. The FCN uses addition for feature fusion, while the U-Net uses concatenation for feature fusion which will increase depth of feature maps, it is proven that concatenation is a better way for feature fusion [296].

Another essential part in U-Net is data augmentation. The data augmentation generate lots of randomly transformed versions of training samples, such as shift, rotation, and random crop, to teach network the desired invariance and robustness properties [296]. The data augmentation can help the training of networks, especially the only limited numbers of annotated images are available. This will be even important for medical images because the size of datasets of medical images are very limited compared to natural images.

U-Net also modified the form of training loss for better segmentation. A weighted map is calculated and added to cross entropy to train the U-Net for cell segmentation, the modified CE loss will better separate borders of cell. Involving other additional loss to CE loss is also used in many image segmentation CNN models, such as Focal Loss [303] and Dice Loss [304].

The U-Net and its variants has been broadly used in medical image segmentation. A review of several applications is shown in Table 3.8.

Table 3.8 Summary of U-Net and its variants on medical image segmentation

Method	**Application**	**Description**
U-Net [296]	Neuronal structures in electron microscopic stacks	Introduce multi-stage feature fusion via concatenation
V-net [305, 306]	Prostate from MRI, HNC from PET/CT	3D version U-Net
Multi-view U-Net [307]	Hippocampal from MRI	Ensemble results from multi-view U-Nets
Deep Dilated U-Net[308]	Rectal cancer and OARs from CT	Introduce dilated convolution modules in deep layers
Two-stage Cascades U-Net [309]	Brain glioma from MRI	Cascade two U-Nets use output of first as input of second
Attention U-Net [300]	Pancreas segmentation from CT	Introduce attention modules into U-Net
U-Net with GAN branch [310]	Vessel segmentation from retinal image	Use a GAN branch to introduce additional loss to U-Net training
Transformer U-Net [311]	Organ segmentation from CT	Introduce multi-head transformer into U-Net

From the review in Table 3.8, the U-Net and its variants have wide applications in medical image segmentation tasks, across sites from rectal to brain cancer, from CT scans to MRI. There are also many other applications thoroughly reviewed in [312, 313].

3.6 Conclusion

This chapter reviewed the applications of computer-based algorithms on the medical image processing. Firstly, many image processing techniques can enhance the quality of raw medical images to reduce the visual challenges of analysing medical images. Furthermore, the computer vision algorithms can provide different types of machine intelligence to automatically recognize the desired patterns from medical images, includes low-level features such as corner and edges, also high-level patterns such as locations. These will all be helpful to reduce workload of clinicians.

In terms of segmentation a target from medical image, there are different solutions. One way is to unsupervised segmentation of an image into pieces and involve manual interaction to guide algorithm to the target. Another way is to produce annotations for medical images and use them to train supervised classifiers. The supervised methods can be extended to two categories: one is the traditional machine learning using hand craft features and classifiers; another one is deep learning use end-to-end models. From the interactive unsupervised segmentation to traditional machine learning, and deep learning, the subjectiveness of segmentation is reducing, the requirements of annotated data and computation power are increasing. An optimum solution for a practical medical image analysis task should be based what resources are available, such as data and facilities.

Chapter 4 Real MRI data and measure metric

4.1 Introduction

Chapter 3 suggested the importance of the quality of data used in algorithm development. Furthermore, it reviewed many image processing techniques aim to enhance the data quality from different aspects.

Data is also a key factor of computer-aided medical image processing. Firstly, there are many causes that can deteriorate the medical image's quality (as covered in Chapter 2 and Chapter 3). Moreover, the collection of medical data for image processing purpose is challenging. The acquirement of medical data demands available facilities, many of them are high cost, such as MRI and PET devices. Unlike natural images, manual delineation of target object's labels in medical images requires expertise of clinicians

This chapter will introduce the available HNC MRI data, which is processed and evaluated in this thesis. Then, some common metrics for measurement of segmentation algorithms are introduced, which will be frequently used in following chapters.

4.2 Real MRI data

The data of this thesis is provided by Beatson West Scotland Centre, at Glasgow. The provided data is axial T1 + Gd (Gadolinium Enhanced T1-Weighted) MRI slices, obtained from three different 1.5Tesla MRI scanners namely Magnetom Avanto from Siemens, Intera Neuro coils from Philips Medical Systems, and Signa HDxt from GE Medical Systems.

T1 + Gd MRI scans were acquired after 15–20 min of intra-venous injection of 0.1 ml/kg, with typical 3–5 mm slice thickness. The range of other imaging parameters were, 3.3–6 mm spacing in between slices, 9.06–20 ms echo time, 542–1066 ms repetition time, 90◦–150◦flip angle, 0.43 × 0.43–0.94 × 0.94 in-plane resolution, 256 × 256–512 × 512 acquisition matrix and 97.65–221 Hz/pixel bandwidth [193].

As introduced in Section 2.2, the axial MRI slices are images taken by scanning from top-down direction, the images are parallel to the ground. In T1 images, the tissues contain fat or water has light intensity, regions such as air, fast-flowing blood. The patients are injected Gadolinium before scanning to get contrast enhanced MRI data. Each slice contains non-isotropic voxels with size of 0.43 X 0.43 – 0.94 X 0.94 mm in xy plane, 3–5 mm slice thickness, but the distances between slices are 3.3 – 6 mm, thus our data is naturally 2D separate slices, additional processing is required to reconstruct the data to 3D form (introduced in Chapter 3). There were MRI data of 17 patients in total received for this thesis, firstly 10 patients were available, then another 7 patients were also collected. In all 17 patients, the head and neck tumours were delineated by clinicians, only 5 patients' cancerous lymph nodes were delineated. The tumours in slices are mainly located among larynx (throat) and tongue of base, some figure examples will be given in next section. The manual delineations are used as 'ground truth' for training and evaluation of computer-aided segmentation models, while actually the ground truth of the tumour position and contour is nearly impossible to be certain. In this work, the so called 'ground truth' are the consensus manual delineations (or gold standard contours) from clinicians to help the development and testing of segmentation algorithms. The procedure to obtain the consensus manual delineations are detailed introduced in [55, 193]. The MRI data is saved in DICOM (Digital Imaging and Communications in Medicine) format. A List of the dataset used in this thesis is given in following Table 4.1.

Table 4.1 List of MRI dataset used in thesis

Dataset ID	Number of slices	Scanner	Modality
MR01052013	8	'Philips Medical Systems'	T1
MR02092014	11	'GE MEDICAL SYSTEMS'	T1
MR06192012	8	'SIEMENS'	T1
MR09082010	13	'Philips Medical Systems'	T1
MR09092010	8	'Philips Medical Systems'	T1
MR10062011	7	'Philips Medical Systems'	T1
MR10062013	11	'GE MEDICAL SYSTEMS'	T1
MR11092012	7	'SIEMENS'	T1
MR12082013	12	'GE MEDICAL SYSTEMS'	T1
MR14012013	8	'GE MEDICAL SYSTEMS'	T1
MR15052014	7	'Philips Healthcare'	T1
MR17102013	12	'SIEMENS'	T1
MR24082012	12	'SIEMENS'	T1
MR27082012	8	'GE MEDICAL SYSTEMS'	T1
MR28082014	18	'Philips Healthcare'	T1
MR29072011	7	'GE MEDICAL SYSTEMS'	T1
MR30052012	9	'SIEMENS'	T1

4.3 Illustration of HNC MRI dataset

This section will give several examples of the MRI data used in this thesis, to visualize the problems and challenges to tackle in this work. The illustration will follow the arrangements 1) MRI slices whose horizontal positions are at throat or base of tongue

(Fig. 4.1 - 4.2), 2) MRI slices those are vertically at centre or side of a tumour (Fig. 4.3), 3) MRI slices with the existence of lymph nodes inside (Fig. 4.4).

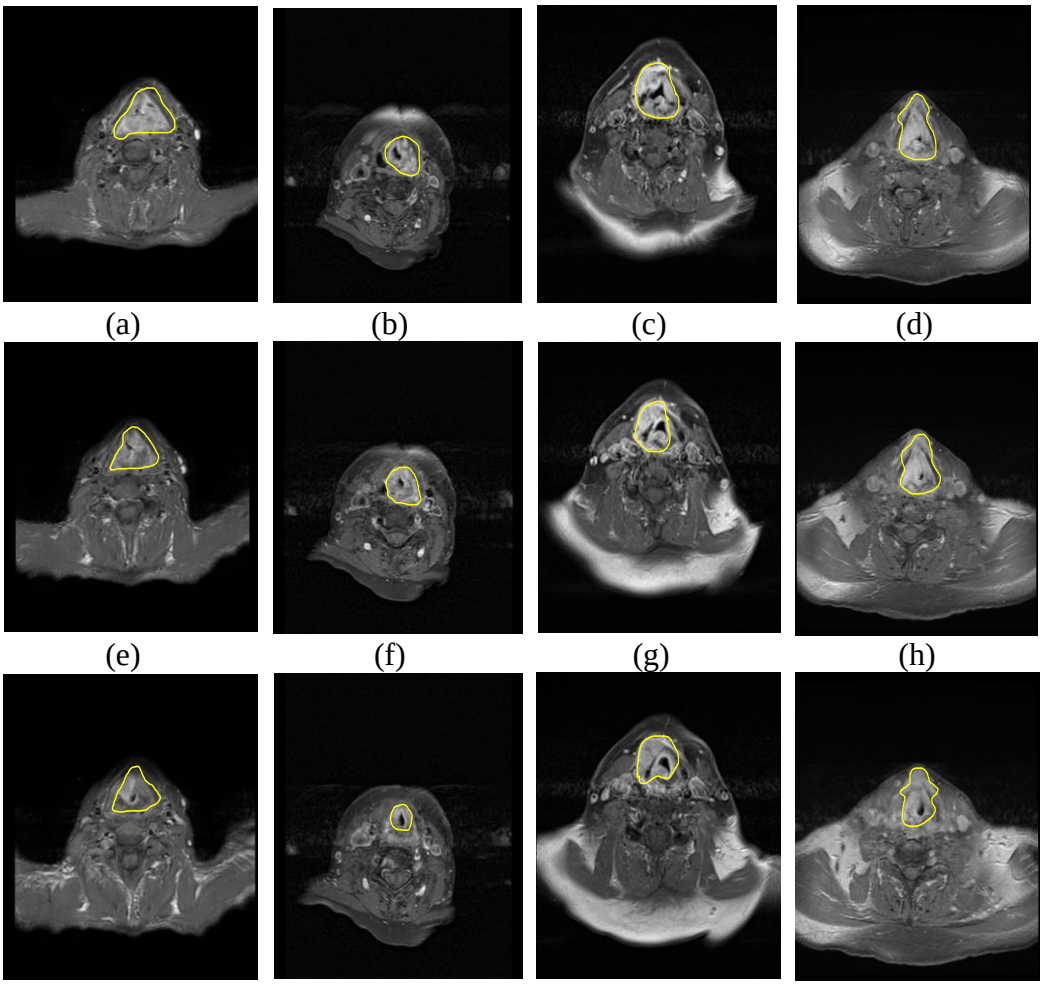


Fig. 4.1 Illustration of figure examples of MRI slices scanned near throat with tumours marked with yellow contours. Each column shows three consecutive slices from one patient, such as (a)(e)(i), in total four patients are shown ((a)(e)(i) from MR 10062011, (b)(f)(j) from MR09092010, (c)(g)(k) from MR06192012, (d)(h)(l) from MR27082012). The example slices show the larynx subarea of head and neck region with existence of head and neck tumour. The yellow contours indicate the larynx tumour regions drawn by experts, which are around throat (black hole inside yellow contours).

Fig. 4.1 gives examples of MRI scans from 4 patients, the slices are around throat area, the tumours here (consensus manual delineations) are marked. The HNCs in this area are majorly around throat, means it can be above, below, left, and right sides of the throat. Call the body of patients as foreground, others are background; the throat tumours are close to the top of the foreground.

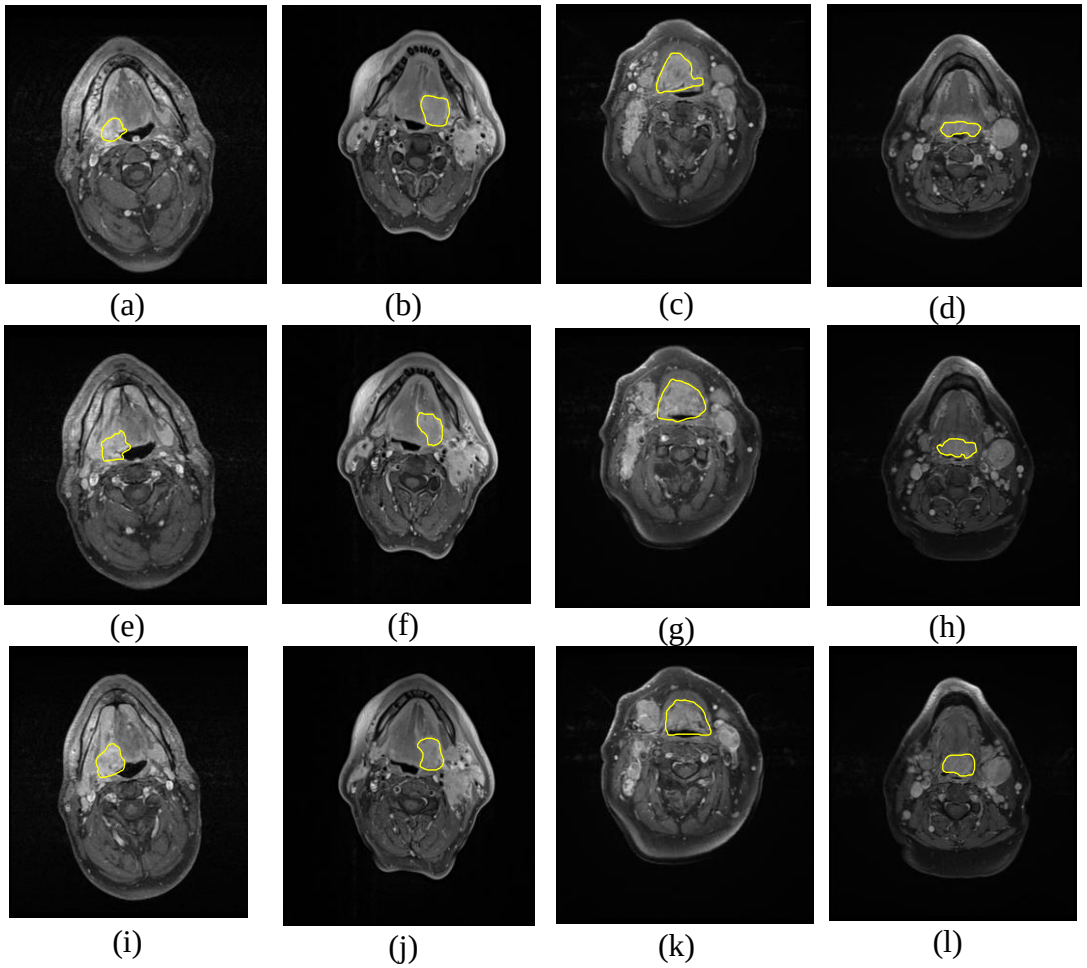


Fig. 4.2 Illustration of figure examples of MRI slices scanned near base of tongue with tumours marked with yellow contours. Each column shows three consecutive slices from one patient, such as (a)(e)(i), in total four patients (a-d) from MR12082013, MR09082010, MR10062013, MR14012013 are shown. The examples show the slices horizontally at base of tongue with presence of HNC marked with yellow contours. The HNCs at base of tongue usually adjacent to one side of throat which is black holes in slices.

Fig. 4.2 gives examples of MRI scans from 4 patients, the slices are around base of tongue area, the tumours here (consensus manual delineations) are marked. The HNCs

here are majorly at one side of throat, mostly adjacent and above the throat. The base of tongue tumours are at the middle region of the foreground in image, where tongues and teeth are at above. There are some neighbour tissues having similar intensities with tumours, such as cancerous lymph nodes (in Fig. 4.2 (d)(h)(l)), and salivary glands (in Fig. 4.2 (c)(g)(k)). Sometimes the enlarged tumour can infiltrate into neighbour tissues, such as in Fig. 4.2 (a)(e)(i). Bias field can be seen in Fig. 4.2 (a)(b)(d)(f)(h)(j)(l).

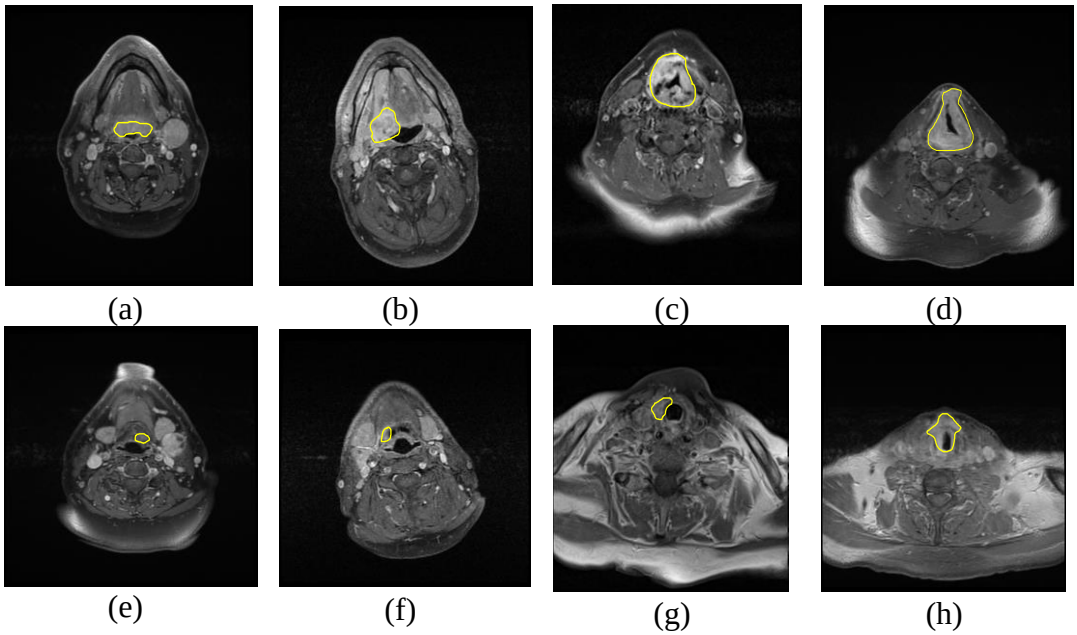
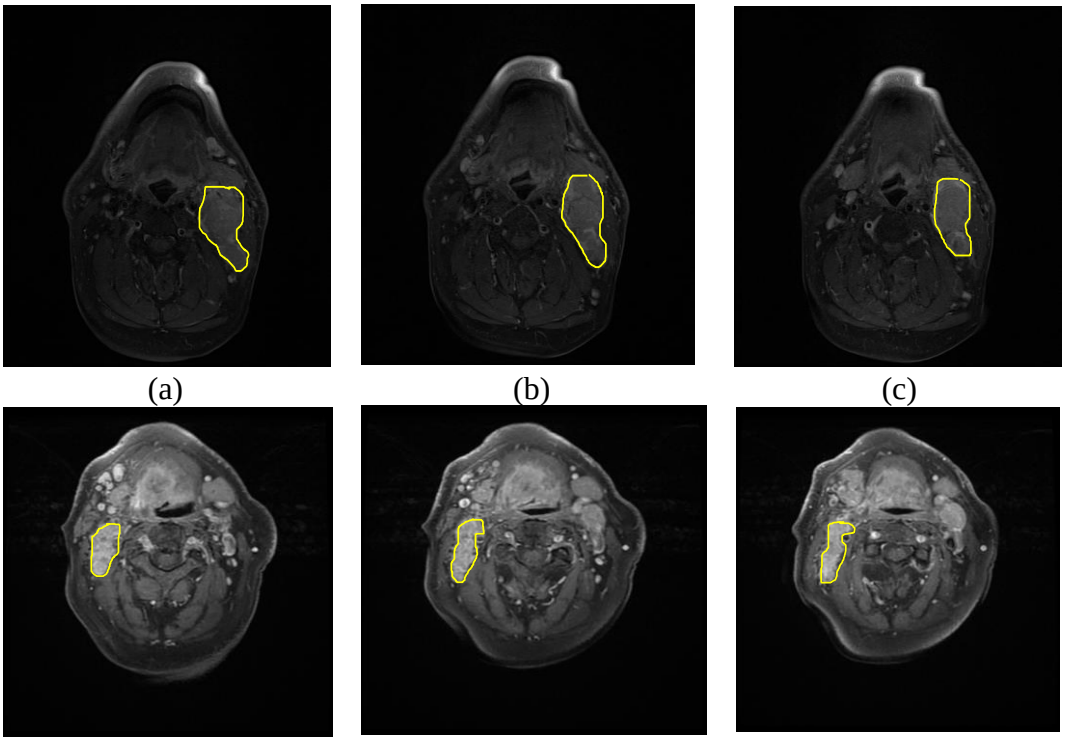


Fig. 4.3 The illustration of MRI scans from four patients (MR14012013, MR12082013, MR06192012, MR27082012), with tumours marked with yellow contours. Each column is scans of same patient, such as (a)(e). The first row (a-d) are slices taken form middle of tumour volume; the second row (e-h) are taken from side (end) of tumour volume.

Fig. 4.3 shows the MRI scans of four patients, with tumours (consensus manual delineations) marked. The tumours have large radius in their centres (middle), but small radius at their ends. Also, from the given four examples, the tumours' shapes are not regular.

In summary, the given examples of HNC slices from Fig. 4.1 to Fig. 4.3 show that the HNCs have some common features on locations, majorly adjacent the throat, and throat tumours can be fully around the throat. The intensities of the tumours are bright,

but there are non-uniform intensities inside tumours. The boundaries of tumours are fuzzy, and there are some neighbour tissues having similar appearance with tumours. The shapes of tumours can be arbitrary.



(a) (b) (c)

(d) (e) (f)

Fig. 4.4 Illustrations of MRI scans from two patients (MR30052012, MR10062013), with cancerous lymph nodes marked with yellow. Each row is a patient's scan such as (a-c).

Fig. 4.4 gives the examples of MRI scans with cancerous HN lymph nodes (abnormal lymph nodes). The ALNs are at two side of throat but not adjacent. The intensities of ALNs are also bright like tumours. Compared to tumours, ALNs have more clear boundaries such as in Fig. 4.4 (d), unless they enlarged too much and infiltrate to neighbour tissues such as Fig. 4.4 (a-c). The ALNs are mostly in balloon and ellipse shapes. More examples of MRI data used in this project will be shown in Appendix B.

4.4 Metrics to measure the segmentation: Dice FP, distance

This thesis is about the auto-segmentation of HNC tumours and abnormal lymph nodes from MRI slices. The algorithm is developed and tested with the guidance of clinician's opinions and consensus manual delineations (labels). Given MRI data and

corresponding labels, the performance of an auto-segmentation algorithm should be measured via the comparison between algorithm output and consensus manual contours. The comparisons between two contours (segmentations) can be quantitatively measured by several metrics, this section will introduce some of them, include confusion matrix, dice score, Jaccard score, and Hausdorff distance.

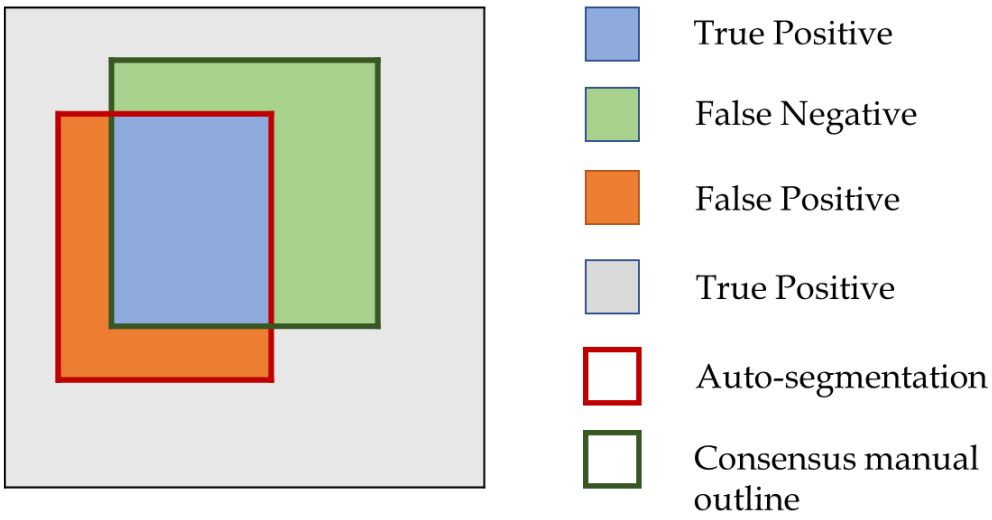


Fig. 4.5 Demonstration of comparison between auto-segmentation and consensus manual contour.

Fig. 4.5 gives an illustration of overlapping algorithm prediction (Red contour) on gold standard label (dark green contour). Several metrics for quantitative measurements of segmentation performance can be explained by this illustration.

The first metrics to introduce is confusion matrix, the confusion matrix can be obtained from the comparison between prediction and label, and in comparison, four circumstances can be concluded by confusion matrix given as the following table,

Table 4.2 Confusion matrix

		Predicted Class	
	Total cases = P + N	Positive (PP)	Negative (PN)
Actual Class	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

As shown in Table 4.2, the actual classes are divided into positive and negative group, the positive corresponds the ground truth of foreground of target in segmentation or detection problem, the negative corresponds to ground truth of background in segmentation and detection. When a part of predicted positive (foreground) fits the actual classes, this part is called as true positive (TP). When a part of predicted positive is negative (background) in actual classes, this part is called as false positive. When a part of predicted negative (background) fits the actual classes, this part is called as true negative. When a part of predicted negative is positive (foreground) in actual classes, this part is called as false negative. So back to Fig. 4.5, given an algorithm output contour (box in red boundary) and consensus manual contour (box in dark green boundary), the TP is the intersection (shown with blue area) between the two predicted area of foreground; the FP is the algorithm mistaking background as foreground (shown with orange are); the FN is the algorithm missing the foreground part in delineation (shown with green area); the TN is the algorithm correctly distinguish the background (shown with Gray in Table 4.1).

The following metrics can be computed from the confusion matrix:

$$\text{Accuracy} = \frac{TP + TN}{P + N} \tag{4.1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 - \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{4.4}$$

where Accuracy is the rate that how algorithm correctly find true positive as well as true negative among entire data, this works well when the sizes of classes (foreground and background in binary case) are symmetric. The Precision is the rate how precisely the algorithm prediction fit true positive, the key to improve precision is to decrease FP. The Recall is the rate how many percent of the foreground object is correctly predicted. The F1-score can be rewritten as,

$$F1 - \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{4.5}$$

$$= \frac{2}{\frac{1}{\frac{TP}{TP + FP}} + \frac{1}{\frac{TP}{TP + FN}}}$$

$$= \frac{2TP}{2TP + FP + FN}$$

The F1-score has same form as another important metrics called Dice Similarity Coefficients (DSCs) [314] or Dice Score, which is given as,

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{4.6}$$

The Dice Score and F1-score consider both true positive and true negative, but compared to Accuracy, the Dice score can handle uneven classes better, here an extreme example is given as,

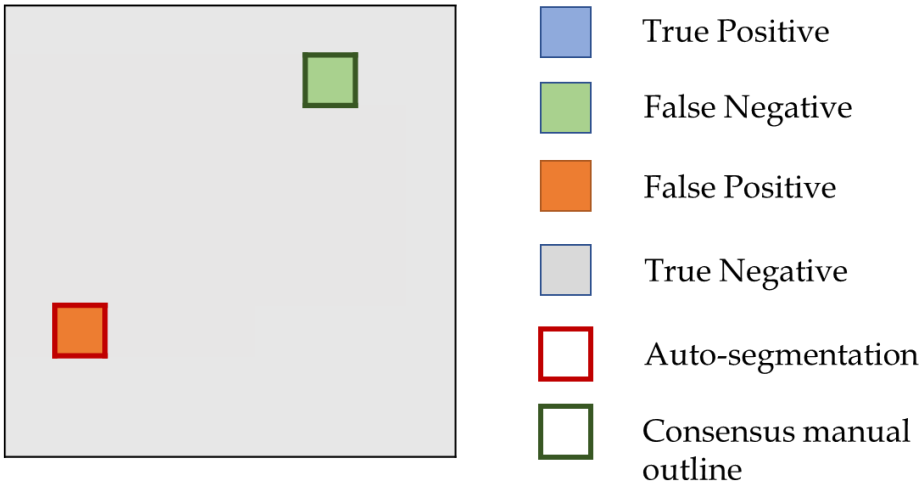


Fig. 4.6 Illustration of a segmentation results on a 9 X 9 image.

Fig. 4.6 shows an example of segmentation on a 9 X 9 image. The predicted contour (red boundary) totally misses the actual object (dark green contour), but by using different metrics it may give different quantitively results. The Accuracy and Dice score can be computed from (4.1) and (4.3) as

$$Accuracy = \frac{TP + TN}{P + N} = \frac{0 + 9\,X\,9 - 2}{9\,X\,9} = 0.975 \tag{4.7}$$

$$Dice\ Score = \frac{2\,X\,0}{2X0 + 1 + 1} = 0$$

From the Eq. 4.7, a totally missed segmentation (Fig. 4.6) can reach 0.975 accuracy. This is because the target image has imbalance classes, where the foreground is only 1 of 81 pixels, other 80 pixels are all background. However, Eq. 4.7 shows that the Dice Score correctly measured the real results of this auto-segmentation. Due to fact that in many segmentation tasks the target (TP) is small compares to entire image, the Dice score is widely used in image segmentation evaluation for quantitatively study of algorithm performance. This thesis also works on reduce the overestimation of tumour segmentation (in Chapter 6), thus FP will be another important measurement.

Apart from metrics that measuring area overlapping rate, another useful metrics for image segmentation is Modified Hausdroff Distance (MHD) [315] which is based on distance measure. The MHD can be used to effectively measure the topological

similarity of two objects. Given two sets of contours $C_A = \{a_1, a_2, a_3\}$ of length M and $C_B = \{b_1, b_2, b_3\}$ of length N shown as following figure,

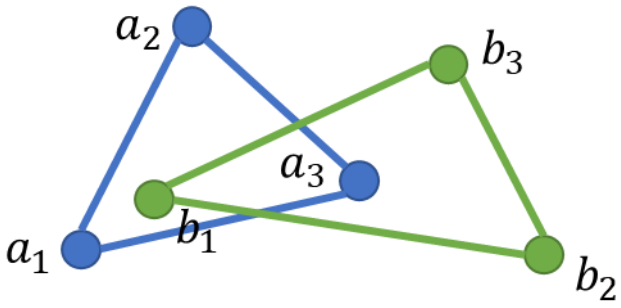


Fig. 4.7 Two contours with three points each.

Fig. 4.7 shows the contours C_A and C_B, so the distance from C_A to C_B is,

$$D(C_A, C_B) = \frac{1}{M} \sum_{c_a \in A} d(c_a, C_B) \tag{4.8}$$

where $d(c_a, C_B)$ is the is minimum distance between point a in set C_A to all points in C_B. M is the number of points in C_A, which is 3 in Fig. 4.7. So similarly, $D(C_B, C_A)$ can be given as,

$$D(C_B, C_A) = \frac{1}{N} \sum_{c_b \in B} d(c_b, C_A) \tag{4.9}$$

where N is the number of points in C_B, and it is 3 in Fig. 4.7. Thus, based on Eq. 4.8 – 4.9, the MHD between C_A and C_B is given as,

$$MHD(D(C_A, C_B), D(C_B, C_A)) = max(D(C_A, C_B), D(C_B, C_A)) \tag{4.10}$$

A good segmentation corresponds to higher dice score, and shorter MH distance.

4.5 Conclusion

This chapter described the MRI dataset used in this thesis and highlighted the frequent challenges of the segmentation of HNCs and ALNs from MRI dataset. The numbers of the MRI slices are limited, and the raw data is in 2D form. The HNCs have no regular shapes, no clear boundaries, and in many cases no uniform intensity distributions. The ALNs have more regular shapes and more clear boundaries compared to HNC, but it can enlarge and infiltrate to neighbour tissue. The development of automatic segmentation algorithms is under the guidance of clinicians' opinions and manual consensus delineations. The following chapters will introduce the developed novel algorithms for auto-segmentation of ALNs and HNCs from real MRI dataset, and the performance of these algorithms will be validated via metrics such as Dice score introduced in this chapter.

Chapter 5 Novel 3D Segmentation Methods for head and neck abnormal lymph nodes from MRI Data

5.1 Introduction

Radiotherapy planning is a complicated and lengthy process requiring detailed defining of complex cancer regions (GTV as prementioned in Section 2.4) as well as affected cancerous lymph nodes, so that decide the PTV to guide RTP. This chapter will present the proposed novel 3D segmentation algorithm for abnormal lymph nodes (ALNs) from MRI data, which will help solve several challenges (subjectiveness [5], time expense) in RTP prementioned in Chapter 2. The challenges of this work also include segmenting tumour as well as abnormal lymph nodes regions with fuzzy boundaries, irregular shapes, non-uniform intensities, and avoiding adjacent anatomical structures. It is essential to determine the intensity range of the target area. also, the size of the initialisation cube for 3D LSM can also impact the result. The proposed work aims to achieve 3D segmentation, quantification, and visualisation of ALN using knowledge-based detection followed by LSM boundary tracking, so that detect accurate location and extract smooth 3D volume of ALNs.

In chapter 3, varieties of algorithms for medical image segmentation have been reviewed. As discussed in Chapter 3, the solution of cancer segmentation should be designed consider available data, computing power, and clinical purpose. Regarding limited numbers of labelled ALN data (introduced in Chapter 4) and the purpose of 3D results, a conventional knowledge-based 3D segmentation algorithm is proposed to extract the volume of ALNs from T1-weighted Gadolinium-enhanced MRI data provided by Beatson West of Scotland Cancer Centre, in Glasgow.

The rest of this chapter will be organized as following: the section 5.2 will review the overall pipeline of the ALN segmentation algorithm; section 5.3 will be the pre-processing steps of input MRI data; section 5.4 will introduce the proposed knowledge-based detection system; section 5.5 will introduce the LSM for volume extraction and post-processing for surface refinement; visual results and conclusion will be in section 5.6 and 5.7.

5.2 System overview

The high-level workflow of the proposed automatic segmentation algorithm is shown in Fig. 5.1.

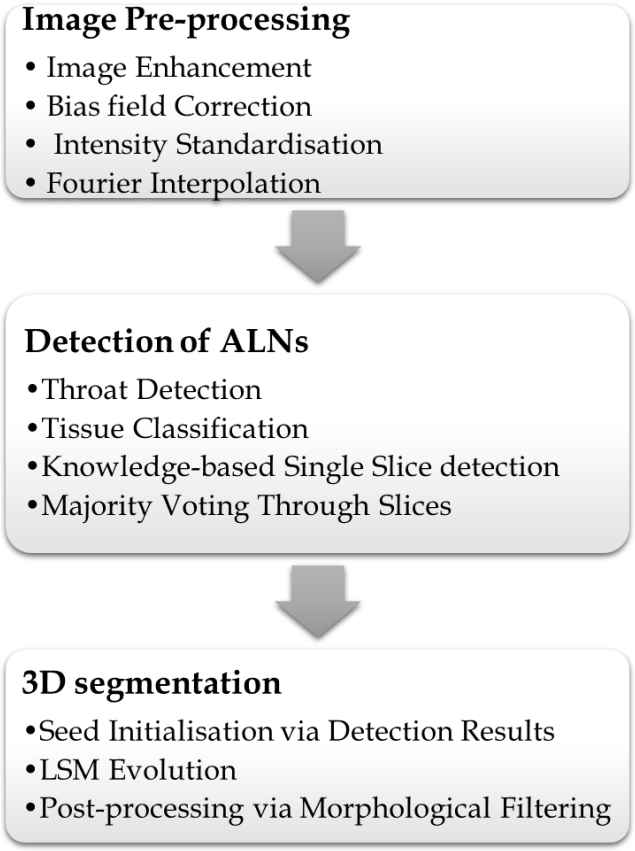


Fig. 5.1 Flow chart of automatic cancerous lymph nodes segmentation [316].

As shown in Fig. 5.1, the proposed ALNs segmentation algorithm consists of two major parts, which is detection (localization) and segmentation (volume extraction). As the conventional image processing methods used in the proposed algorithms majorly use image-level feature, a pre-processing pipeline is designed to improve input MRI data first before later analysis. The details from flowchart in Fig. 5.1 will be described in the following sections.

5.3 Pre-processing of T1 Gd-enhanced head and neck MRI data

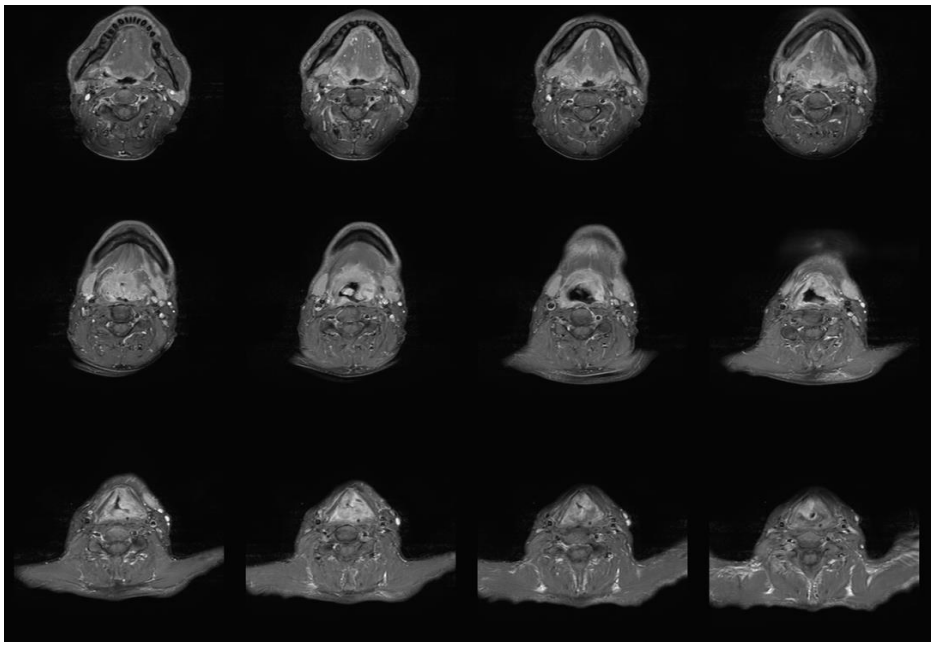
As described in section 2.3, a variety of MRI artefacts occur during data acquisition and signal reconstruction. The artefacts in MRI data lead to difficulties for manual and

automatic segmentation, which includes but not limited to poor image quality, noise, and intensity inhomogeneity

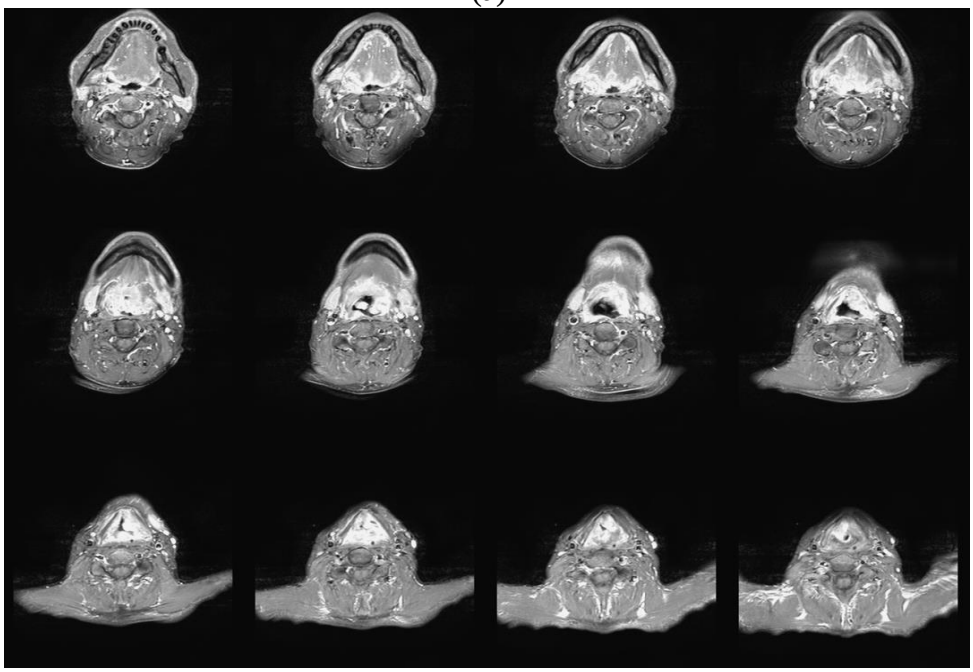
5.3.1 Image enhancement

The first step for image pre-processing in our work is image enhancement. This contains contrast enhancement. Contrast enhancement stretches images' histogram so that intensities of different tissues are more significantly different visually and numerically. The methods of contrast enhancement were introduced Section 3.2.1.1, which is histogram equalization (HE). here CE is applied on our data to improve the quality of images.

As introduced in Section 3.2.1.1, the HE can change the contrast of input image I_{in} by mapping its CDF of histogram $c_{h0}(I_{in})$ to $c_{h1}(T_c(I_{in}))$, where $T_c(I_{in})$ is the transformed image and $c_{h1}(T_c(I_{in}))$ refers to the CDF of transformed histogram, the T_c can be obtained by minimize $|c_{h1}(T_c(I_{in})) - c_{h0}(I_{in})|$ where $c_{h0}(I_{in})$ and $c_{h1}(T_c(I_{in}))$ are known. In addition to that, the transformation must follow several rules: 1) T_c must be monotonic for each pixel in original image, 2) The order of the input intensities is maintained in the output images [52]. Also, when the HE is for improve contrast of image I_{in} itself rather than make it has similar histogram with another image, a flat histogram ($c_{h1}(T(x))$) used as the target image's histogram. The flat histogram means that all the range is used, and all levels are represented by the same amount of pixels [125]. A visual comparison of HNC slices from one patient before and after HE will be given in Fig. 5.2



(a)



(b)

Fig. 5.2 (a) Raw MRI slices from a patient (MR10062011). (b) Contrast enhanced MRI slices from (a).

In Fig. 5.2, we show slices of one patient in (a) and show contrast enhanced slices in (b). After contrast enhancement via HE with flat histogram as target, firstly the overall luminance of slices is improved; then the contrast is more significant so that the different tissues are more visually distinguished.

Fig. 5.3 displays the MRI slices and their histograms of intensity, before and after contrast enhancement, respectively.

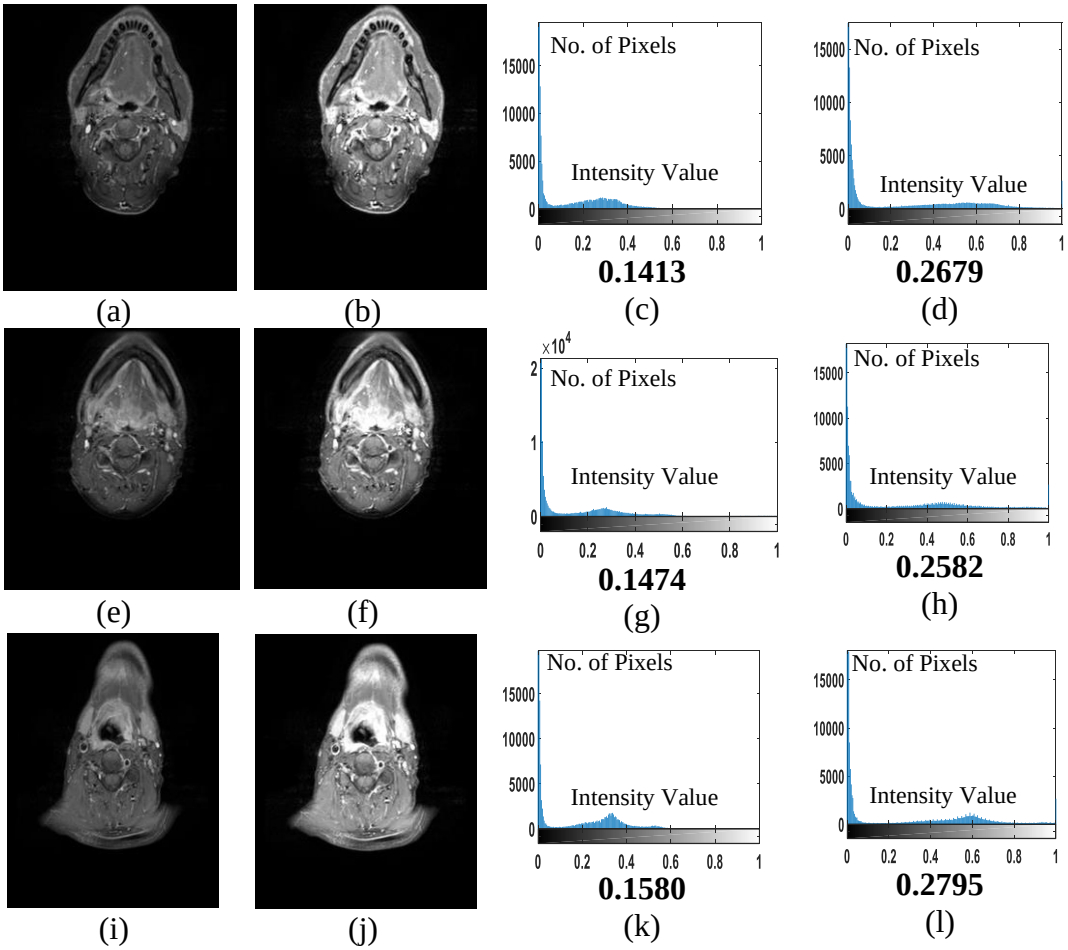


Fig. 5.3 Image (a)(e)(i) shows three slices of MRI data (MR10062011), (b)(f)(j) after contrast enhancement, (c)(g)(k) are histograms of the slices, and (d)(h)(k) histograms after contrast enhancement. The y axis represents numbers of pixels, the x axis represents the intensity value (normalized to [0,1]). The Bold number under each histogram is the C_{RMS} describe the contrast which has been introduced in Eq. 2.2.

It can be seen from Fig. 5.3 (a)(e)(i) that before the contrast enhancement the slices have poor luminance across all tissues. From Fig. 5.3 (c)(g)(k) the histograms prove that before CE the intensity of different tissues is mostly in the lower half. Fig. 5.3 (b)(f)(j) shows that after CE the slices get better luminance, then the histograms in Fig. 5.3 (d)(h)(l) prove that the distribution of intensity are transformed from lower half to wider ranges. The given C_{RMS} also quantitatively shows the improvement of contrast.

From changes of intensity distribution shown in slices, histograms, and C_{RMS}, the CE modified the contrast between tissues to make them easier to distinguish. However, this CE processing only works on discrete values and cumulative histogram distribution, it will enhance noise effect in image in some circumstances.

5.3.2 Bias field in magnetic resonance imaging

As introduced in Section 2.3.3.4, bias field is the background intensity not homogenously distributed in medical images. In this section, it specifically refers to that, in one single slide the same tissue in MR slices has different range of intensity values. The methods for bias field correction have been reviewed in Section 3.2.1.4, the method used in this work is from [193]. Concisely, the initial bias field is estimated by polynomial function with least square approximation. Then the estimated bias field is further refined using bicubic spline model. The parameter of spline model (spline/knot distance) is determined via technique in [317]. The MRI slices before and after bias field correction are shown in Fig. 5.4, corresponding with the estimated bias field.

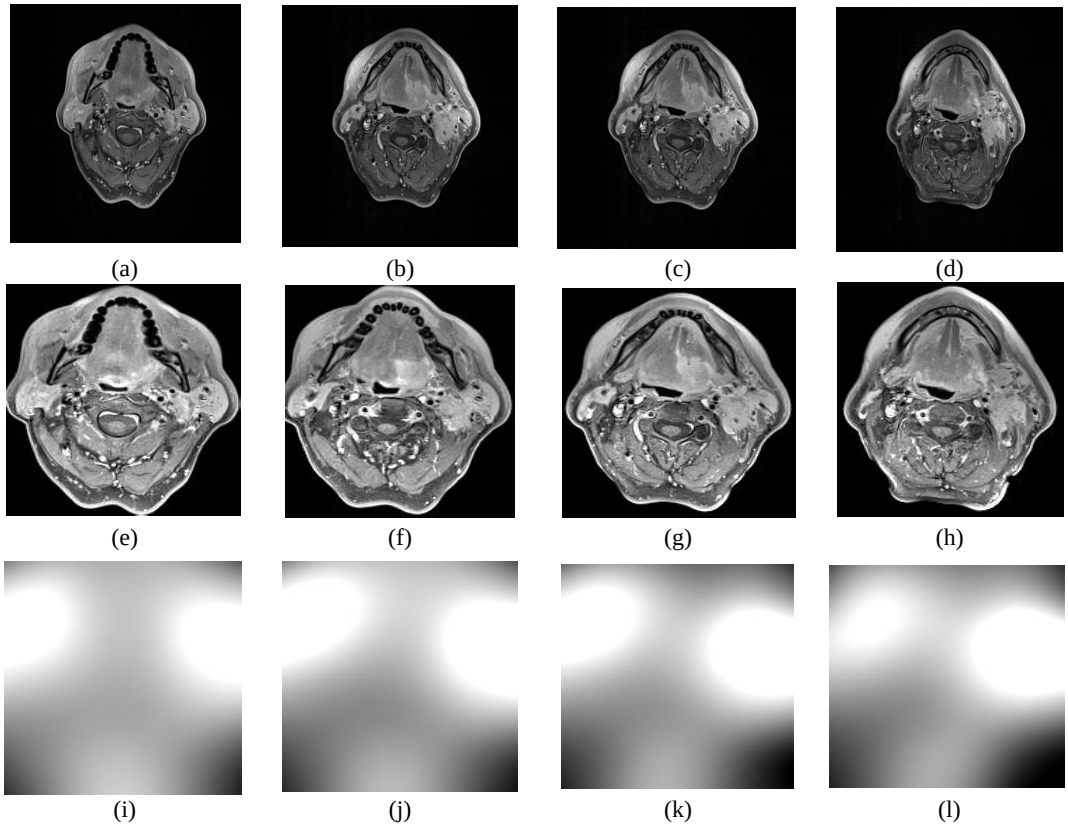


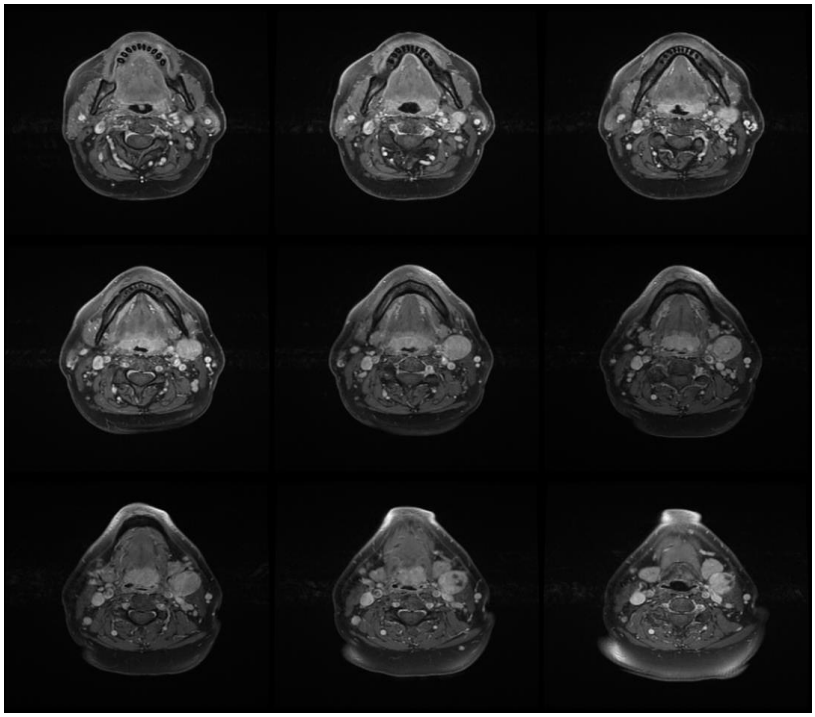
Fig. 5.4 Images (a)-(d) show four raw MRI HNC slices (from MR09082010), (e)(f)(g)(h) after bias field correction, and (i)(j)(k)(l) are estimated bias fields.

Fig. 5.4 (i-l), shows the bias fields (intensity inhomogeneity) in raw MRI HNC slices, the illuminance in left and right top regions inside images (Fig. 5.4 (a-d)) are more significant than other regions inside same images. After the correction of bias fields, the images (Fig. 5.4 (e-h)) have more homogenous intensity across regions in one same image. The bias fields are corrected by methods described in [193]. After the correction of bias fields, the images (Fig. 5.4 (e-h)) have more homogenous intensity across regions in one same image.

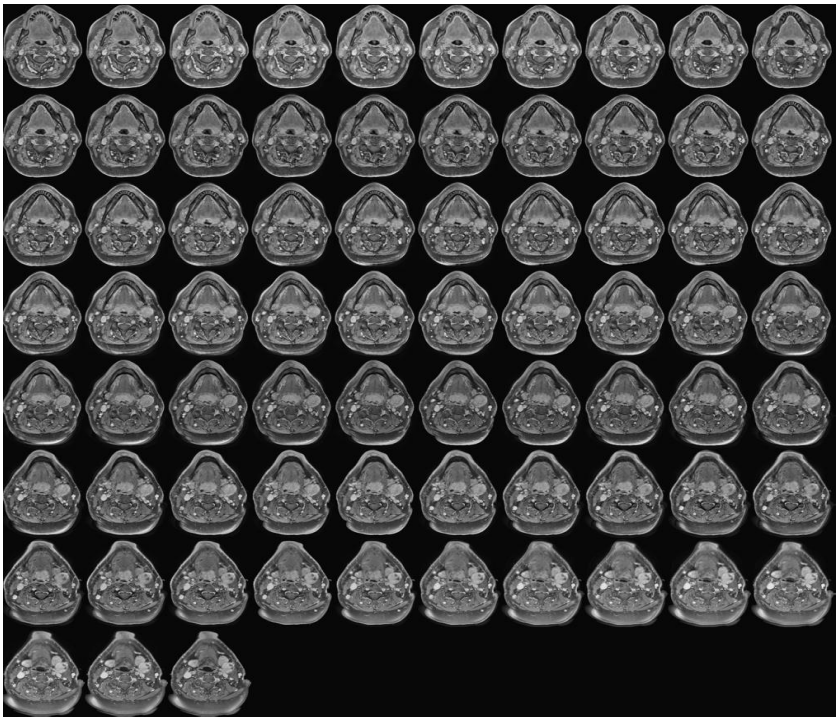
5.3.3 Fourier interpolation

After pre-processing, the quality of each raw HNC MRI slices is improved. On the other hand, the data we used has anisotropic voxels whose resolution are 0.47x 0.47x3.5mm^3, and here are about 4 mm distance between slices. To apply voxel-based 3D segmentation algorithm on MRI data, the anisotropic voxels should be converted to isotropic voxels first, thus we could obtain reconstructed 3D MRI volume from

separate 2D MRI slices. As introduced in 3.2.1.6, the Fourier interpolation is computational economical, besides does not involve inherent blurring and distorted edges. This section demonstrates the Fourier interpolation (FI) of pre-processed T1 MRI slices, the method is from [181]. Briefly, the interpolation on image domain is done by append zeros in the centre of spatial domain.



(a)



(b)

Fig. 5.5 Images (a) shows the raw HNC MRI slices and counts of slices from two patients (MR14012013), (b) shows interpolated slices.

As shown in Fig. 5.5 (a)(c), the two sets of original obtained MRI data have 9 and 8 slices, then as shown in Fig. 5.5 (b)(d) here are more than 70 slices after interpolation, the interpolated slices has isotropic voxels around 0.5x0.5x0.5 mm^3 and thus MRI volumes are built for potential 3D work.

5.4 Knowledge based 2D detection of throat, tumour, and abnormal lymph nodes

After the pre-processing module, the raw 2D MRI data is reconstructed as a 3D HN volume with improved image quality. The workflow is shown in Fig. 5.1, based on the reconstructed 3D volume data, we proposed and implemented methods for detection and segmentation of throat and abnormal lymph nodes. In this section the algorithm of detecting abnormal lymph nodes will be introduced. Apart from pre-processing of HNC MRI data, this section will display the steps of throat detection, clustering of HN tissues, and finally the detection of abnormal LN.

5.4.1 Throat detection using fuzzy probability map

As prementioned, the HNC is mostly around throat region. This anatomical information is very helpful for clinicians to quickly find HNC among complex anatomical structures in HN MRI scans. Thus, motivated by this idea, our algorithm also takes throat region as important landmark for locating of region of interest. In this work, throat regions are detected using fuzzy rule based methods from [193], which has been mentioned in Section in 3.2. This section will remind the details of the detection process.

The fuzzy rules for detection of throat region are built on intensity and spatial information. In terms of intensity, the throat regions have dark appearance (low intensity value around 0). In terms of spatial location, throat regions are mostly horizontally in the middle of one MRI slice, and vertically in the top half of slice.

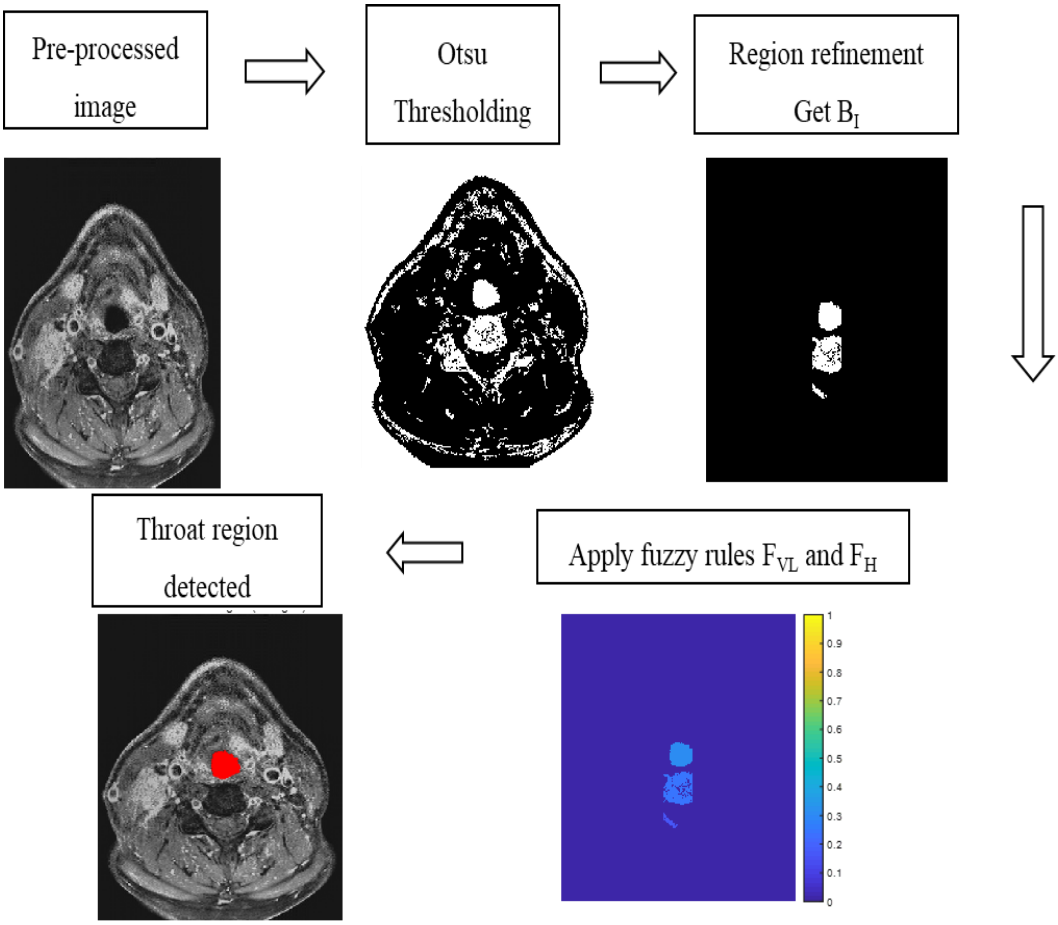


Fig. 5.6 Workflow of fuzzy rule-based throat region detection from pre-processed MRI slice

As shown in the workflow in Fig 5.6, firstly the dark regions which include throat regions are roughly picked using the Otsu's thresholding method [318]. Briefly, given an intensity threshold T to globally categorize pixels in an image into foreground I_f and background I_b, the Otsu's method is to find a T that maximize the inter-class variance between I_f and I_b.

Then, by defining signatures of a binary object as distances from pixels on an object's boundary to object's centre, large objects (with biggest signature over 30 based on an empirical setting) are removed from the top half of the image. Then, as the throat is located around only horizontally middle part of image, the binary objects outside middle area are also removed, and the middle is defined by weighted centre of image.

This is termed the refined binary regions B_I. Finally, the fuzzy rules are horizontally and vertically designed and applied on B_I.

To detect throat region, first spatial fuzzy rule is that if pixels are closer to mid (horizontally), they have higher possibilities to be throat region. Thus, given centre column c_y, around this vertical line fuzzy rules are built as,

$$F_{VL} = \{(i,j), \mu_{VL}(i,j)\} \qquad i = 1 \dots r_n, j = 1 \dots c_n \tag{5.1}$$

with

$$\mu_{VL}(i,j) = f_{VL}(j) = \begin{cases} 1, & |c_y - j| \le a \\ 1 - \dfrac{|(c_y - j) - a|}{b - a}, & |c_y - j| > a \delta\delta |c_y - j| \le b \\ 0, & Otherwise \end{cases} \tag{5.2}$$

where r_n represents number of rows, c_n represents number of columns, a and b are parameters we empirically set to 5 and 20 respectively in practice. The Eq. (5.2) indicates that throat is highest likely in the region (with $\pm a$ columns at sides) around centre column c_y, and not likely in the region two far away from c_y (further than $\pm b$ columns at two sides). The pixels close (less than a) to c_y has highest membership value 1; then the membership values of pixels are decreasing by the distance (between a and b) from pixels to c_y; if distances between pixels and c_y are over b, the membership values will be zero.

Another spatial fuzzy rule is that, if some pixels are closer to first row, they are more likely to be in throat region. Thus, the second rule can be described as:

$$F_H = \{(i,j), \mu_H(i,j)\} \qquad i = 1 \dots r_n, j = 1 \dots c_n \tag{5.3}$$

with

$$\mu_h(i,j) = f_H(i) = \begin{cases} 1, & i == c \\ \dfrac{|d - i|}{|d - c|}, & i > c \delta\delta i \le d \\ 0, & Otherwise \end{cases} \tag{5.4}$$

where c is the first nonzero row of B_I, and membership values are set to 1 for those pixels at row c, and then keep decreasing from row c to row d until 0. Here d is set as r_n.

Two fuzzy rules are combined by 'min' operator, so that only regions have high membership value in horizontal and vertical rules are selected, the combination given as,

$$F_{VLH} = F_{VL} \cap F_H \tag{5.5}$$

$$\mu_{F_{VLH}} = \min\big(w_1 * \mu_{VL}(i,j), (1 - w_1) * \mu_H(i,j)\big) \tag{5.6}$$

$$i = 1 \dots r_n, j = 1 \dots c_n$$

where w_1 and w_2 controls the importance of horizontal and vertical fuzzy rules. And they are set as 0.3 and 0.7 in this work. So finally, candidate pixels for the throat region in B_I are detected as,

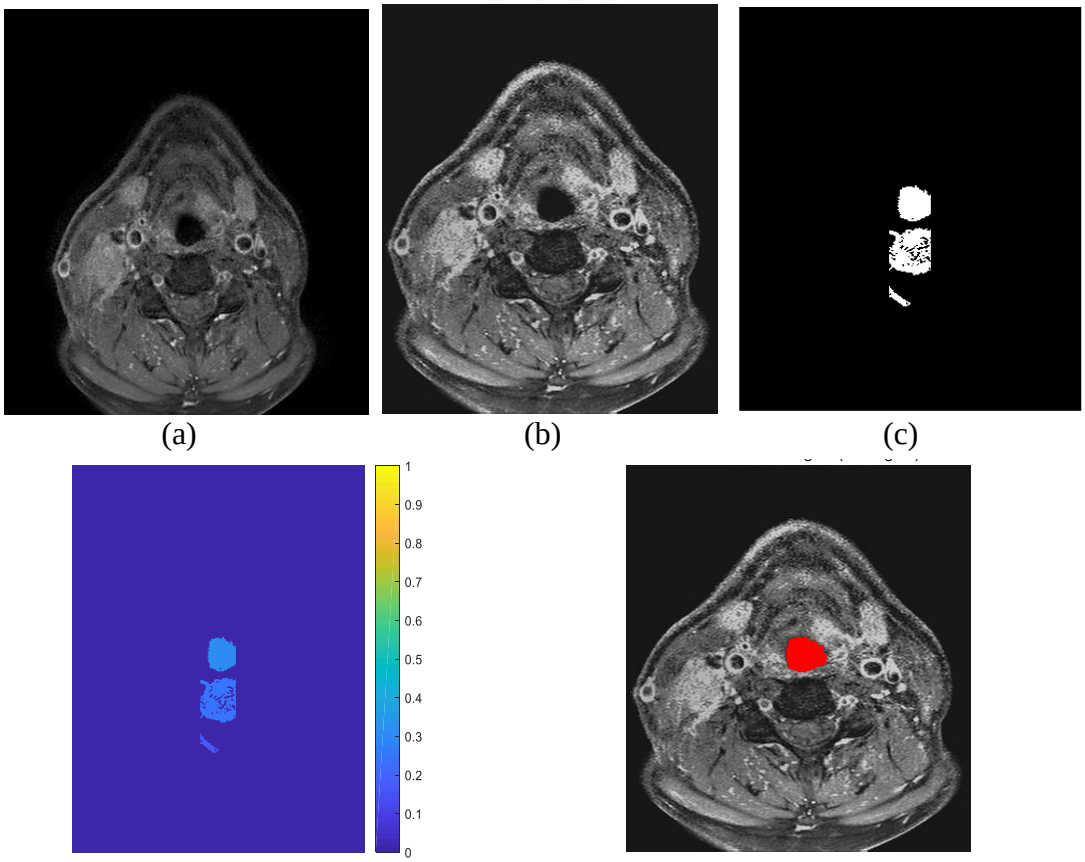
$$F_{intersect} = F_{VLH} \cap B_I \tag{5.7}$$

with

$$\mu_{F_{intersect}}(i,j) = \min\left(\mu_{VLH}(i,j), B_I(i,j)\right) \tag{5.8}$$

$$i = 1 \dots r_n, j = 1 \dots c_n$$

$F_{interest}$ gives the rules of throat detection, and so that pixels in B_I will be assigned membership values which represent their possibilities of locating in throat region. The detection will be region-wised rather than pixel-wised, so that inside each region, pixels' membership values are taken max operation so that a value is obtained to represent the possibility of a region to be throat.



(a) (b) (c)

(d) (e)

Fig. 5.7 (a)(b) shows an original and pre-processed MRI slice (from MR28082014), (c) is the selected binary image B_I, (d) is the possibilities of each region in B_I to be throat region, and (e) shows the detected throat region on slice which is marked as red.

Fig. 5.7 shows the workflow in Fig. 5.6 is applied on an MRI slice, throat region of an MRI slice Fig. 5.7 (a) is detected and shown in Fig. 5.7 (d).

The throat detection and segmentation yield 0.87 mean Dice score, with minimum of 0.81 and maximum of 0.94, which makes it reliable for further usage [55, 181]. More examples of the throat detection are given in Fig. 5.8

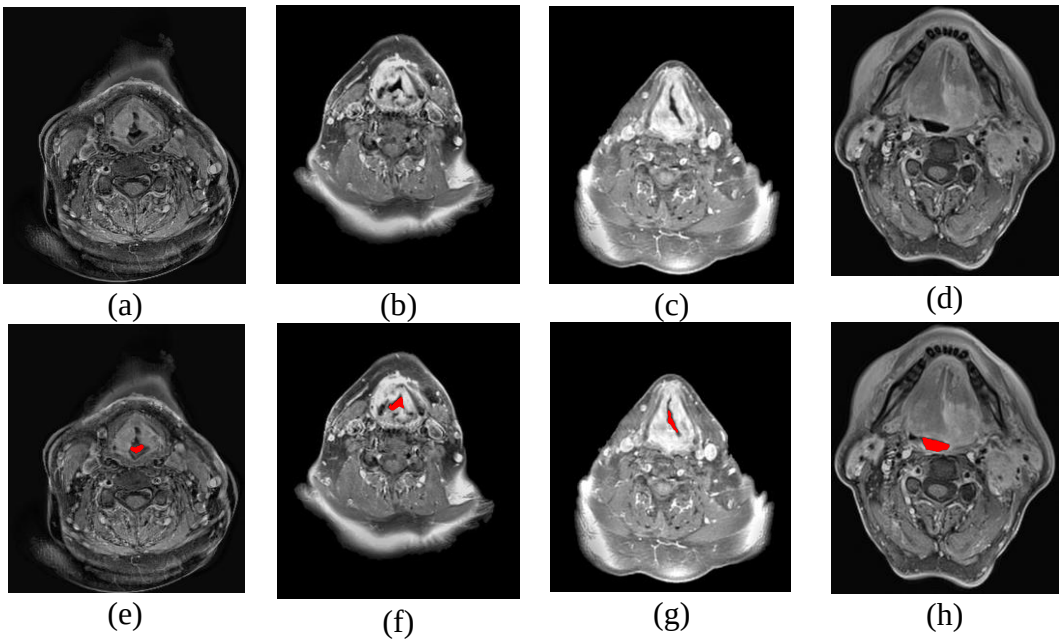


Fig. 5.8 Throat detection examples. (a-d) Input MRI slices (MR09112012, MR19062012, MR27082012 MR09082010) (e-h) Input MRI slices with detected throats marked with red.

5.4.2 Modified fuzzy c-mean head and neck tissue classification

As described in Section 2.2.2, the head and neck tissues can be classified into four categories: fatty tissues, cancer tissues, normal tissue, and normal muscle tissue. This section introduces the application of modified fuzzy c-mean (MFCM) on HNC MRI data for rough tumour detection. As T1 MRI data is grayscale, a standard fuzzy c-mean can define an objective function based on pixels' intensity value, then each pixel will be assigned a membership value and updated along with the optimization of objective function. The objective function is given as,

$$J_T = \sum_{i=1}^{N_c} \sum_{k=1}^{N} \mu_{ik}^{m_f} \, d^2(v_i, I_k) \tag{5.9}$$

where v_i is the cluster centre of centre of cluster i, and N_c represents total numbers of clusters.I_k represents the intensity of k^{th} pixel in image I. Then, $\mu_{ik} \in [0,1]$ is the membership degree of k^{th} pixel belonging to i^{th} cluster. The value m_f (larger than 1, which is 2 in this work) controls the fuzziness of membership function $\mu_{ik}^{m_f} \, d^2(v_i, I_k)$,

which is calculated as the Euclidean distance from value of cluster centre v_i to pixel intensity I_k. The membership functions are constrained by following rules:

$$\sum_{k=1}^{N} \mu_{ik} > 0 \; \forall i \; \in \{1, \dots, N_c\}; \sum_{i=1}^{N_c} \mu_{ik} = 1 \; \forall k \; \in \{1, \dots, N\} \tag{5.10}$$

Then based on the throat information we have acquired in last step, a spatial membership function can be added into objective function for modification, given Euclidean distance of between pixels in a region to throat centre as,

$$d_s^2(i,j) = \sum_{j=1}^{c_n} \sum_{i=1}^{r_n} ((i - t_r)^2 + (j - {t_c}^2)) \tag{5.11}$$

where i is i^{th} row out of r_n rows, and j is j^{th} column out of c_n columns, t_r and t_c represents the row number and column number of throat centre. Then $d_s^2(i,j)$ of all pixels are normalized and so that a spatial-based feature vector is constructed for clustering. Then, the spatial membership function is,

$$J_S = \sum_{i=1}^{N_c} \sum_{k=1}^{N} \mu_{ik}^{m_f} \, d_s^2(i,k) \tag{5.12}$$

In Eq. (5.12) i represents the i^{th} cluster out of C clusters, and k represents spatial feature of the k^{th} pixel in image. Then, $\mu_{ik}^m d_s^2(i,k)$ is calculated as the sum of Euclidean distances between pixels' spatial feature to cluster centres. J_S is also constrained by (5.10). Which means after clustering, no cluster will be empty, and membership value of a pixel to all clusters will be summed as 1. Finally, the objective function of the modified fuzzy c-mean is,

$$J_{ST} = \sum_{i=1}^{N_c} \sum_{k=1}^{N} \mu_{ik}^{m_f} \, d_s^2(v_i, I_k) + \sum_{i=1}^{N_c} \sum_{k=1}^{N} \mu_{ik}^{m_f} \, d_s^2(i,k) \tag{5.13}$$

The pixels are clustered by minimizing the objective function in Eq. (5.13). The optimization is by iteratively updating membership function μ_{ik} and cluster centres v_i as follows:

$$\mu_{ik} = \frac{(d(v_i,I_k)+d_s(i,k))^{-2/(m_f-1)}}{\sum_{j=1}^{N_c}(d(v_i,I_k)+d_s(i,k))^{-2/(m_f-1)}} \text{ and } v_i = \frac{\sum_{k=1}^{N} \mu_{ik}^{m_f} I_k}{\sum_{k=1}^{N} \mu_{ik}^{m_f}} \quad (5.14)$$

through the MFCM, the pre-processed MRI images will be divided into five classes as shown,

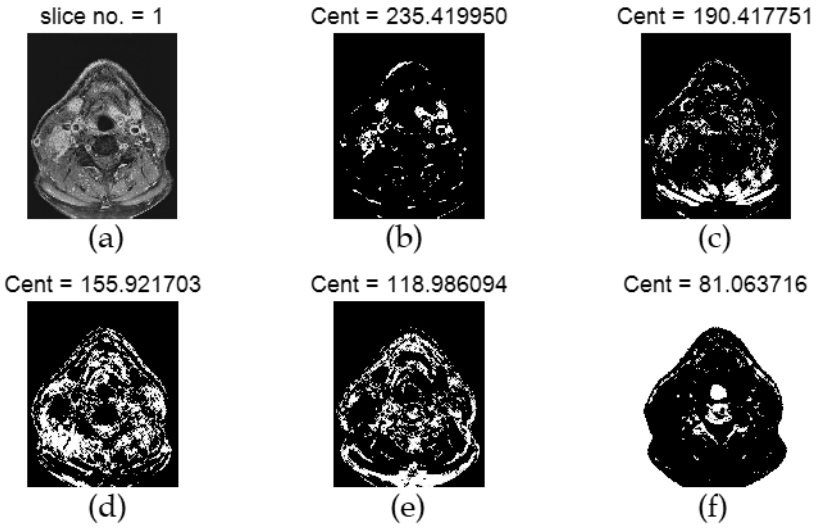


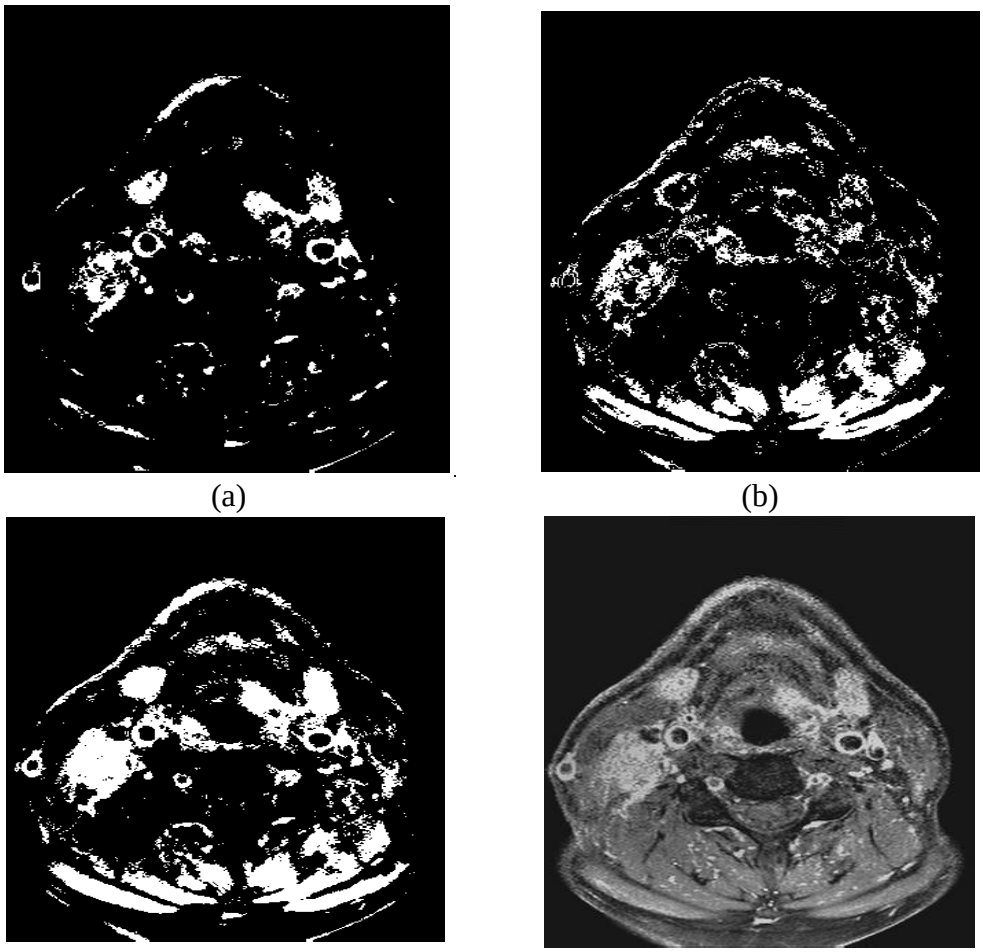
Fig. 5.9 Modified fuzzy c-mean clusters pixels in an MRI slice (first slice of MR28082014) into five categories, left top image show the slice, and rest five images show five clusters with their cluster centres (mean intensity value of each cluster, in this case around 235, 190, 156, 119, and 81 in the range of [0, 255]).

As shown in Fig. 5.9, low intensity regions such as throat and bones are clustered into the same class with lowest cluster centre (Fig. 5.9 (f)), and high intensity regions such as Edema and body fat are clustered into another same class with highest cluster centre (Fig. 5.9 (b)). The tumours and Edema are mostly distributed in clusters with first and second highest cluster centre, these two clusters will be taken as ROI for further processing. Thus, classification of head and neck tissues are done, this step helps us obtain rough ROI and get rid of non-related regions for HNC segmentation. The next section will introduce the further operations of segmentation.

5.4.3 Knowledge-based identification of abnormal lymph nodes in 2D MRI slice

After the previous MFCM, we obtain an image with rough binary ROI (denoted as $IB = \{Ri \mid i = 1, \ldots, N\}$) where the tumour and abnormal lymph nodes exist. In order to get accurate segmentation, more refinements required for identifying location of tumour and abnormal lymph nodes. This section will introduce the knowledge-based method for refinement of rough ROI, the knowledges are from clinical experts' experience, which includes some general rules about localization of HNC cancer in medical scans.

This section we will also present the new localization method of abnormal lymph nodes. Considering the intensity distribution, ALNs have similar or higher intensity values compared to tumours. Based on the classification from Fig 5.9, the pixels of ALNs belong to clusters of pixels with first and second highest cluster centres in intensity values. So, firstly, related clusters are merged as shown in following Fig. 5.10,



(a) (b)

(c) (d)

Fig. 5.10 (a) are clustered regions with highest centre (out of 5), (b) are regions with second highest centre, (c) are (a) and (b) merged (d) is the original image used for clustering (MR28082014).

where Fig. 5.10 (a) are regions with highest clustering centre (intensity value around 235, also shown in Fig. 5.9 (b)), Fig. 5.10 (b) are regions with second highest clustering centre (intensity value around 190, also shown in Fig. 5.9 (c)), and Fig. 5.10 (c) are merged from Fig. 5.10 (a) and Fig. 5.10 (b). The merging is finished by bitwise OR operation.

Based on the selected regions in Fig. 5.10 (c), we propose a knowledge-based probability map for region refinements and ALN detection. The map is given as,

$$W = \alpha W_{location} + \beta W_{size} + \theta W_{eccentricity} \quad (5.15)$$

where W is the probability map for ALN detection, ALNs will have highest W values in this map. The map comprises $W_{location}$ (probability deduced by region location), W_{size} (probability deduced by region size), and $W_{eccentricity}$ (probability deduced by eccentricity, i.e., shapes). Parameters α, β, and θ control the weights between each probability, and they fit the condition that $\alpha + \beta + \theta = 1$.

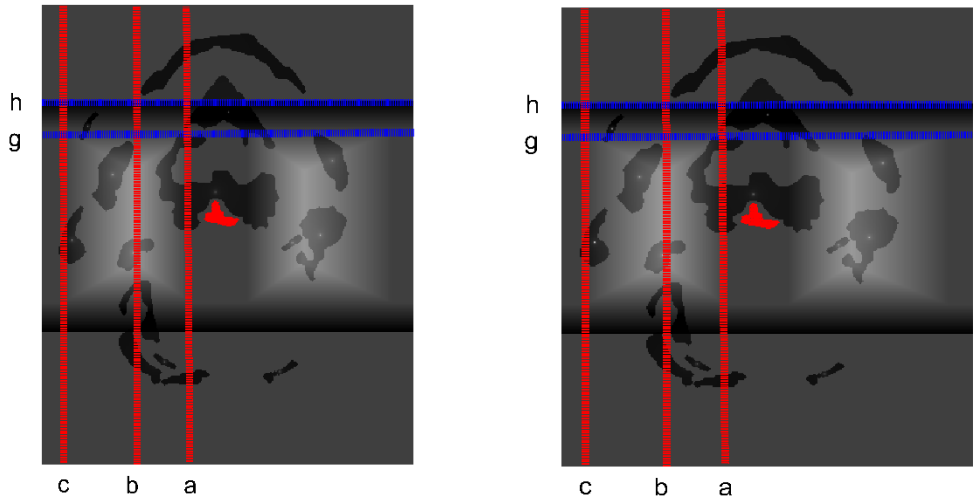
The value of $W_{location}$ is set according to the prior knowledge that the head and neck lymph nodes are located at two sides of the throat region but not closely adjacent to the throat. The construction of $W_{location}$ maybe written as:

$$W_{location} = w_c \cap w_r \tag{5.16}$$

$$w_c = \begin{cases} 0, & |C_n - y| < a \\ \dfrac{|C_n - y - a|}{|a - b|}, & a < |C_n - y| < b \\ 1 - \dfrac{|C_n - y - b|}{|c - b|}, & b < |C_n - y| < c \\ 0, & else \end{cases} \tag{5.17}$$

$$w_r = \begin{cases} 1 - m\dfrac{R_n - x}{g}, & |R_n - x| < g \\ (1 - m) - \dfrac{|R_n - x - g|}{|h - g|}, & g < |R_n - x| < h \\ 0, & else \end{cases} \tag{5.18}$$

where R_n and C_n are row and column coordinates of pixels in probability map. The $W_{location}$ consists of spatial information from columns (w_c) and rows (w_r). The parameters x and y are coordinates of the throat centre (shown in Fig. 5.7 (d)). Intersection symbol in Eq. (5.16) denotes fuzzy AND operator. Parameters $a - c$, g and h are adaptive thresholds based on distance away from throat centre. The parameter m is a control coefficient which ranges from 0 to 1. Two examples of the probability maps are given as,



(a) (b)

Fig. 5.11 Probability maps with spatial thresholds. Red lines with texts show the horizontal thresholds a-c, and blue lines with texts show the vertical thresholds g-h. Red regions in (a) (b) indicates the throat regions. Here only display one side of a-h on single slice, a-h are symmetry on the other side of same slice [319].

As shown in Fig. 5.11, the spatial thresholds $a-h$ define the slices into several sub-areas, and in sub-areas the probabilities of existence of ALNs are given in Eq. 5.19. The values of $a-h, m$ are adaptively set based on throat region's size. For example, if on an MRI slice the throat height is less than 0.1 of image height, a will be set as 0.8 of throat height; of throat height is over 0.2 of image height, a will be set as 0.5 of throat height. The selected empirical setting is given as,

$$a = \begin{cases} c_{t\_x} \pm 0.8 * w_t, & if\ w_t \in [0, 0.1 * c_n) \\ c_{t\_x} \pm 0.6 * w_t, & if\ w_t \in [0.1 * c_n, 0.2 * c_n) \\ c_{t\_x} \pm 0.5 * w_t, & if\ w_t \in [0.2 * c_n, c_n) \end{cases}$$

$$b = \begin{cases} c_{t_x} \pm 1.5 * w_t, & if\ w_t \in [0, 0.1 * c_n) \\ c_{t\_x} \pm 1.5 * w_t, & if\ w_t \in [0.1 * c_n, 0.2 * c_n) \\ c_{t_{t\_x}} \pm 1.5 * w_t, & if\ w_t \in [0.2 * c_n, c_n) \end{cases}$$

$$c = \begin{cases} 0.6 * w_t \parallel c_n - 0.6 * w_t, & if\ w_t \in [0, 0.1 * c_n) \\ 0.4 * w_t \parallel c_n - 0.4 * w_t, & if\ w_t \in [0.1 * c_n, 0.2 * c_n) \\ 0.3 * w_t \parallel c_n - 0.3 * w_t, & if\ w_t \in [0.2 * c_n, c_n) \end{cases} \quad (5.19)$$

$$g = \begin{cases} c_{t\_y} \pm 0.8 * h_t, & if\ h_t \in [0, 0.1 * r_n) \\ c_{t\_y} \pm 0.6 * h_{t,} & if\ h_t \in [0.1 * r_n, 0.2 * r_n) \\ c_{t\_y} \pm 0.4 * h_{t,} & if\ h_t \in [0.2 * r_n, r_n) \end{cases}$$

$$h = c_{t\_y} \pm 0.4 * r_n$$

where c_n and r_n represent total number of columns and rows of an MRI slice, c_{t_x} and c_{t_y} represents centre x and y coordinates of detected throat region on this slice, and w_t and h_t are width and height of throat region. The throat region's height and width compared to an entire scan can be seen as a marker of the pattern, such as the scan is from near throat, or near tongue base. The ALNs have different spatial relationships to throat in these different patterns.

The second component W_{size} is accurately decreasing the probability value obtained by large object/regions in detection. If a candidate region/object is not ALN but very large, it can accumulate large probabilities when it is overlapped with probability map $w_{location}$ only. Here a probability map component is given as,

$$w_{size} = \sum_{i=1}^{n} \frac{1}{\sqrt{(x_i - x_c) + (y_i - y_c)} + eps} \quad (5.20)$$

where (x_i, y_i) is location of i^{th} point in a binary region R_i, (x_c, y_c) are location of centre point of R_i and eps is a small positive number (e.g., machine epsilon) to ensure non-division by zero. This makes R_i's w_{size} mostly depend on its centre point, and points away from centre of R_i will have smaller w_{size}. In this way, large false positives will be less likely occurred in the detection process.

The Eccentricity is calculated as,

$$W_{eccentricity} = \frac{C_e}{A_e} \tag{5.21}$$

where C_e is the distance from the region centre to the focus along the major axis of ellipse, and A_e is the length of major axis. If R_i has $W_{eccentricity}$ close to 0, it is likely to be a circle; if it is close to 1, the region is likely to be elongated.

The Construction of probability map from Eq. (5.15) to Eq. (5.21) are also shown in Fig. 5.12.

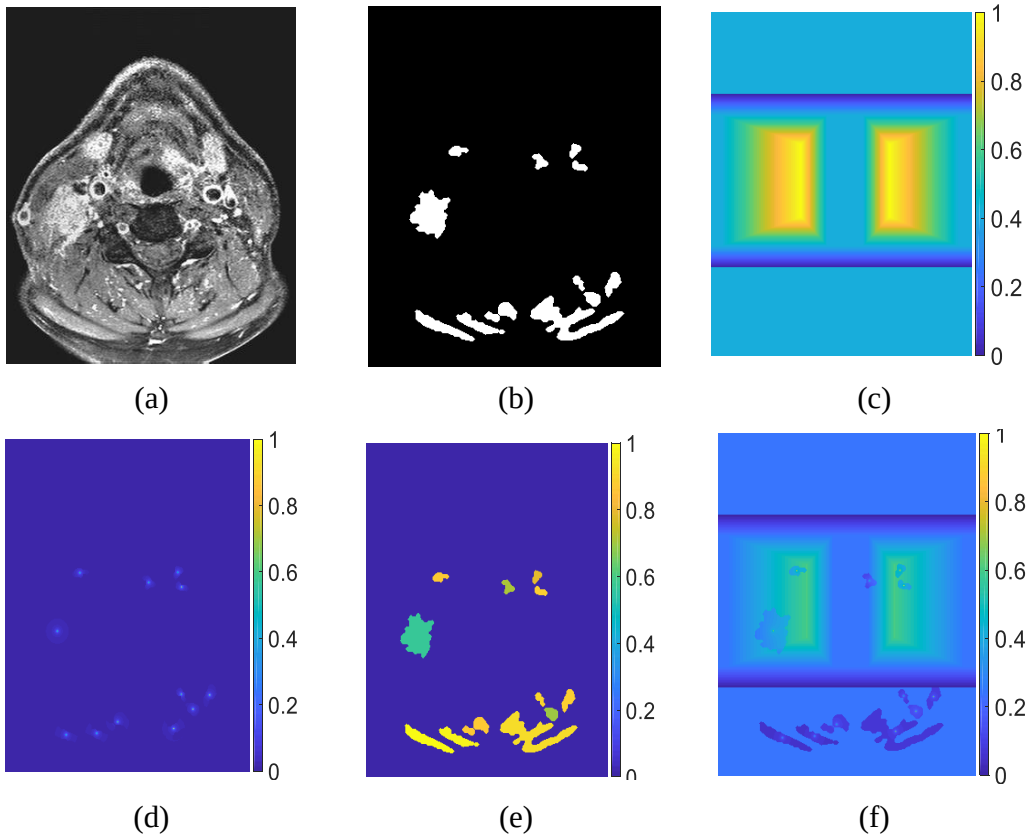


Fig. 5.12 Construction of Probability map. (a) Input MRI slice (MR28082014) (b) Rough ROIs on I_B (c) Probability map distribution deduced from $W_{location}$ (d) Probability map distribution deduced from $W_{location}$ (e) Probability map distribution deduced from W_{size} (f) Probability map distribution deduced from $W_{eccentricity}$ (f) Final probability map combines all factors. The colour bar on right side of each map represents the probability values.

In Fig. 5.12, the construction of probability map for an input MRI slice (Fig. 5.12 (a)) is demonstrated. Given the rough ROI image I_B (Fig. 5.12 (b)), probability map W for detection of ALN is built according to $W_{location}$ (Fig. 5.12 (c)), W_{size} (Fig. 5.12 (d)), and$W_{eccentricity}$ (Fig. 5.12 (e)).

Then, the detection of ALNs is done by accumulating the probability values inside each candidate ROI in I_B, and results can be gained as indicated in Fig. 5.13.

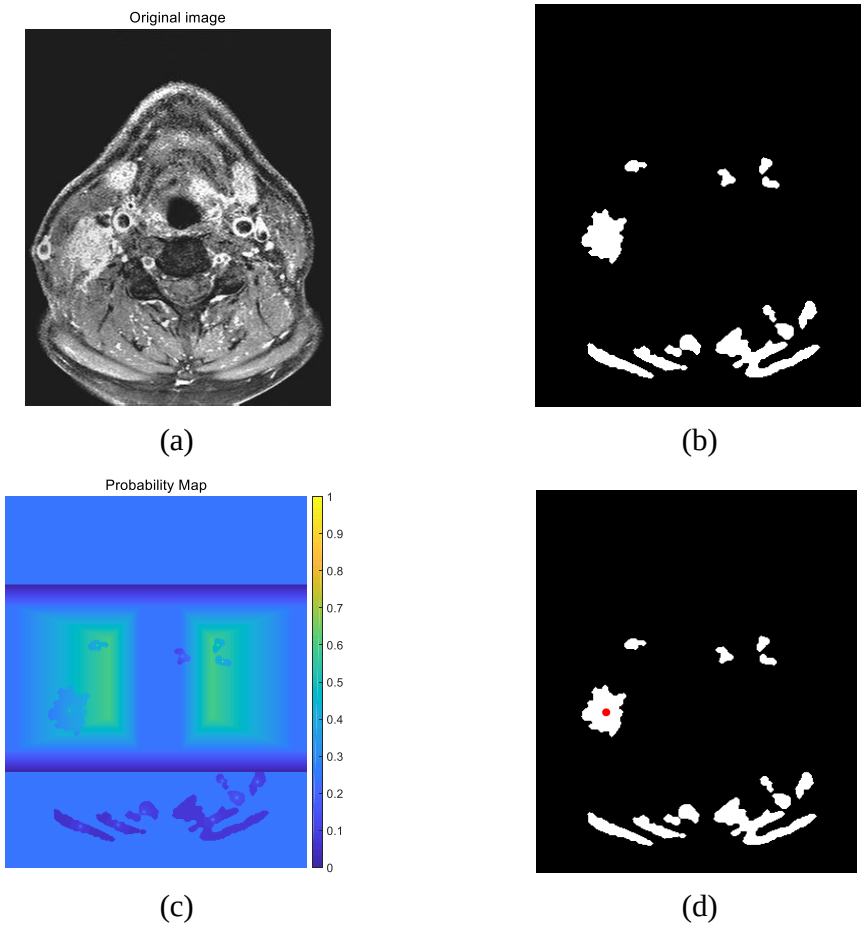


Fig. 5.13 Detection of abnormal lymph nodes using probability map *W*. (a) An original MRI slice (MR28082014). (b) Rough ROIs in I_B. (c) Map *W* applied on I_B. (d) Binary regions in I_B with detected ALN (marked with red point).

As shown in Fig. 5.13 (c), the probability map *W* gives probability values to each point of ROIs. By accumulating the probabilities of each ROI, a region who has the highest sum probabilities should be the target ALN, which is marked red in Fig. 5.13 (d).

5.4.4 3D localization of ALNs from 2D detections

The detection of ALNs is applied on single MRI slice. This section will introduce further localization for 3D detection and segmentation. An algorithm is proposed to remove false positives from 2D detection, so that keep reliable 2D detection considering the inter-slice information. From these remained multiple 2D detections, 3D coordinates of ALN can be calculated. This algorithm is illustrated in Fig. 5.14.

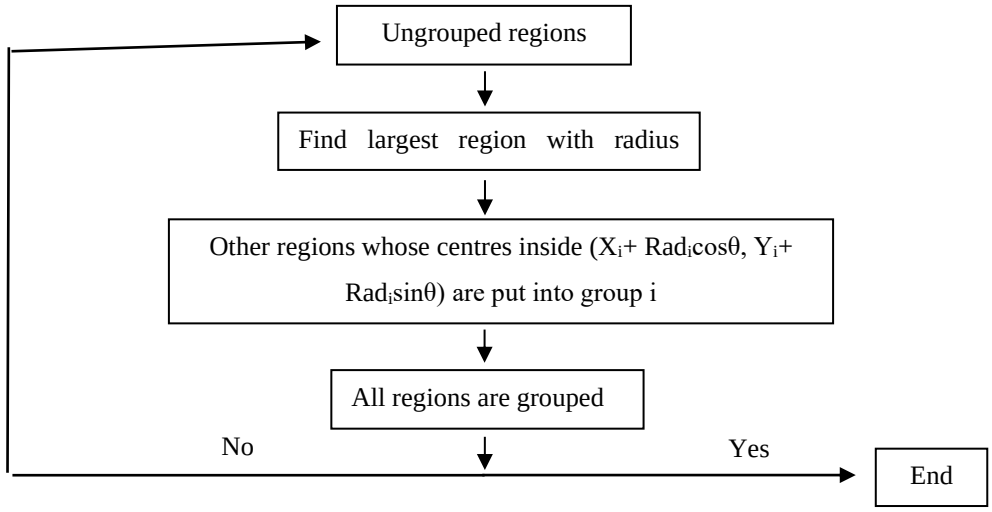


Fig. 5.14 Workflow for grouping 2D detections

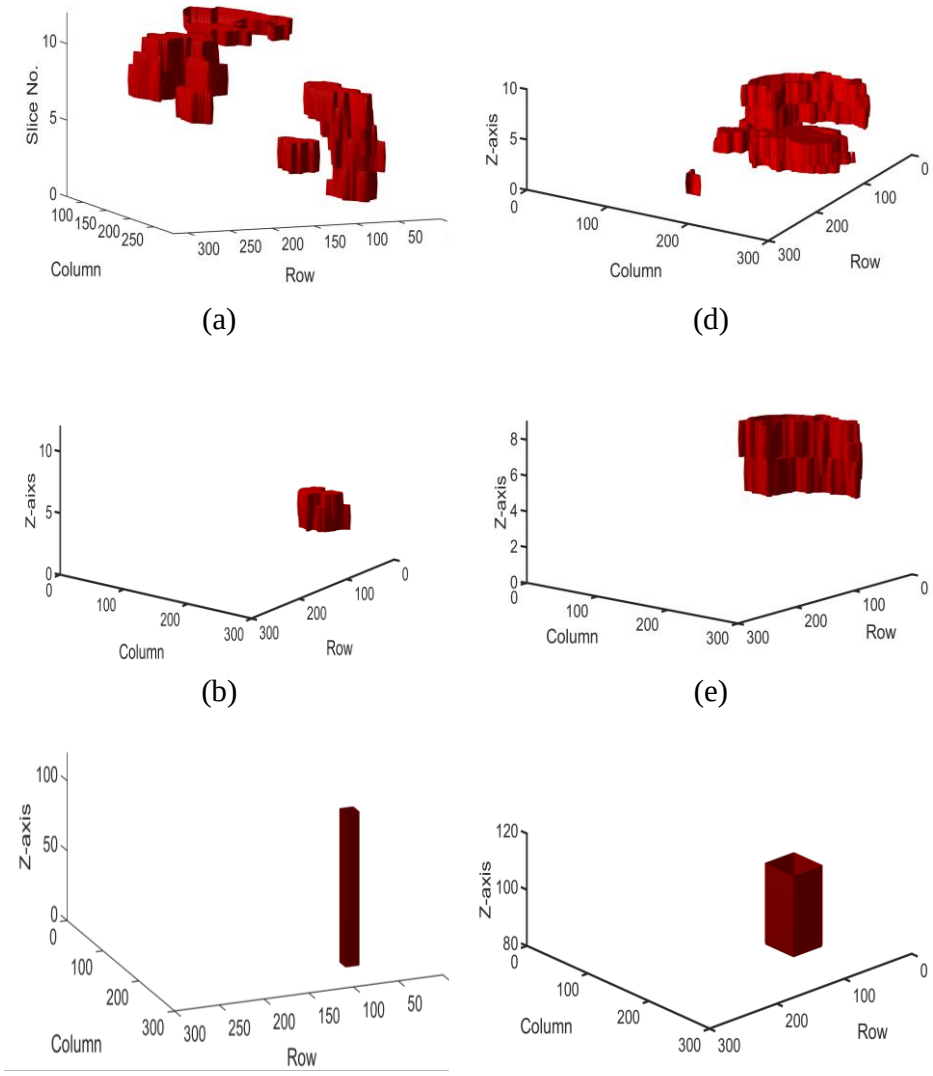
As shown in Fig. 5.14, firstly the 2D detections ((Fig. 5.13 (d))) are sorted based on their radii Rad. Then, the region with largest radii Rad_i and centre (X_{c_i}, Y_{c_i}) is selected; other regions whose centres is inside $(X_{c\_i} + Rad_i cos\theta, Y_{c\_i} + Rad_i sin\theta)$ are assigned to the same group *i*. Then process is repeated on the rest of regions (i.e., ungrouped regions), until all regions are grouped. The sorting and searching are performed slice by slice. By completing this process, all 2D detections are clustered into groups. Group G_i with majority of N regions R is kept, and regions in other groups are discarded. And The lymph node centre can be calculated by taking the average of the regions' coordinates in the persevered group. Thus, 2D ALN detections are refined and located in 3D space. The calculation of horizontal centre of ALN is described as,

$$C_{g\_i} = \ mean\left(\sum_{j=1}^{N}\left(C_j(x), C_j(y)\right)\right) \tag{5.22}$$

where,

$$C_j \in G_i, j = 1 \ldots \ldots N$$

An example of the 3D localization is shown in Fig. 5.15.



(a) (d)

(b) (e)

(c) (f)

Fig. 5.15 Two examples of grouping of 2D detections. (a)(d) All 2D ALNs (red parts) detections shown in 3D space. (b)(e) Refined (grouped) 2D ALNs (c)(f) Located position of ALN (red part) in interpolated 3D space.

In Fig. 5.15 (a) the 2D detections are displayed in 3D axis, and among them false detections are existed. Through the majority voting grouping process in Fig. 5.15, the false detections are eliminated, thus the correct 2D detections of ALN are determined as shown in Fig. 5.15 (b). Then an initial seed (Fig. 5.15 (c)) for 3D LSM is built based on 2D ALNs. The horizontal centre of this seed is from average of centres of 2D ALNs, and the height of this seed equals to the height of interpolated MRI volume.

After the refinement of 2D detections (Fig. 5.15 (a)), a 3D location (Fig. 5.15 (b)) of ALN volume is obtained for starting a segmentation process.

5.5 Automatic lymph nodes contouring using active contours

5.5.1 Automatic definition of 3D LSM function

Taking the previously determined 3D location of ALN, this section will use 3D LSM to segment 3D volume of ALN. As introduced in Chapter 3, the LSM function required several settings to evolve, generally including start seed, speed function, and stop criterion.

The start seed has been given in last section and shown in Fig. 5.15 (b). The start seed is a 3D cube. The cube's horizontal centre is calculated from Eq. (5.22), its length and width are 5% of length and width of entire MRI volume, and its height equals to MRI volume's height. This setting ensure that the cube (start seed) is horizontally small enough to be inside the target (ALN), and vertically go through the MRI volume to ensure all slices be considered for 3D reconstruction.

In this work, the speed function F is from [320],

$$\begin{aligned} F &= \lambda(\varepsilon_{LN} - |I(x,y,z) - \mu_{LN}|) + (1-\lambda)\nabla.\frac{\nabla\emptyset}{|\nabla\emptyset|} \\ &= F_{ext} + F_{int} \end{aligned} \tag{5.23}$$

where $\nabla.\frac{\nabla\emptyset}{|\nabla\emptyset|}$ is the average curvature of evolving curve; μ_{LN} is mean intensity of detected lymph nodes regions;ε_{LN} is standard deviation of all pixels inside the detected lymph nodes region, and λ is a weighting factor $0< \lambda < 1$. The right-hand side of Eq.

(5.23) comprises and external force F_{ext} which drives the curve to the boundary and an internal force F_{int} that keeps segmentation result smooth. This level set function is based on intensities of pixels, and curvature of evolving curve.

Here the defined evolution of level set function will stop in two cases: first, after it evolves more than certain times; or second, the changes of curve are smaller than certain numbers of voxels.

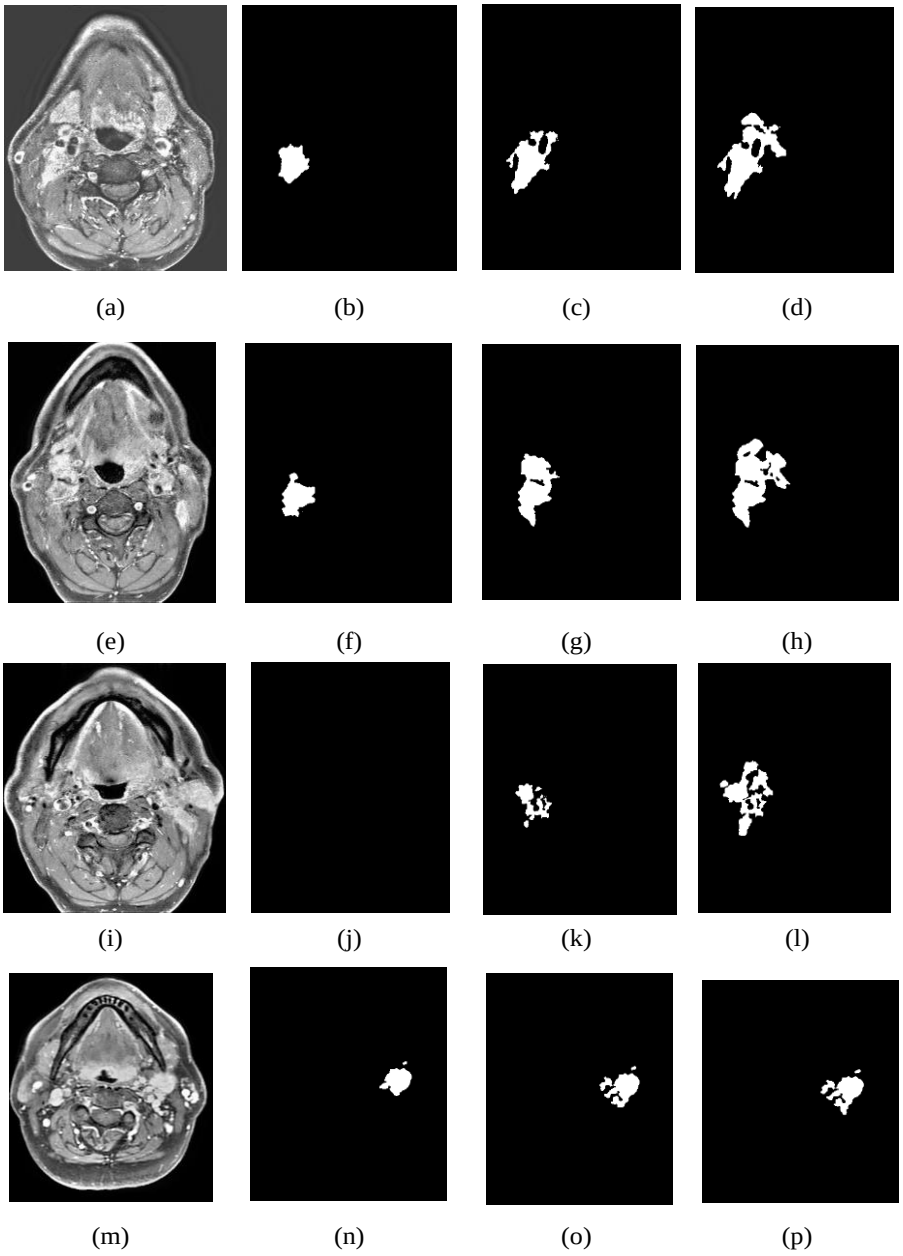


Fig. 5.16 The evolution of LS function in MRI volume. (a)(e)(i) are three interpolated MRI slices at top, middle and bottom of same MRI volume (MR28082014). (b-d) are sliced volume ALN segmentation of top position changed following iterations. (f-h) at middle, and (j-l) at bottom. (m) is MRI slice from other interpolated MRI volume (MR14012013), (n-p) show their evolution.

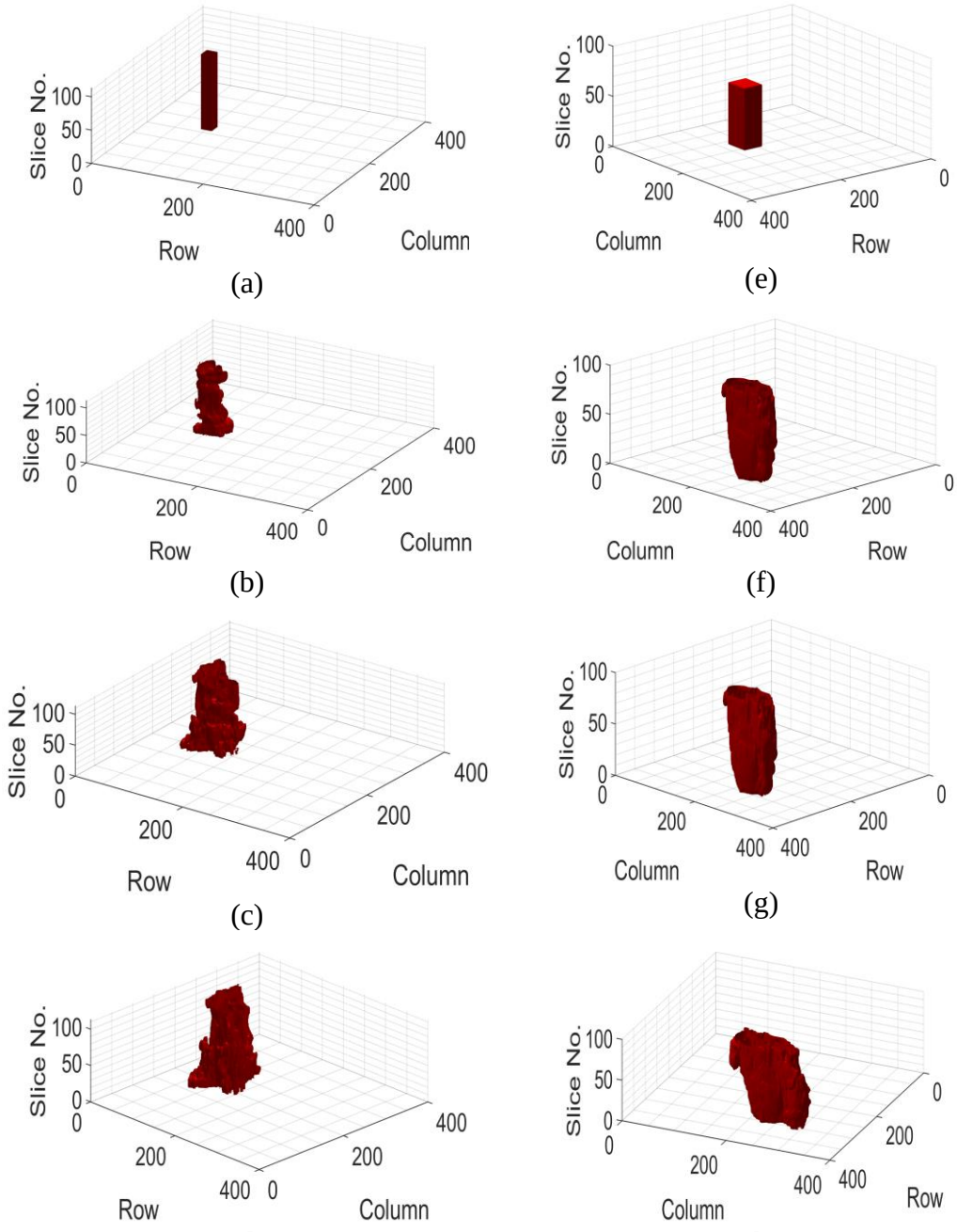


Fig. 5.17 (a-d) Evolution from initial seed to final 3D volume of ALN from MR09082010. (e-h) Another example (MR11062013) of this process.

As shown in

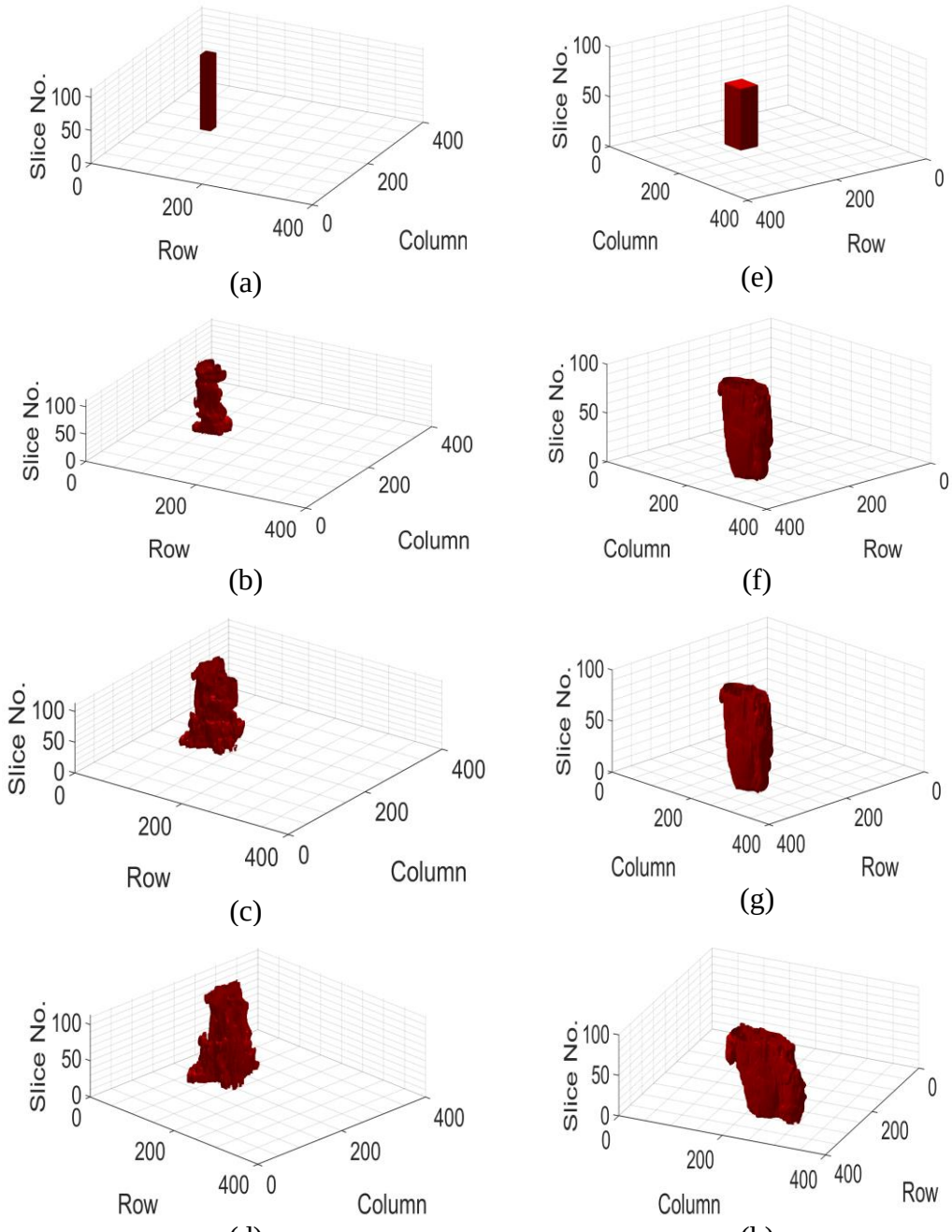
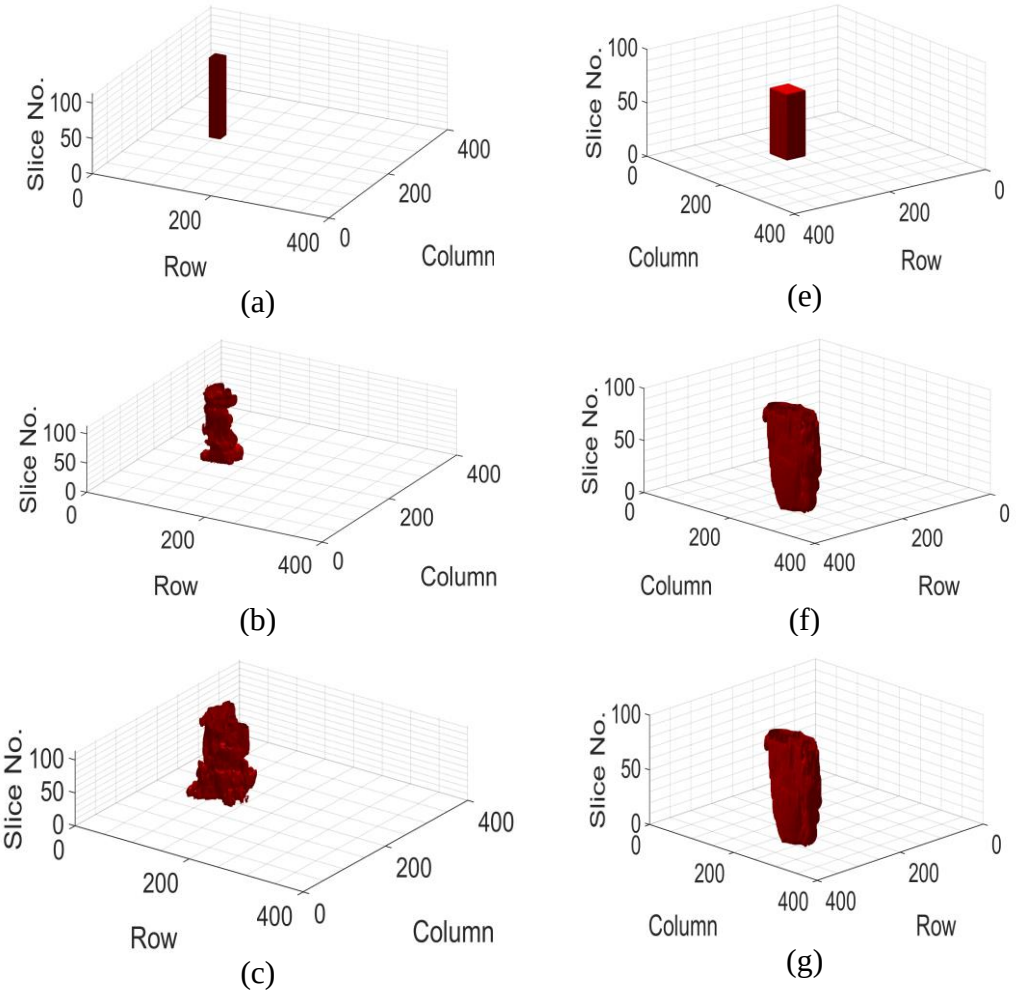


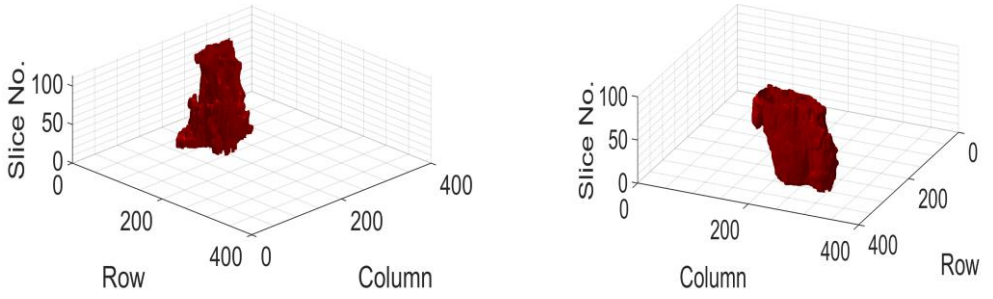
Fig. 5.17 (a-d) and (e-i), the 3D start seed evolves by iterations in 3D space, and finally got a volume for 3D ALN. From the left to right perspective in Fig. 5.16 (a-l), the 3D ALN volume is evolving in horizontal direction; and from top to bottom perspective, the 3D volume is also evolving vertically. Thus, it shows that the defined LSM function has ability to track boundary and surface of 3D object in MRI volume, so that with good detection as start seed, 3D volume of ALN is segmented.

5.5.2 Post processing of extracted 3D ALN volume

The post processing steps are necessary in this work due to two main reasons: 1) As introduced in chapter 3, the level set method based curve evolution may have boundary leakage at weak boundaries, where intensity changes (gradient) are subtle; 2) In HN MRI data, there are still some tissues having similar intensities with ALN (as shown in Fig. 5.10 (c)), such as tumours, body fat, and salivary glands. So, this section will introduce the morphological operation as post processing for refining 3D volume obtained by 3D LSM.

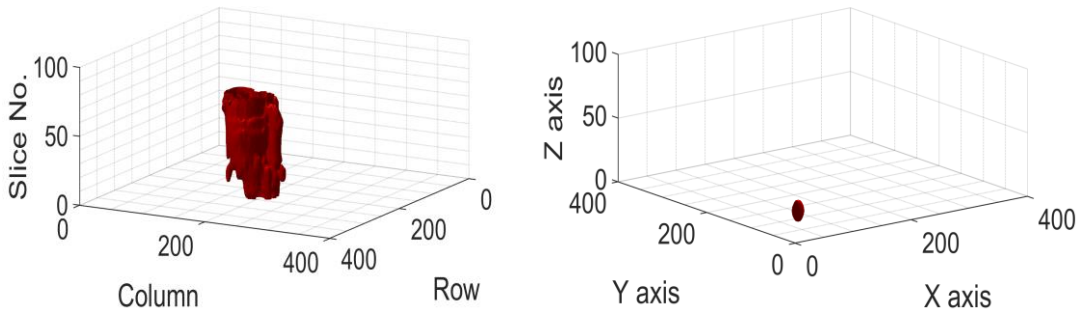
We design a post processing procedure with several steps. Firstly, a 3D erosion operation (with a sphere structure element of radius in 8 pixels) separates the obtained 3D ALN (





(d) (h)

Fig. 5.17 (d)(i)) into smaller objects, which means some narrow connections are cut; then, all separated 3D objects are sorted and only the largest one is kept. Finally, a dilation operation (a same sphere structure element of radius in 8 pixels) is applied on the kept object, so that its volume is recovered to the size before erosion, and at the same time all neighbour small objects are removed.



(a) (b)

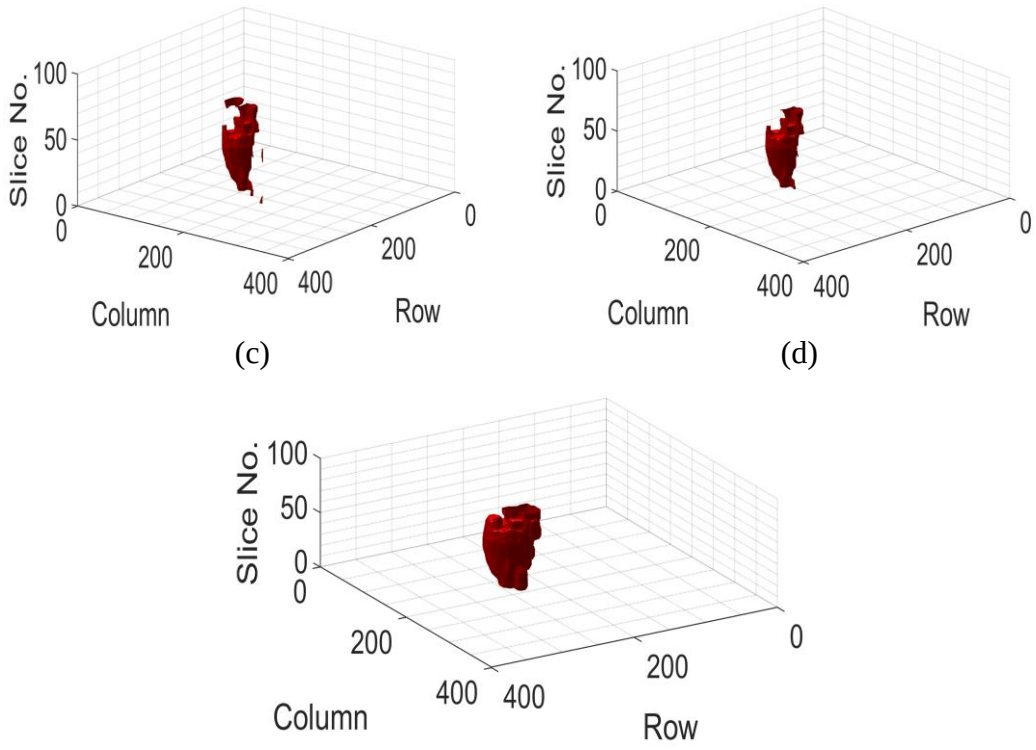


Fig. 5.18 Post processing of 3D LSM segmentation of MR14012013. (a) ALN volume obtained by 3D LSM (Red), (b) Circular 3D structure element used (radius is 8 pixels), (c) Erosion applied, (d) Only largest object preserved, (e) Dilation applied, final segmentation obtained.

As introduced in Section 3.2.1.3, the morphological operations are capable to topologically separate 2D connected objects or merge discrete objects. When extended to 3D, Fig. 5.18 (c) shows that 3D erosion breaks weak connections inside 3D objects so that separates them. Then, we pick the largest object among separated objects (Fig. 5.18 (d)), so that the major component of 3D LSM is preserved, and leakage parts of LSM evolution are eliminated. Finally, a 3D dilation is applied on preserved part, so that the size of this part is recovered (Fig. 5.18 (e)). The recovered 3D object is the final 3D segmentation result of ALNs in HN MRI data.

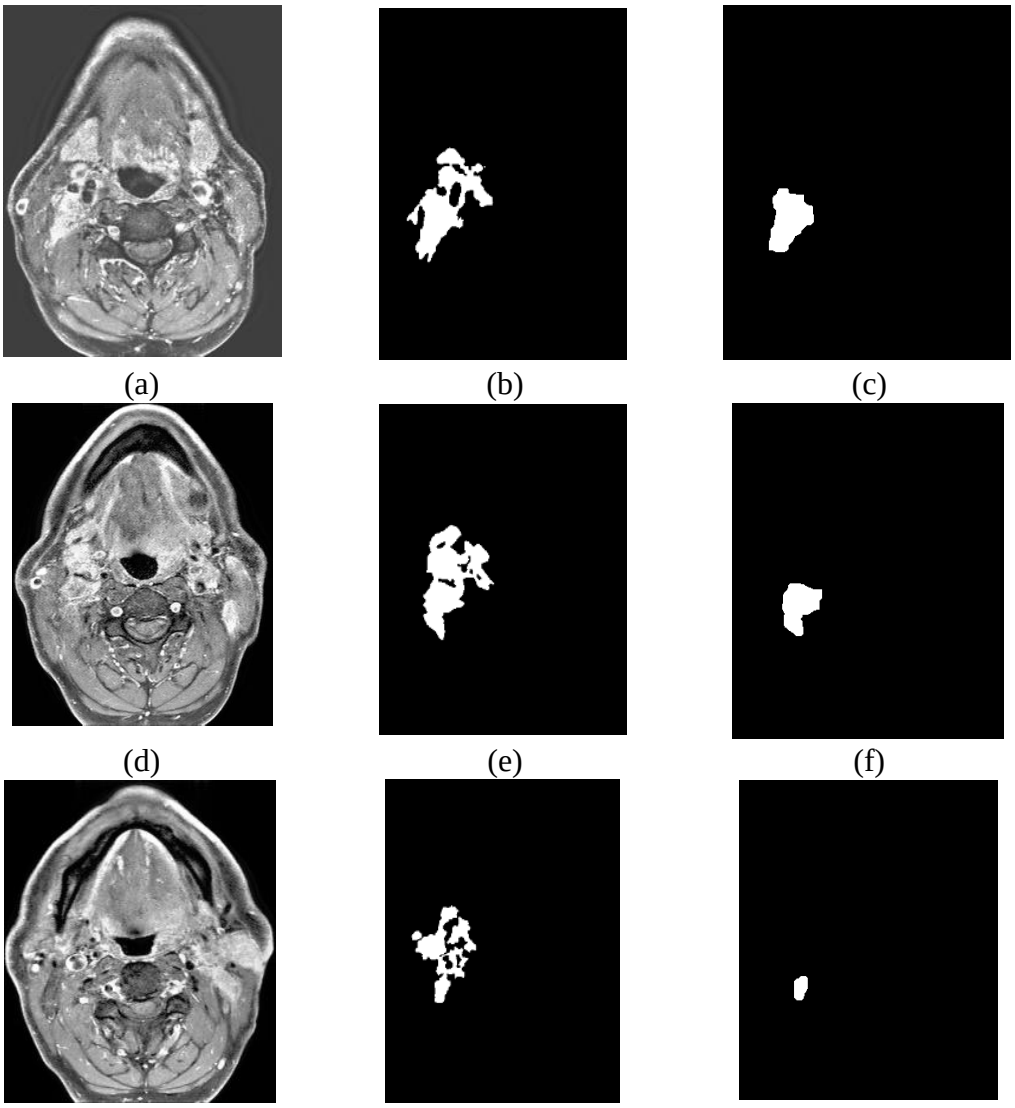
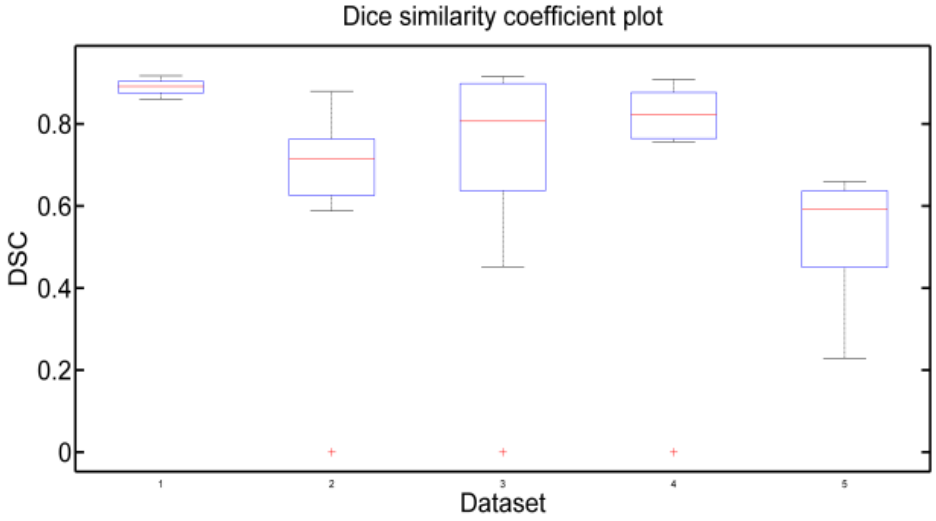


Fig. 5.19 The improvement of post processing of LSM segmentation. (a)(d)(g) are three HN MRI slices at top, middle, bottom position of MRI volume (all from MR28082014), (b)(e)(h) are sliced LSM segmentation output. (c)(f)(i) are segmentations after post processing.

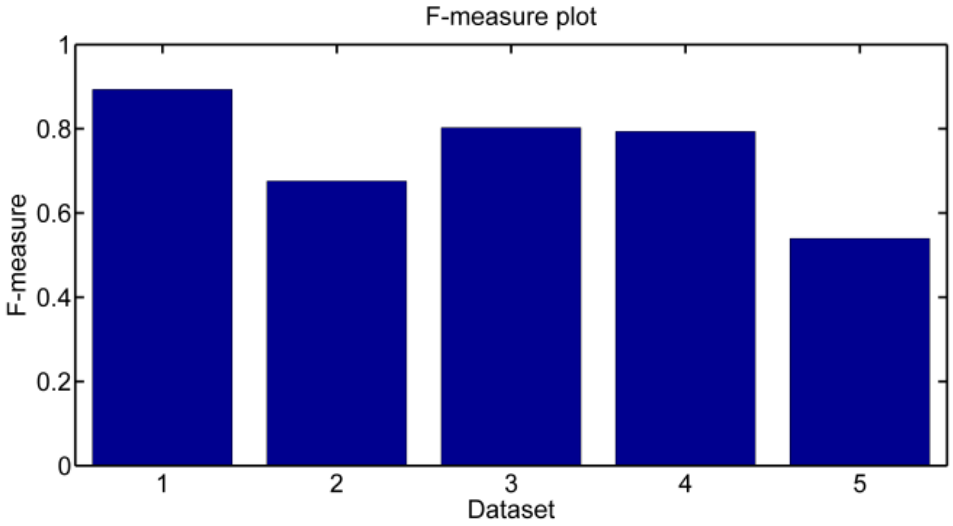
Fig. 5.19 gives the 2D view of the post processing. In 2D perspective, (e)(h) are hard to be distinguished which part is the main target for segmentation. But they are well solved by 3D post processing, and a 2D as well as 3D topologically complete ALN is segmented.

5.6 Results

To quantitatively measure the segmentation results, we use previously introduced metrics called Dice similarity coefficient and F-measure to validate our algorithm. As prementioned in Chapter 4, the validation was on HN datasets provided by Beatson, West Scotland Centre, and on the datasets the gold standard ground truth contours of abnormal lymph nodes are delineated by clinicians from Beatson. The DSC and F-measure are calculated by comparing algorithm output with manual segmentation.



(a)



(b)

Fig. 5.20 ALN segmentation validation (a)Dice similarity coefficient on 5 head and neck MRI datasets, (b) F-measure on 5 head and neck MRI datasets. Both in order: MR30502012, MR17102013, MR14012013, MR10062013, MR09082010

Fig. 5.20 shows the DSC and F-measure plots of segmentation results. The mean DSC through 5 datasets is 70%. The average false negative rate is 0.0025, and the average false positive rate is 0.2023. Based on the FN and FP rate, the fuzzy boundaries of ALNs are likely to lead to over estimation in the LSM segmentation. On the other side, the start and end (vertically) of ALNs are smaller than middle (vertically) part of ALNs, which make them hard to be found (both in localization and delineation) on those slices and might lead to lower Dice score there. The proposed majority voting algorithm of 2D detections (Fig. 5.14) can compensate information from high confidence detections (such as large significant ALNs) to low confidence detections (such as small ALNs at two ends), so that help the localisation. The average F-measure is 0.7. All datasets have above 60% DSC, the highest can reach 90%, and the lowest is around 60%.

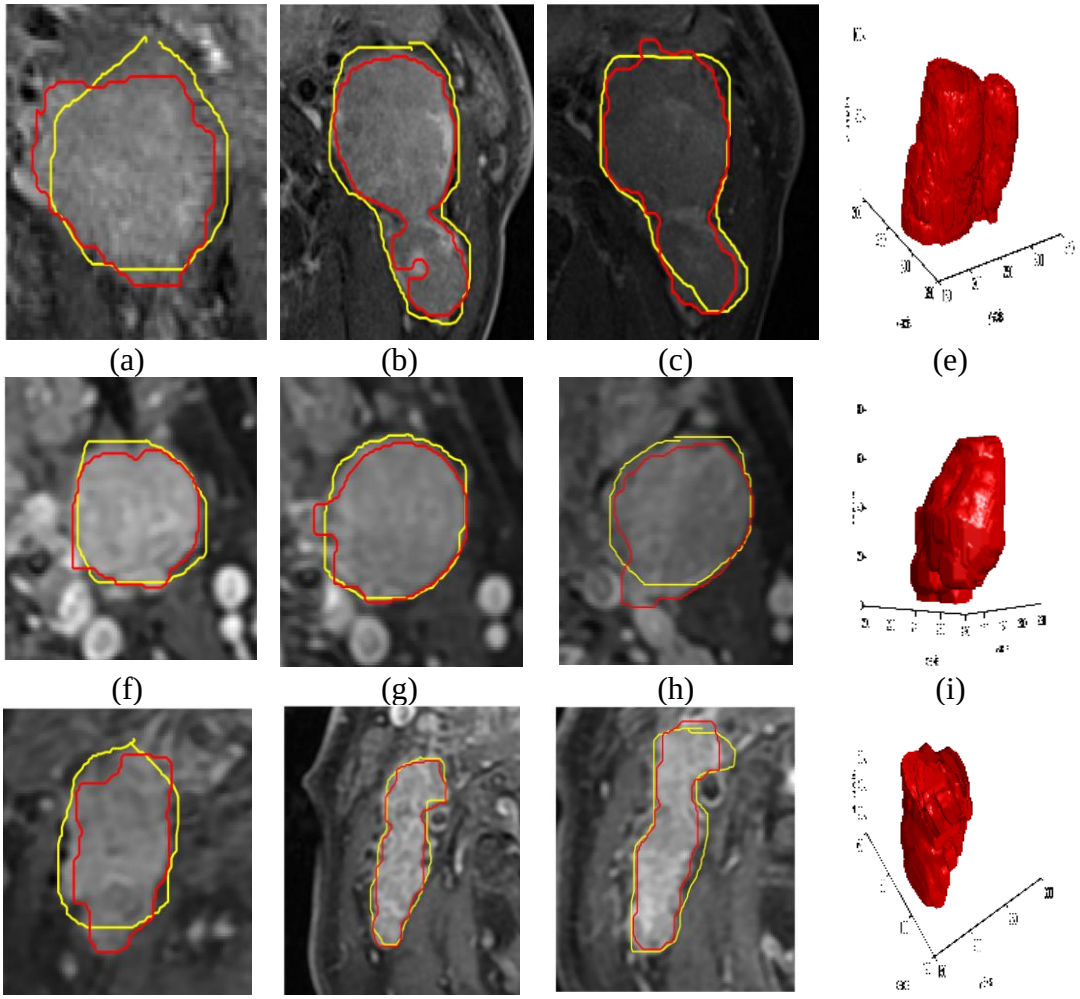


Fig. 5.21 Visualisation of segmentation results. (e)(i)(m) are 3D volume of 3D ALNs. Each row in (a-c) (MR30052012), (f-h) (MR14012013), (j-l) (MR10062013) red contours are 2D sliced 3D output at top, middle and bottom of MRI volume, yellow contours are corresponding gold standard contours.

The algorithm was implemented with Matlab on 16G RAM Desktop. Each dataset has 8-10 MRI slices; the average processing time of proposed algorithm on each data set is 250 seconds, and time cost on each slice is seconds.

5.7 Conclusion

This chapter presented a new algorithm for 3D automatic detection and segmentation of abnormal lymph nodes in head and neck MRI data. The results show that the

proposed algorithm can yield 70 % Dice score, which is comparable to manual segmentation.

The proposed algorithm is based on knowledge-based localisation and intensity-based level set method segmentation. There are several advantages of this method. The whole process is automatic, which can greatly reduce user's workload. The result of segmentation is a 3D model, which can help clinicians visualise and quantify 3D volume from 2D scans. Each step in the proposed algorithm is interpretable, which makes it explainable for clinicians to make final treatment plan so that avoid bad cases created by computer algorithm. Furthermore, it enables clinicians to participate in some steps of the algorithm to get create certain modifications. The algorithm does not require supervised training, so that it can works on small amount of data, wherein the data acquisition and labelling is always a major difficulty in medical image processing area. The hardware requirement is not significant as no GPU is required for this method.

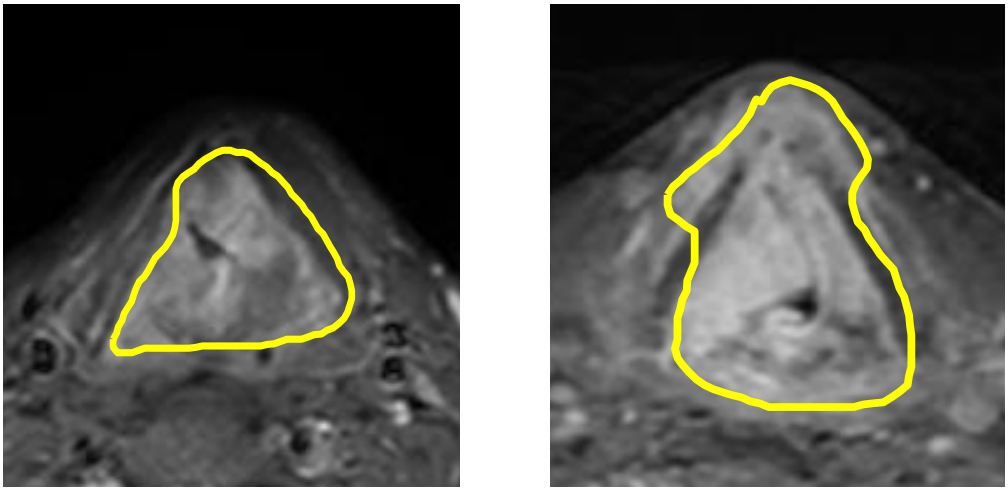
There are certain drawbacks with the proposed algorithm. As this method is a chain of many steps, there are many parameters that need to be set. The segmentation is currently only on the most obvious ALN and does not capture all ALNs in the MRI volume. The ALN has relevant regular shape and appearance, however head and neck tumours can be arbitrary shape and has more complicate texture.

In the next chapter we will introduce the modified algorithm which can be used to automatically delineate the tongue base and throat tumours.

Chapter 6 Novel 3D Segmentation Methods for head and neck tumours from MRI Data

6.1 Introduction

The last chapter presented the 3D detection and segmentation of abnormal lymph nodes in head and neck. Similar to abnormal lymph nodes, the larynx and pharynx tumours are also around the throat, which can be roughly located using throat as landmark. Compared to the abnormal lymph nodes, the larynx and pharynx tumours can have arbitrary shapes, fuzzy boundaries, and complex textures, which cause additional challenges for effective segmentation.



(a) (b)

Figure 6.1 Two example MRI slices with tumour drawn by yellow contours. (a) from MR10062011, (b) from MR27082012. In Both slices, the tumour areas have non-uniform intensity. Shape of two tumours varies. In (b) it can be seen that the bottom and top part of tumour boundaries are fuzzy.

This chapter will present the novel algorithm for 3D segmentation of head and neck tumours from MRI data, which includes a further improved knowledge-based detection procedure, and a spatial-weighted level set function.

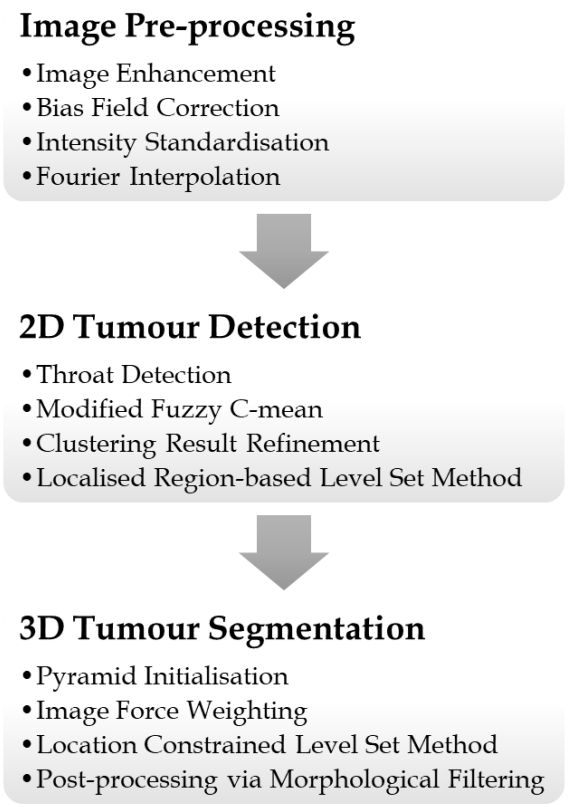


Fig. 6.2 Workflow of proposed 3D automatic segmentation of head and neck cancer from MRI data

The steps of the proposed 3D auto-segmentation framework are shown in Fig 6.2, the rest sections of this chapter will be organised as, 1) pre-processing of MRI data for HNC segmentation, 2) locating of HNC position, 3) segmentation of HNC using modified LSM.

6.2 Pre-processing of MRI data for 3D auto-segmentation of HNC

Similar to the method described in Section 5.3, the raw MRI data needs to be pre-processed (shown in Fig. 6.2) before applying detection and segmentation steps. The pre-processing steps follow the same workflow as the 3D segmentation of abnormal lymph nodes. Firstly, the raw MR images' contrast are enhanced by adjusting the histograms. Then, the background noise is removed by chains of morphological

operations. The intensity inhomogeneity in each slice is removed by method introduced in [193], which has been mentioned in Section 5.3.2.

Before the pre-processed images are interpolated into 3D volume for 3D segmentation, one more pre-processing is introduced for the purpose of reducing inter-slice intensity variations. As aforementioned in Section 2.3.5, the inter-slice intensity variations means that in a series of MRI slices, the same tissue has different intensity ranges across all slices. To improve 3D segmentation of HNC, the slices' histograms of a patient are standardised before interpolation and the subsequent segmentation steps.

In this work, the standardisation of the histogram is achieved by mapping histograms of all slices to central slice, so that to ensure same types of tissues in each slice have similar range of intensities. The mapping is using histogram equalisation (HE) [124], which has been introduced in Section 5.3.1, where it is used for contrast enhancement with a flat histogram as target of mapping. To standardise the histogram, the HE is using HE with histogram of reference (central slice) image as target of mapping. The typical result of the HE processes is illustrated as Fig. 6.3.

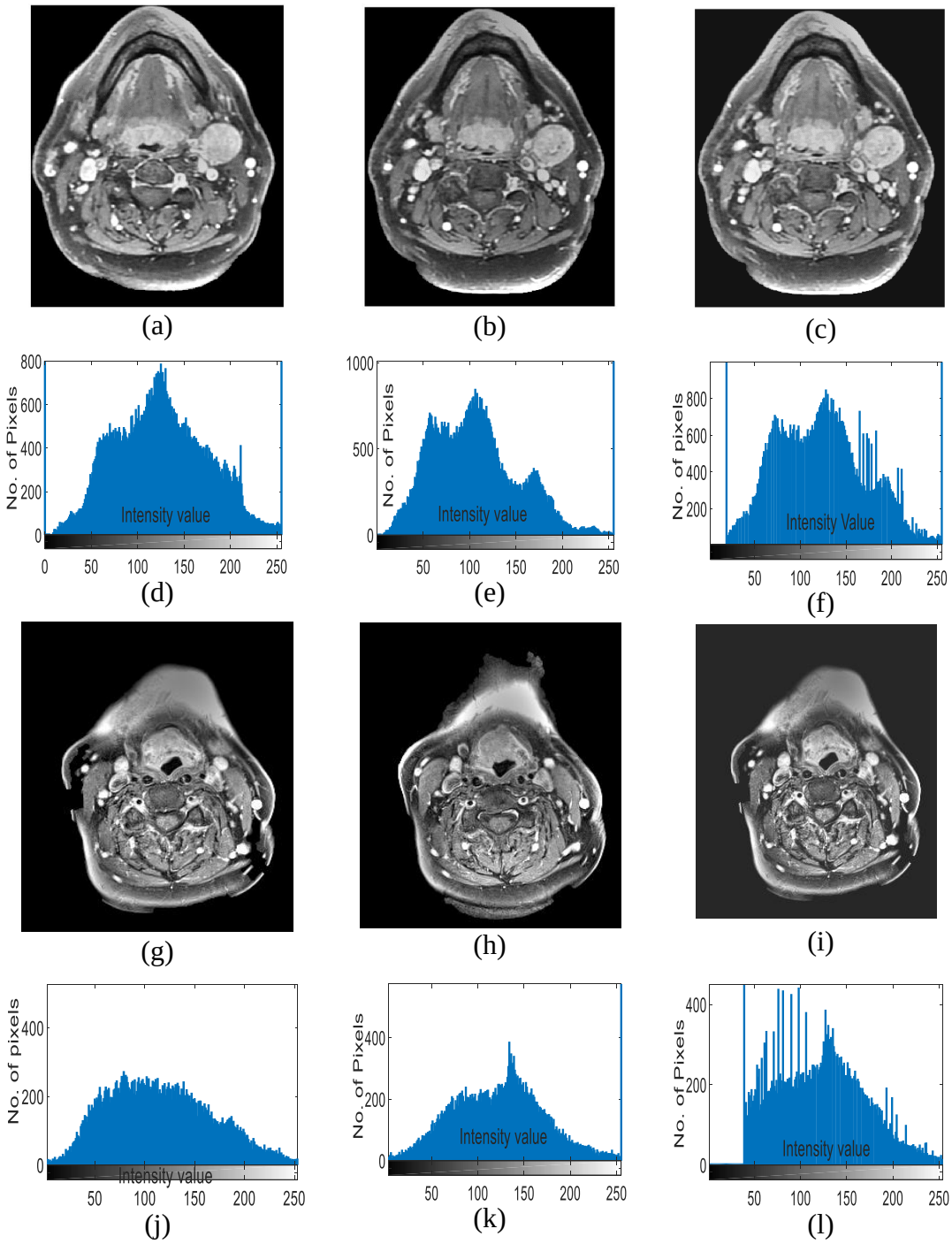


Fig. 6.3 Intensity standardisation by histogram equalization. (a) (from MR14012013) (g) (from MR09112012) are reference MRI slices and (d)(j) are their histograms. (b)(h) are pre-processed MRI slices and (e)(k) are their histograms. (c)(i) are intensity standardized version of (b)(h) and (f)(l) are equalised histograms.

Fig. 6.3 (c)(i) show intensity standardized version of (a)(g), under the reference of (b)(h) respectively. From the images and the histograms, the standardized images have closer intensity distribution to reference images before standardization.

6.3 2D detection of tumours at the central slice

As shown in Fig. 6.2, before the 3D segmentation of tumours, 2D detection is necessary to locate the tumour in MRI volume. The localization of tumours has similarities with the localization of ALN (introduced in Chapter 4). The similarities include: 1) using throat as landmarked to guide the detection process, 2) using fuzzy c-means to cluster tissues based on their intensities 3) refine the clustering results to select binary candidate regions I_B which possible are targets (ALN or tumours here). About the further processing of I_B, last chapter the probability maps methods are applied for detection of ALNs. However, the detection algorithms of tumours are differently designed due to several reasons. Firstly, compared to ALNs, head and neck tumours can have arbitrary shapes, which make them cannot be distinguished through shape features. Then, the tumours can have non-uniform intensities. In addition, the tumours' intensities and locations are close to many adjacent tissues such as lymph nodes and salivary glands, which can likely lead to segmentation errors.

These features of HNC appearance will influence the development and implementation of auto-segmentation algorithms (3D LSM here). As introduced in Chapter 3, the accurate evolution of 3D LSM for segmentation of a target needs a good initial seed, and a well-defined speed function. Considering these factors, a 2D segmentation of HNC on central slice is involved first to locate and guide the evolution of 3D LSM. The process of detection and segmentation of head and neck tumours on central slices is illustrated in Fig. 6.4.

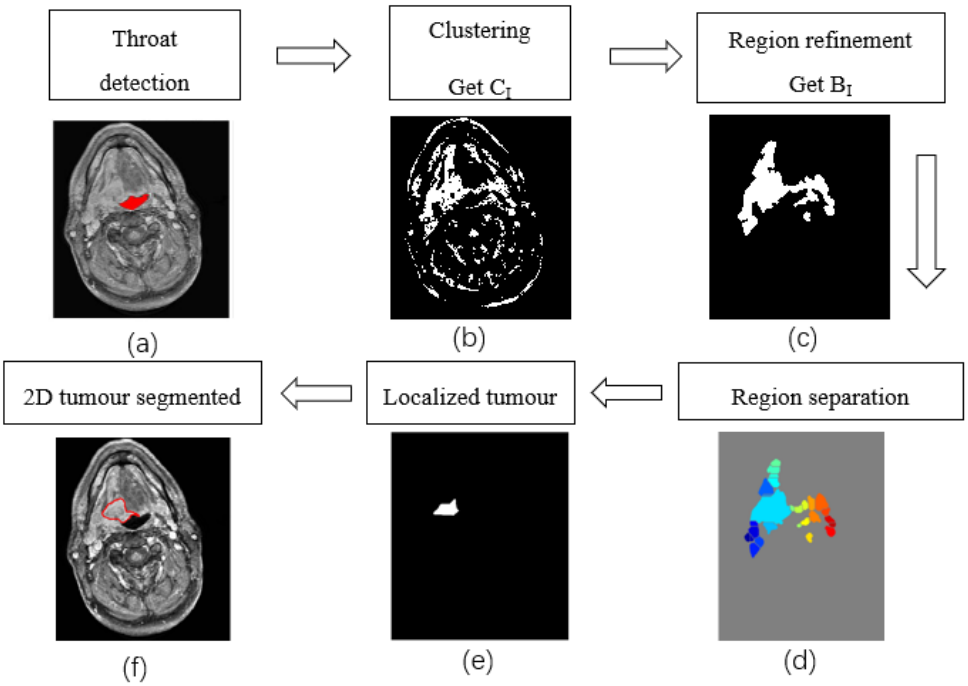


Fig. 6.4 Workflow of 2D detection and segmentation of head and neck tumour on central slices of MR12082013.

Fig. 6.4 shows the workflow of segmentation on central slices. The first step is throat detection, which described in Fig. 5.6. Then the modified fuzzy c-mean described in Section 5.3.2 [321] is applied to classify all pixels into five clusters (based on the knowledge of existence of fatty tissues, cancer tissues, normal tissue, normal muscle tissue, and background pixels in MRI slices). The clusters with the first and second highest mean intensity values are taken (name the combined clustering results as C_I), where the tumours are mostly among these regions. The clustering and combination of clusters were described in Fig. 5.10, the process of refining the C_I to get the refined Binary region B_I is illustrated in Fig. 6.5.

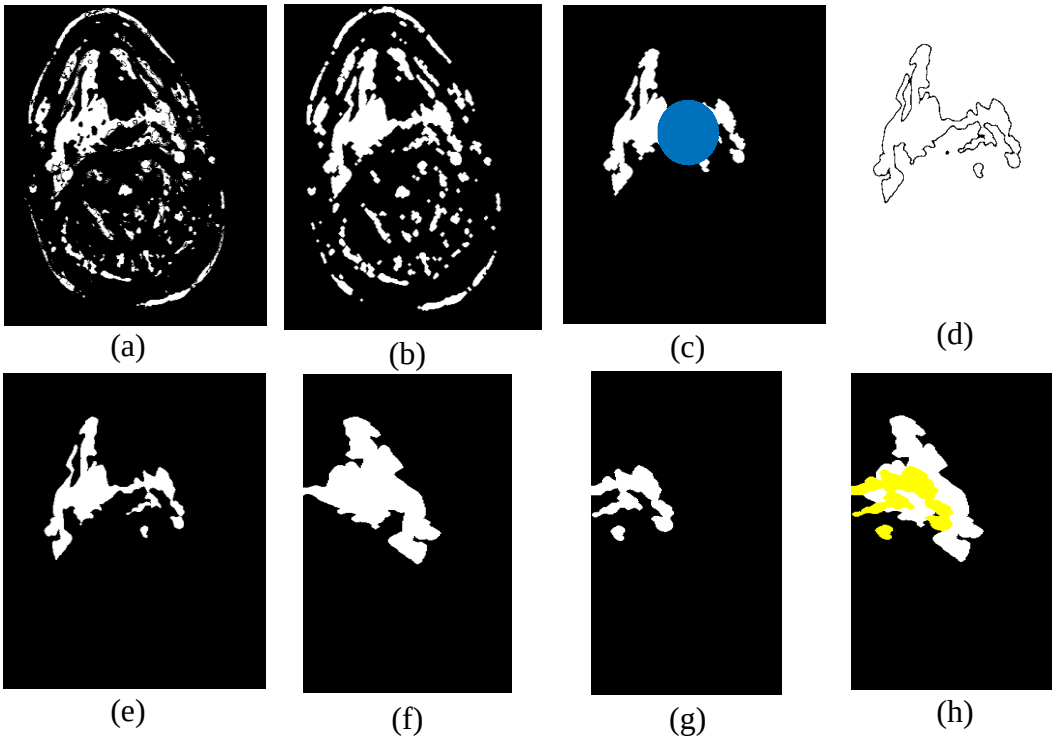


Fig. 6.5 Illustration of processing from clustering results C_I to refined binary regions B_I. (a) Clusters with first and second highest mean intensities combined by bitwise AND (i.e., C_I). (b) Holes and small gaps in C_I filled by morphological operation. (c) Only regions around throat are persevered, the blue circle is the search area around throat. (d) The edge information used for further separations. (e) The refined regions B_I. (f) The left half part of B_I flip to right side. (g) The right half part of B_I (h) The overlay of left (white) and right (yellow) part.

As shown in the (b) to (c) of Fig. 6.5, the small regions in clustering results are removed, only the large regions around throat are kept (called B_I). In this refinement stage, firstly the small holes and gaps in Fig. 6.5 (a) are filled, thus a better binary image is obtained as shown in Fig. 6.5 (b). Then the binary regions away from throat are removed (as shown in Fig. 6.5 (c)), this is implemented by applying a circular mask (the blue one in Fig. 6.5 (c)) around throat centre on binary regions. The regions around throat are further examined, edge information on original grayscale MRI image is used to separate the regions. The edges are extracted by SUSAN [252] detector which mentioned in Section 3.4.1. Briefly, SUSAN algorithm builds a circular detector (like a window) with a nucleus in centre, the SUSAN detector slides through the MRI image and compare pixels inside detector with pixel at nucleus, so that

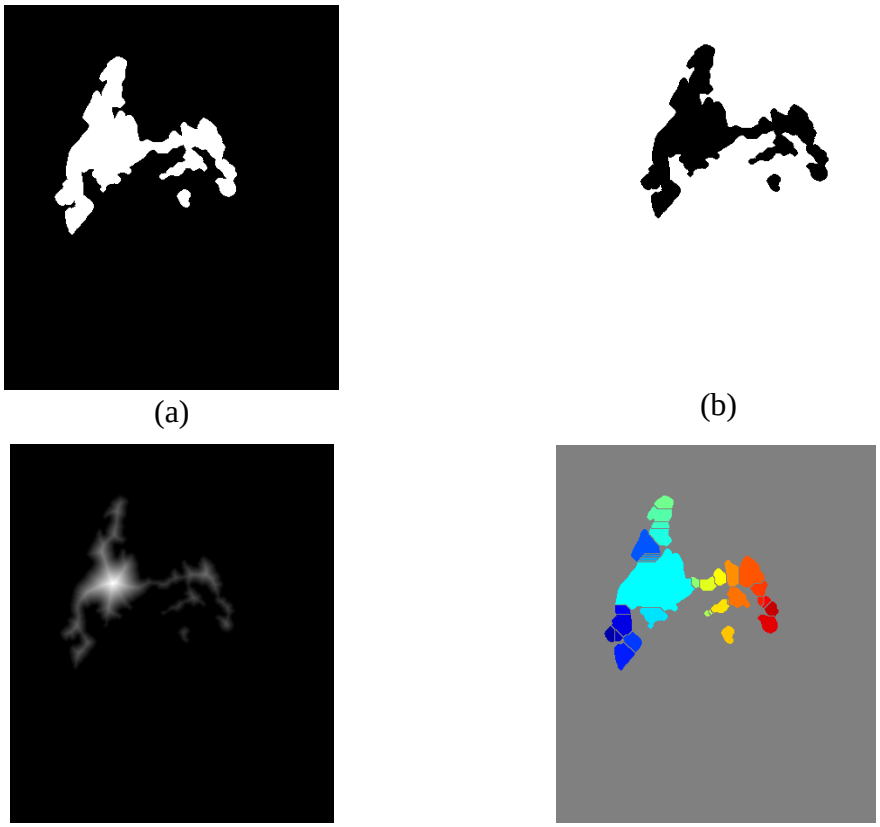
classify these pixels into two classes. In this way, edges and corners can be found, then used in separation of regions.

Separated regions away from throat are removed again (as shown in Fig. 6.5 (d), but in this case no regions are further removed). Then the algorithm checks the remain regions on both sides (left start, right start) and remove regions with high symmetricity, because they are very likely to be salivary glands [319]. The symmetry information is measure by following equation,

$$SymScore(B_I) = \frac{R(B_I) \cap Flip(L(B_I))}{\max\left(R(B_I), Flip(L(B_I))\right)} \tag{6.1}$$

The left side of Eq. 6.1 stands for the symmetry score of binary images B_I, which measure the symmetry information in this image. On the right side, firstly the split the clustering results C_I into left and right part, the mirror line is the centre column of throat. A comparison is applied between two parts, this comparison takes left part (shown in Fig. 6.5 (f)) to flip (horizontal mirroring) to right part (shown in Fig. 6.5 (g)), and the regions in left part will overlay (shown in Fig. 6.5 (h)) with regions in right part, thus based on the Eq. (6.1), the symmetry score of each region is calculated. This score will be used later.

After the refinement process shown in Fig. 6.5, candidate regions B_I for tumours detection are obtained. The tumours are expected inside B_I, but more accurate localization is required to achieve the detection of tumour. A watershed algorithm [322] is applied to further separate B_I, which is shown in the right bottom of Fig. 6.4. The process of watershed region separation is pictorial illustrated in Fig. 6.6.



(a) (b)

(c) (d)

Fig. 6.6 Watershed transform for splitting entire region into small objects. (a) The binary image B_I (b) Complement of B_I. (c) Distance map generated. (d) Small regions obtained by applying watershed (W_I).

As shown in Fig. 6.6, the binary regions B_I (Fig. 6.6 (a)) are split into small segmented sub-regions (Fig. 6.6 (d), call them W_I) by using watershed. The watershed algorithm generates sub-regions in an image by finding 'watershed ridge lines' and 'catchment basins (valleys)'. In this process, the high intensity pixels are regarded as high elevations and dark pixels are regarded as low elevations. Firstly, the binary regions B_I are taken its complement (as shown in Fig. 6.6 (b)). Following, the geodesic distance map is calculated by each pixel to its nearest nonzero pixel in Fig. 6.6 (b). The distance map is shown in Fig. 6.6 (c), by applying watershed on the distance map, the sub-objects W_I are segmented as shown in Fig. 6.6 (d).

The regions in Fig. 6.6 (d) are well divided, which enables more accurate refinement for the localization of tumour. This is referring to process from Fig.6.4 (d) to Fig. 6.4

(e). To make this clearer, as shown in Fig. 6.7 (a-f), the extracted B_I and C_I may have several symmetry binary regions, which includes small body fats, larynx tumours, and salivary glands. When only consider regions around throat, the larynx tumours and salivary glands are major existed symmetrical regions (marked with dash boxes in Fig. 6.7). Then, cancerous regions can be symmetrical around throat (Fig. 6.7 (a)(b)) or majorly above throat (Fig. 6.4 (a)). According to these, several simple rules are designed which aims to eliminate salivary glands-like regions and initial the 2D segmentation start seeds based on the patterns we found, the steps are illustrated in Fig. 6.7.

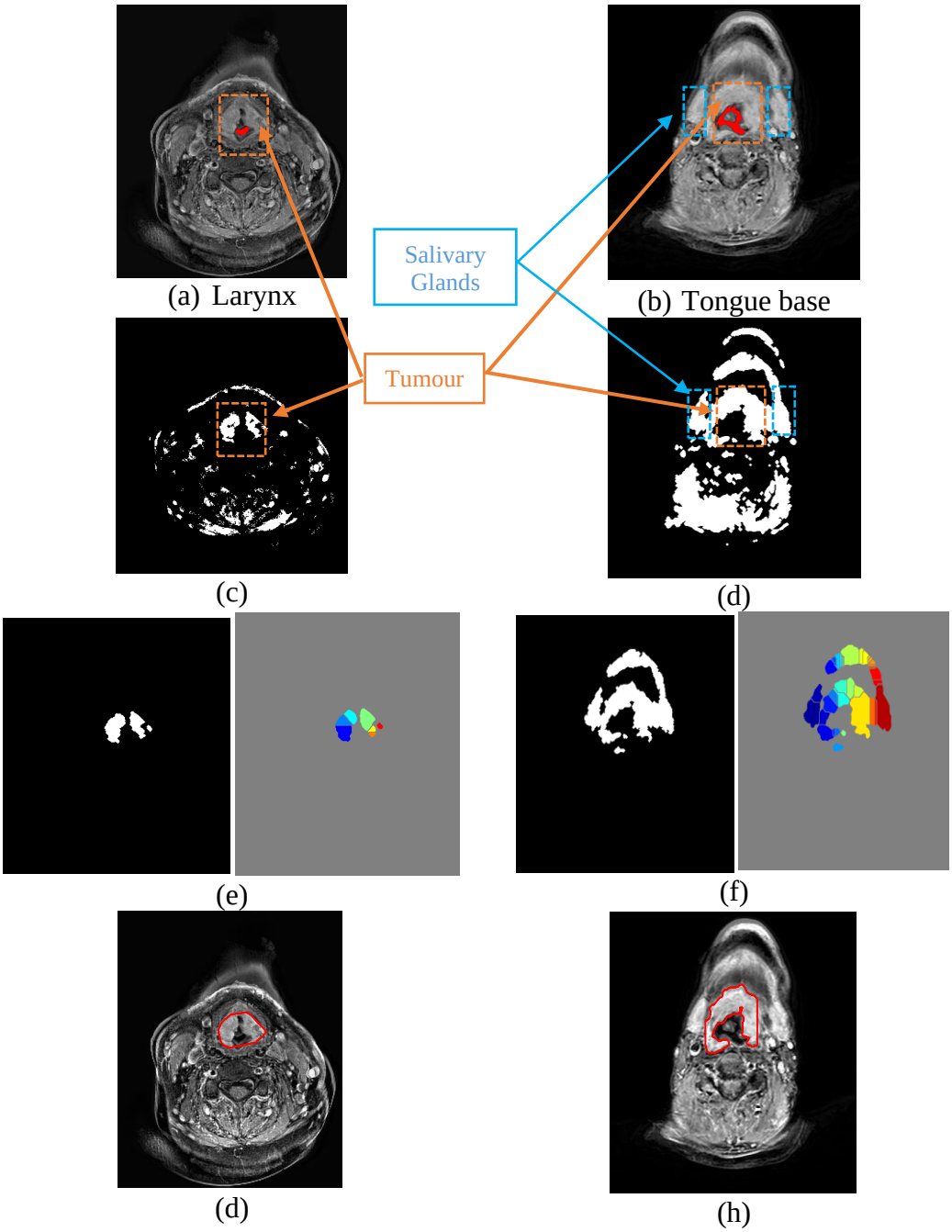


Fig. 6.7 Localize the tumour and initial the 2D segmentation algorithm. (a)(b) Two central slices from two patients (MR09112012, MR10062011). (c)(d) Clustering results (C_I) of them. (e)(f) Refined binary regions (B_I) and watershed transformed applied (W_I). (d)(h) Localized tumour regions and initialized contours for LSM segmentation.

The first refinement rule is that the HNC should be around throat. Then a rule is designed to eliminate salivary glands-like regions, the rule is given as: 1) The B_I has

high symmetry score (set as over 0.5); 2) The regions away from throat has symmetry score over 0.5. If the input B_I fits the rule, then the regions on two side away (but not adjacent) of throat need to be removed, as there might be salivary glands or other regions not wanted. This step aims to preserve regions adjacent throat and find rough centre of tumour, so if regions not salivary glands (FN) are removed, it will not affect results. When the size of objects in W_I are sorted, two patterns should be considered:

1) The first pattern is that one object's size is significantly larger than others, and basically the regions are not symmetrical distributed (such as Fig. 6.4 (a), and Fig. 6.6 (d)).

2) The second pattern is that two or more objects have similar sizes, and some of the regions show symmetricity (such as Fig. 6.7 (e)(f)).

In the first pattern, only the largest object is preserved as the initial seed for LSM segmentation. In second pattern, the symmetrical regions adjacent to throat are persevered, and a convex hull is used to combine them as initial seed for LSM segmentation. As shown in Fig. 6.7 (e-h), the initial seed (contour) are given based on these steps, the LSM will start from here to delineate the contours of HNC.

The level set method used here is localized region-based level set method [323]. In this algorithm, the evolution function is given as,

$$\frac{\partial \phi}{\partial t} = -\delta(\phi(x,y))\nabla F + \lambda\delta(\phi(x,y))\text{div}\left(\frac{\nabla\phi(x,y)}{|\nabla\phi(x,y)|}\right) \tag{6.2}$$

where the first term on the right-hand side of Eq. 6.2 is the external force (i.e., image force) which drives level set evolving to desired boundary and second term is internal force, term div $\left(\frac{\nabla\phi(x,y)}{|\nabla\phi(x,y)|}\right)$ is the curvature of is the curvature of level set function ϕ which keeps ϕ smooth. Here, external force is given as [323]:

$$\nabla F = \sum_{(x,y)\in C} \delta(\phi(x,y)) \sum_{(x,y)\in M_l} \left(\frac{(I(x,y)-\mu_{in})^2}{A_{\mu_{in}}} \right) - \sum_{(x,y)\in C} \delta(\phi(x,y)) \sum_{(x,y)\in M_l} \left(\frac{(I(x,y)-\mu_{out})^2}{A_{\mu_{out}}} \right) \quad (6.3)$$

where μ_{in} is the mean and $A_{\mu_{in}}$ is the area of interior and μ_{out} and $A_{\mu_{out}}$ of the exterior local region $M_l(x,y)$ of each point on the evolving curve (zero level of level sets, i.e., ϕ =0). $(x,y) \in M_l$represent the points inside localized area M_l. $(x,y) \in C$ represent the points on evolving curve (shown in green contour in Fig. 6.8). The $\delta(\phi(x,y))$ takes only point around $\phi = 0$, which is C. Based on Eq. 6.2 and Eq. 6.3, the level set evolution will stop when foreground and background have maximally separate mean intensities. The illustration of the definition of level set function is also shown in,

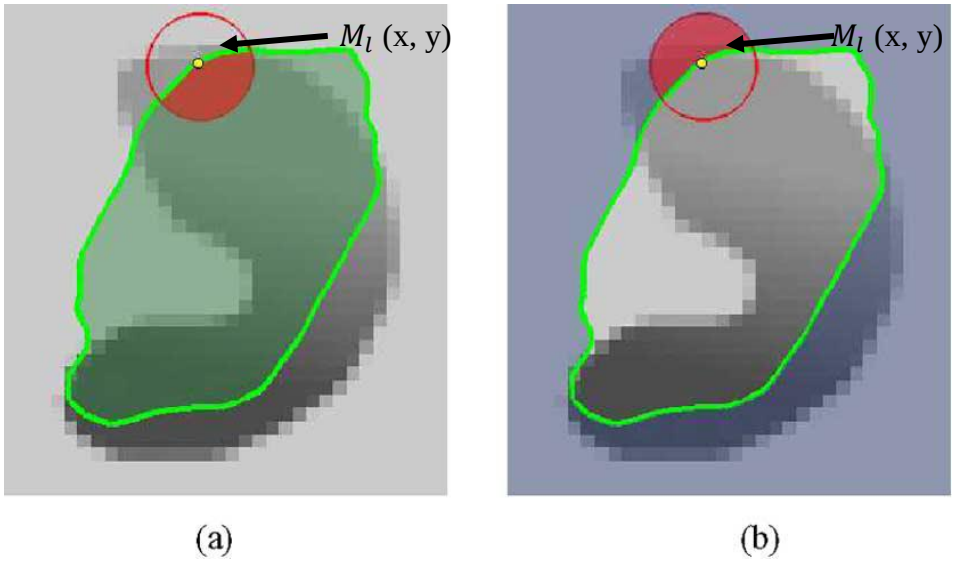


Fig. 6.8 Illustration of local region-based level set method. The evolving contour (i.e., ϕ =0) of level set is represented by green curve. The dark regions in (a) and (b) are targets to segment. In shaded areas in (a) and (b) represent interior (ϕ <0) and exterior (ϕ > 0) of evolving contour (ϕ =0). The red circle denotes the localized region M_l for deducing the evolving force of white point on the green curve [323].

As shown in Fig. 6.8, the white dot stands for point (x, y), its local region $M_l\ (x, y)$ is in red circle. In (a), the local interior is the shaded part of the circle and in (b), the shaded part of the circle indicates the local exterior. Start at the initial contours given in Fig. 6.7, the level set function given in Eq. (6.2) evolves the contours to get the 2D segmentation results. Several examples of evolution of 2D contours are shown in Fig. 6.9,

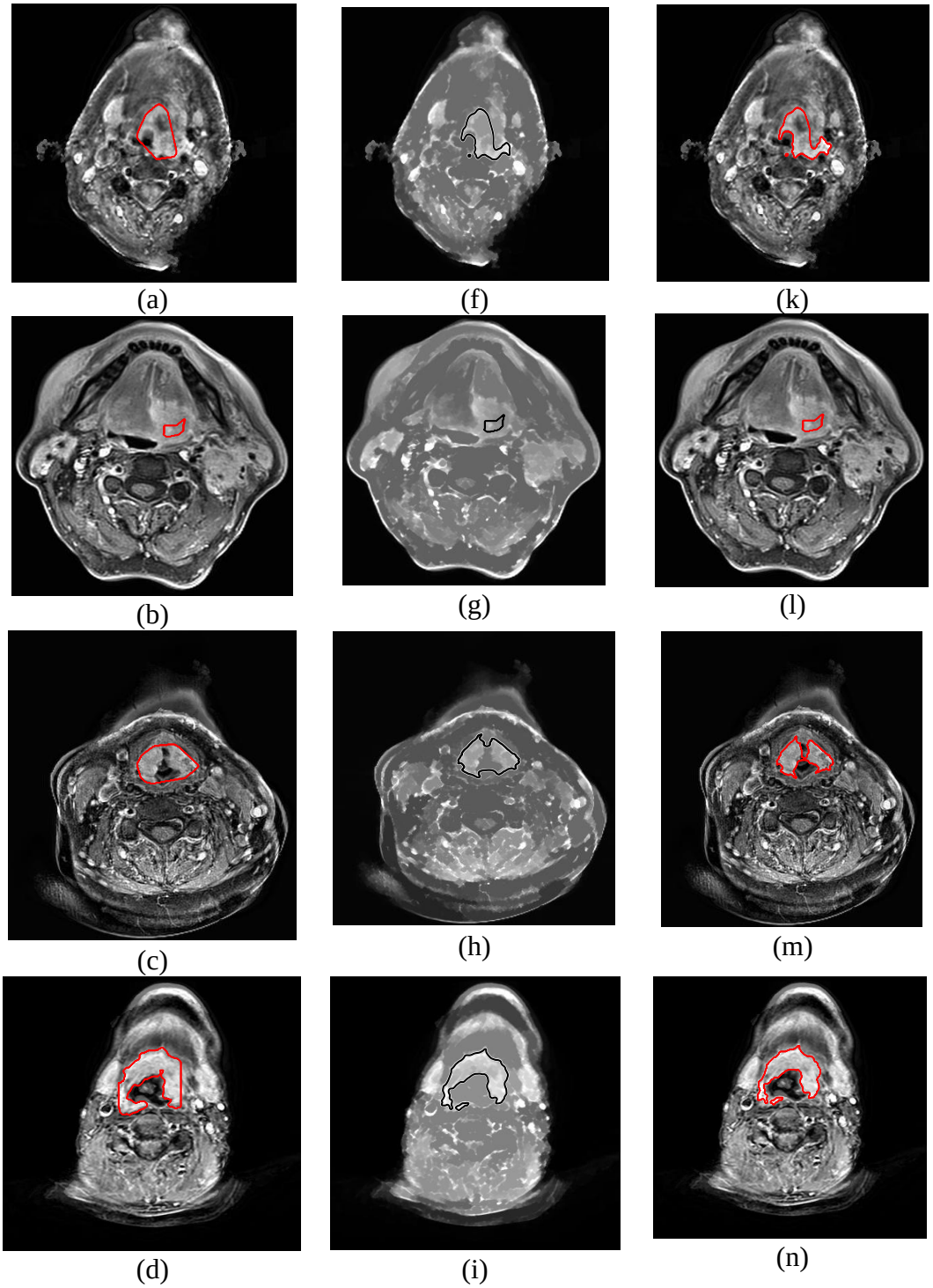


Fig. 6.9 Evolution of level set to segment 2D tumours. (a-d) 2D HN slices (from MR09092010, MR30502012, MR09112012, MR09112012) and initial contours (in red). (f-i) Images morphological closed and level set curve (ϕ =0) evolves at halfway (in black). (k-n) The segmented 2D tumours (in red), name as Seg2D.

As shown in Fig. 6.9, the localized region-based level set method finally extracts the accurate contours *Seg2D* (Fig. 6.9 (l-o)) based on the initial seed (Fig. 6.9 (a-e)). It should be noted that before the starting of level set evolution, the MRI images are firstly processed by morphological close operation, in this way the images are converted to piece-wise constant style (as shown in Fig. 6.9 (f-j)), thus avoiding localized region-based LSM being affected by useless details or noise. In addition, if the initial contour is too small (Fig. 6.9 (b)), it will be taken as the 2D tumour contour (Fig. 6.9 (g)(l)), because the under LSM function the small contour might be vanished. The 2D segmentation results *Seg2D* will lead the localization of initial seed of 3D LSM and influence the setting of 3D LSM function and evolution.

6.4 Definition of spatial constrained level set method

Results of the level set evolution rely on the initialisation, which includes the position and size of initial volume. In this work, the position is set based on Seg2D, the proposed work takes central of *Seg2D* as horizontal position, and the vertical position is half of whole volume's depth due to *Seg2D* is on the central slice. A symmetric pyramid is taken as initial volume (as shown in Fig. 6.10 (a)(d)(g)). The pyramid's max length and width are 5 percent of MRI volume's axial size; and the length and width will be 1 on the top and bottom of MRI volume. These chosen values guarantee that the initial pyramid is not distant from tumour region as discussed in Section 3.2.2.

The level set method algorithm used for 3D HNC segmentation is modified based on [244]:

$$F = \alpha(I - \mu) + \beta \mathrm{div}\,(grad \nabla \phi) \tag{6.4}$$

where the first term on right hand side of Eq. 6.4 is external force (i.e., image force), I is image data and μ is predefined lower bound of the gray-level of the target object. The external force drives the LSM surface evolving based on the intensity values. In this work lower bound of *Seg2D* is taken as μ, which defines the expected intensity ranges of HNC. The second term is internal force, and $\mathrm{div}(grad \nabla \phi)$ is curvature flow weighted by gradient feature map $grad$. The internal force keeps the smoothness of 3D LS surfaces and minimizes the leakage of boundaries. The gradient feature map is given as $grad = \exp(-|\nabla I|^2)$. The level set method (Eq. 6.4) is capable of 2D and

3D tasks, and the gradient map g provides additional boundary information to help LSM detect the edges. However, this level set function is majorly sensitive to intensities and can produce many false positives if the background is complex. Thus, in this work, the speed function is modified as following:

$$F = \alpha\left(\omega_L(I) - \mu_{seg2D}\right) + \beta div(grad\nabla\phi) \tag{6.5}$$

where $\omega_L(I)$ is the location constrain of image force, which de-weights image force of regions far from throat and initial volume, but keep weights of *Seg2D* region. In this way, the occurrences of false positives are constrained. The de-weighting is based on two Gaussian distribution:

$$f_s\left(\left[x_p, y_p\right] \mid [x_s, y_s], \sigma_s^2\right) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{([x_p,y_p]-[x_s,y_s])^2}{2\sigma_s^2}} \tag{6.6}$$

$$f_t\left(\left[x_p, y_p\right] \mid [x_t, y_t], \sigma_t^2\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{([x_p,y_p]-[x_t,y_t])^2}{2\sigma_t^2}} \tag{6.7}$$

Eq. 6.6 shows the de-weighting rule for regions away from initial volume centre (i.e., centre of *Seg2D*). Eq. 6.7 shows the de-weighting rule for regions away from throat region centre. These two rules are same for each layer of MRI volume. $\left[x_p, y_p\right]$ is the coordinate of pixel P, $[x_s, y_s]$ is coordinate of centre of Seg2D and $[x_t, y_t]^2$ is coordinate of centre of throat region. σ_s^2 and σ_t^2 are automatically and adaptively set according to radius of Seg2D and throat region. Based on these two distributions, and the rule that keeping weight of Seg2D region, the $\omega_L(I)$ is given as following:

$$\omega_L(I) = I.* max(\text{norm}\ (f_s.* f_t), \text{Seg2}D) \tag{6.8}$$

where norm means normalise the product of f_s and f_t, *Seg2D* is a binary map which also fits the range [0,1]. In this way image force of pixels away from both *Seg2D* and throat region will significantly decrease, while the weight of *Seg2D* will not be influenced. Using Eq. 6.8 to Eq. 6.5, a location-constrained level set function is acquired.

6.5 Evolution and post processing of 3D LSM segmentation

Based on the Eq.6.5 and initialisation (Fig. 6.10 (a)) of level set, the 3D tumour volume can be segmented as illustrated in Fig. 6.10:

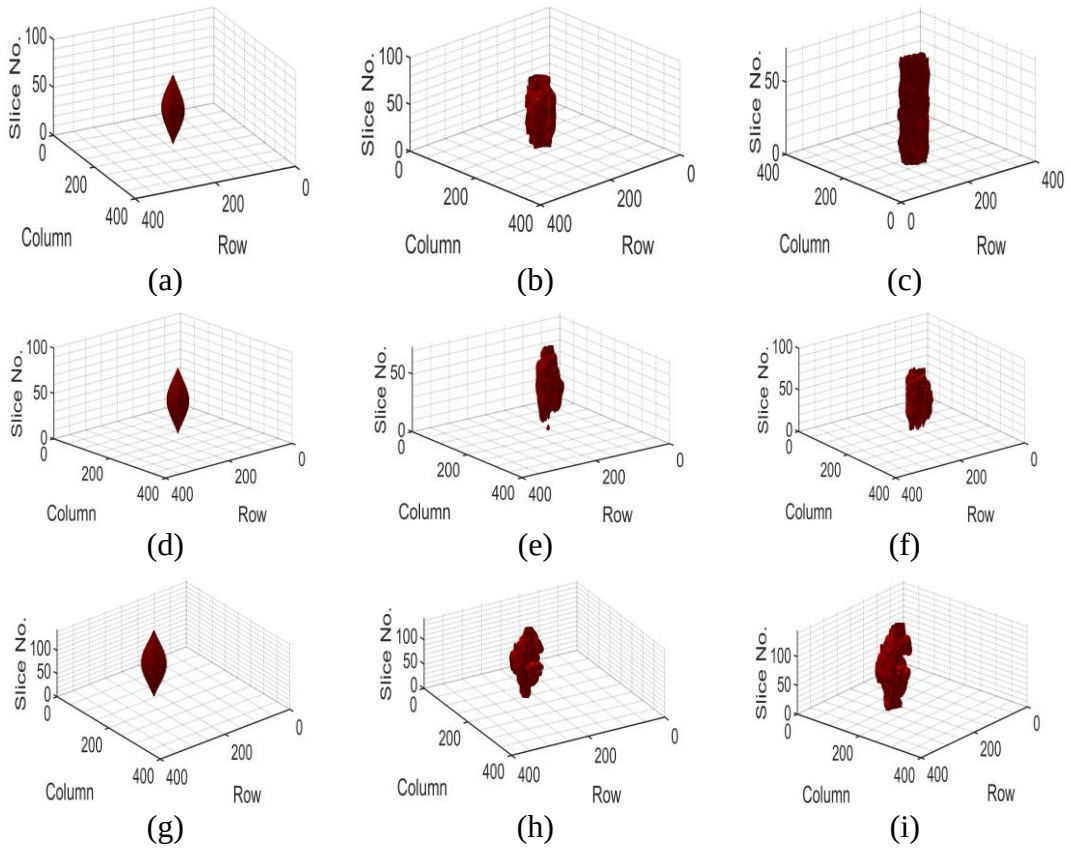


Fig. 6.10 Evolution of 3D surface of HNC under the force of 3D LSM. (a) (MR09112012) (d) (MR14012013) (g) (MR10062011) are initial pyramid of 3D LSM. (b)(e)(h) are surface evolved halfway. (c)(f)(i) The finished evolution of 3D tumour surface.

where the 3D LSM start to evolve at Fig. 6.10 (a)(d)(g), and tracks the surface of 3D HNC, from halfway (Fig. 6.10 (b)(e)(h)) to final segmentation ((Fig. 6.10 (c)(f)(i)). The surfaces are expanding in 3D direction, which can segment the tumours in central positions but also two ends. However, as introduced before, the 3D intensity-based level set method can be too sensitive to intensity value so that produce unsmooth segmentations. Thus, the post-processing is still required to refine the 3D LSM segmentation.

The post-processing is by cascading of 3D morphological operations. First step is using 3D structure elements (sphere with radius of 3 pixels) to dilate the segmented volume to fill small gaps and holes inside tumour volume. Then, 3D morphological erosion (sphere with radius of 8 pixels) is applied to separate weak connections between sub-objects inside segmented tumour. Among the separated sub-objects, only the largest 3D object is preserved. Finally, 3D morphological dilation (sphere structure element of 5 pixels) is applied on preserved object to recover (expand) its volume. In this way, the post-processing refined the segmentation from 3D SLM. An illustration of this procedure is shown in Fig. 6.11.

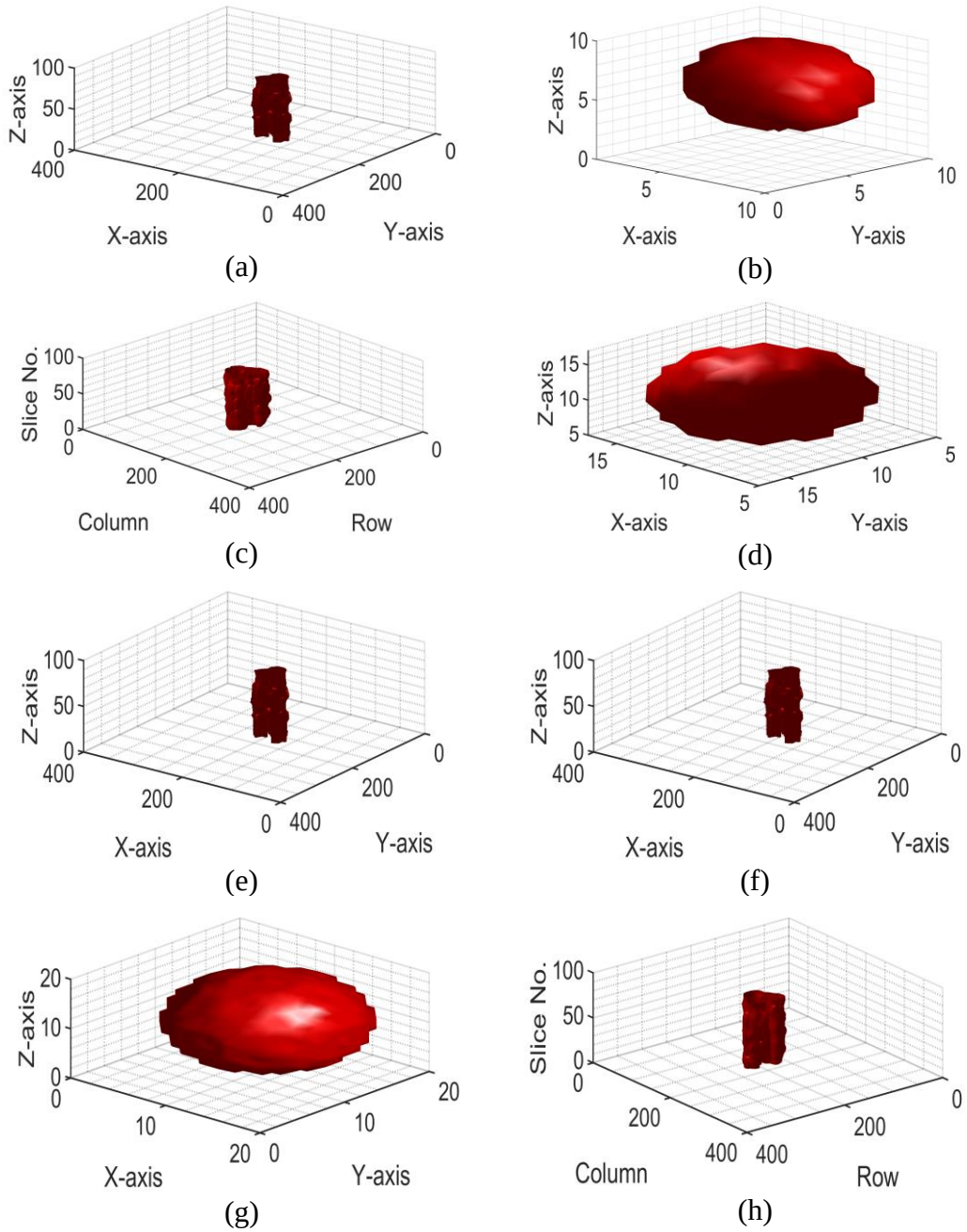


Fig. 6.11 Post processing of 3D LSM segmentation result of MR09112012. (a) 3D tumour volume obtained from 3D LSM. (b) Circular Structure element (radius 3 pixels). (c) Volume dilated and holes filled. (d) Structure element (radius 8 pixels). (e) Volume eroded and separated. (f) largest object preserved. (g) Structure element (radius 5 pixels). (h) Preserved objected dilated to recover volume.

Then in Fig. 6.12, the improvements of post-processing in terms of 2D masks and 3D volumes are demonstrated,

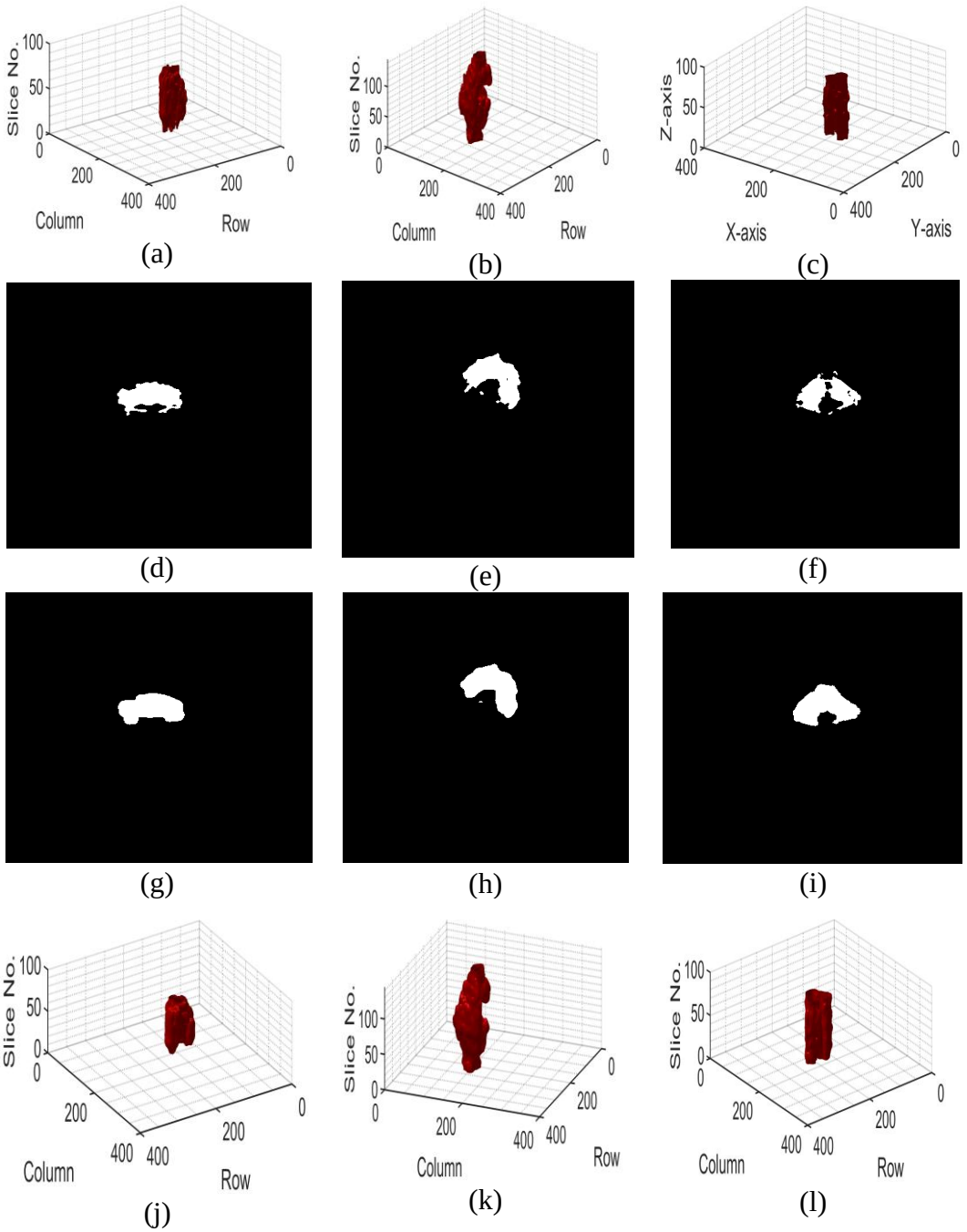


Fig. 6.12 (a-c) Tumours' volume (MR14012013, MR10062011, MR09112012) obtained from 3D LSM. (d-f) Sliced 2D masks of 3D LSM results. (j-l) Refined 3D tumour volume. (g-i) Sliced 2D masks of refined 3D volumes.

As displayed in Fig. 6.12, the comparisons between (a)(j), (b)(k), and (c)(l) show that the 3D tumour surfaces are smoother after post-processing refinement. The comparisons between (d)(g), (e)(h), and (f)(i) show that the post-processing not only smooth the 3D surface, but also improve the sliced 2D segmentations, such as holes inside (Fig. 6.12 (f)), small outliers (Fig. 6.12 (e)(f)), and unsmooth contours (Fig. 6.12 (d)(e)(f)). The post-processing refined 2D and 3D results are taken as the final output of the proposed 3D HNC segmentation algorithm.

6.6 Results

The proposed new algorithm was implemented in MATLAB running on a PC with 16G RAM, 3.2GHz Intel(R) Core (TM) i7-8700 CPU. Experiments were conducted on real MRI datasets from Beatson West of Scotland Cancer Centre to test the performance of the new algorithm. The data contains T1-weighted Gadolinium enhanced MR images of HNC. This section will demonstrate 2D and 3D results of proposed algorithm on real data, and also some quantitative study compared with 2D plus interpolation algorithm [193], and other 3D algorithm [324]. The comparison between results acquired from two methods (proposed and [193]) is shown in Fig. 6.13.

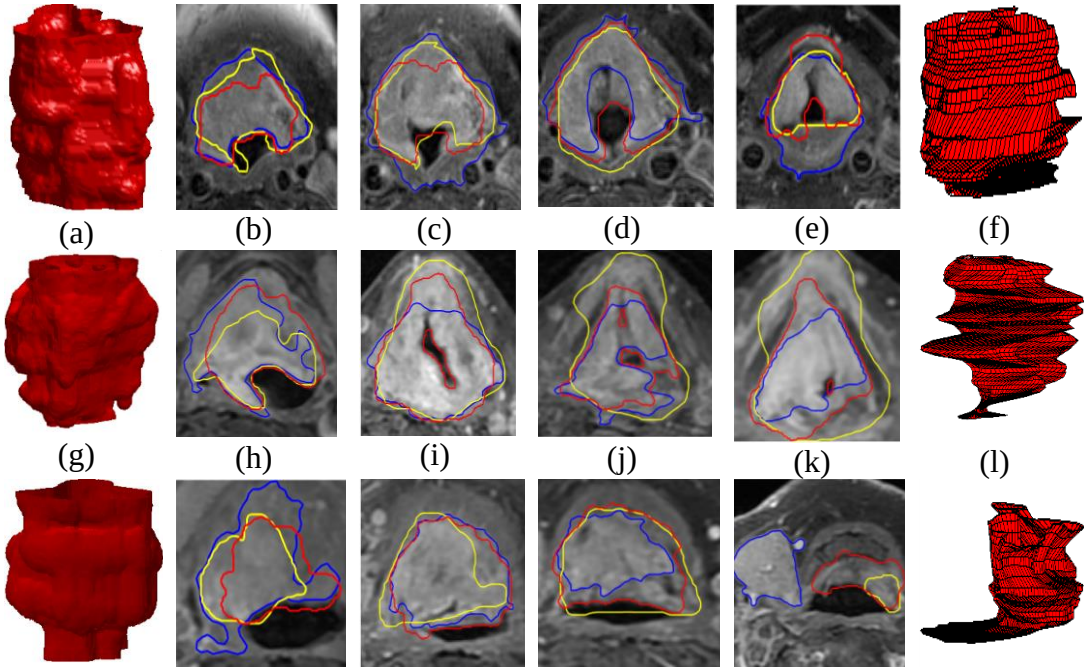


Fig. 6.13 3D tumour volume segmented from real MRI data each row uses the same dataset. (a)(MR11092012) (g) (27082012) ((m) (MR10062013) are volumes obtained by proposed algorithm; (f)(l)(r) are volume acquired by 2D method; (b)(c)(d)(e), (h)(i)(j)(k) and (n)(o)(p)(q) are 2D contours on separate axial slices. Yellow contours are from gold standards (consensus manual outline), red contours are from proposed algorithm, and blue contours are from 2D approach [193].

In Fig. 6.13, each row shows the results on one MRI dataset, and the left parts of each line ((a)(g)(m)) are 3D volume from proposed algorithm, right parts ((f)(l)(r)) are from 2D approach [193]. In the middle of each row are comparisons of 2D contours among proposed algorithm, 2D method, and gold standards. The gold standards are consensus tumour outlines on 2D axial slices according to clinicians from Beatson West of Scotland Cancer Centre.

From the given 2D contours comparison the proposed algorithm has similar segmentation compared with gold standards. Also, the visualisation of 3D tumour volumes in Fig. 6.13 (a)(g)(m) shows that the proposed 3D method can produce tumour volume with smooth surface and consistent structure. On the other hand, the volumes extracted by 2D approach have sharp edges (Fig. 6.13 (f)(l)(r)), and if there

is inaccuracy in 2D segmentation (Fig. 6.13 (q)), the structure of 3D volume will be inconsistent (Fig. 6.13 (r)). More visual results are in Appendix C.

Apart from visual analysis, here are more quantitative measurements of proposed algorithm's performance. The measurements include Dice measure, false negative, and false positive. The bar graph which compares the Dice score between proposed algorithm and method [193] are shown in Fig. 6.14.

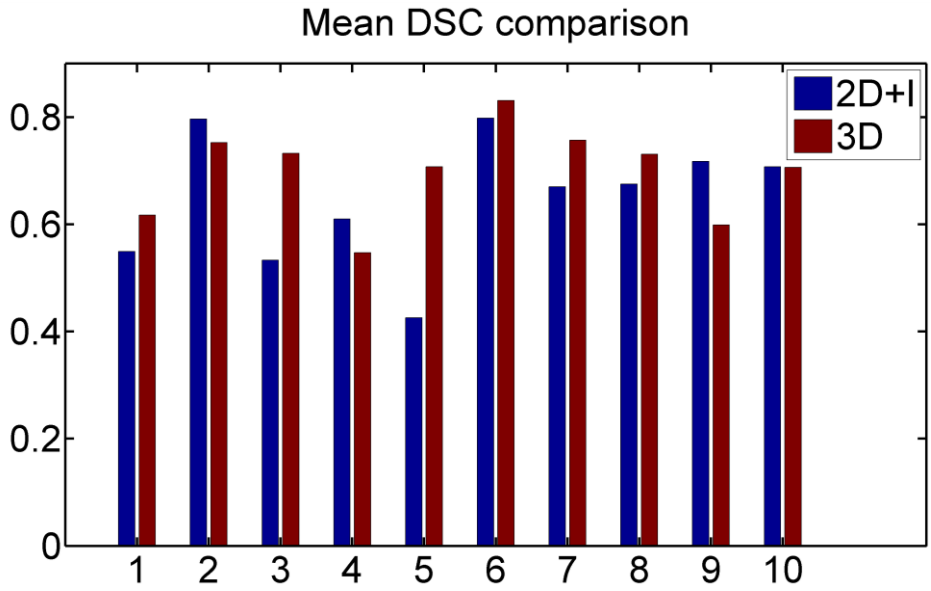


Fig. 6.14 Comparison of DSC between 2D approach verses gold standards (consensus manual outlines), and proposed 3D method verses gold standards. The x axis represents dataset MR29072011, MR14012013, MR10062013, MR09082010, MR12082013, MR09112012, MR27082012, MR19062012, MR10062011, MR09092010.

The bar graph shows that for 70%of the datasets the proposed method performs better. In cases 2, 4, and 9, the 2D+I method is only marginally better. More precise measures of performance among the 3 methods are shown in Table 6.1,

Table 6.1 DSCs, false positive, and false negative rate comparisons.

Methods	DSCs	FP	FN
Proposed	0.6983	0.2401	0.0039
2D+I [193]	0.6484	0.3187	0.0033
3DLSM [324]	0.6302	0.4199	0.0020

According to the bar graph Fig. 6.14 and Table 6.1, the proposed 3D HNC segmentation has better mean Dice similarity score compared to 3D level set method [324] and 2D plus Interpolation approach [193]. On the other hand, the significant reduction of false positive also proves that the idea of location-constrained level set method achieves the expected goals.

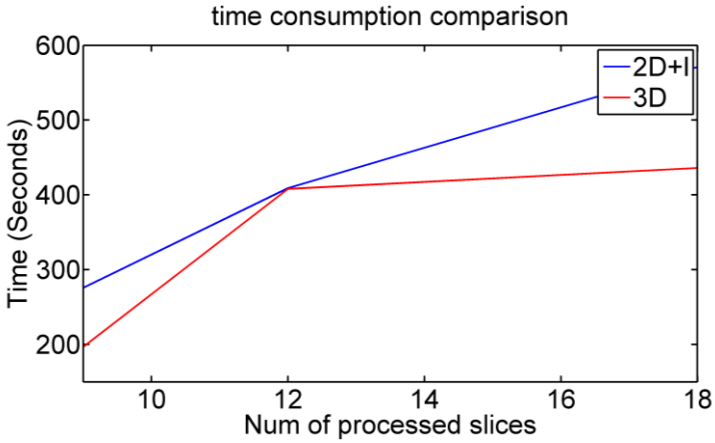


Fig. 6.15 Time consumption of 2D and proposed 3D method, the x axis is the number of input slices, the y axis is the time cost in seconds

The time cost of 3D and 2D+ I methods are shown in Fig. 6.15, it illustrates that the proposed 3D methods also save time consumption for HNC segmentation task. As the 2D methods need to process each slice, when the slices number increase, the difference of time consumption between 3D and 2D approaches will be larger.

6.7 Conclusion

This chapter presented a novel 3D segmentation algorithm for HNC tumour volume extraction. The proposed method uses 2D localized region based LSM to segment the 2D tumour region on central slices, so as to localize the rough position of tumour, and deduce key parameters of 3D segmentation. A location-constrained 3D LSM is

proposed to reduce the false positive generated during the segmentation of HNC from complex background. 3D morphological operations are applied for post-processing of segmentation to improve the smoothness of extracted 3D tumour volume. The proposed method was validated on real MRI data from Beatson west Scotland centre, Glasgow. The validations show that the proposed algorithm achieves Dice score around 70% when evaluated by consensus manual contours, the Dice score is higher than 2D + I [193] and 3D LSM [324] method. The location constrained LSM guided the evolve of tumour surface and reduced the false positive in segmentation, which is about 24%, also improved compared to [193] and [324]. And the proposed 3D algorithm does not lead to more time consumption compared to 2D approach, on the contrary, the time cost will be saved when the slices increase.

The results and performance measure illustrate several advantages of the proposed 3D HNC segmentation algorithm. Firstly, the 3D segmentation is a more natural approach to extract 3D volume of a tumour, which has similarities to the growth of tumour volume in real cancer cases, and practically it can generate uniform volume and avoid unrealistic structure. Secondly, this method introduces 2D segmentation on central slice, which helps the locating of tumour in complex background, so that gives better initialisation of 3D LSM segmentation. Thirdly, the proposed location constrained 3D LSM introduces 2D segmentation results for improved parameters setting, which is able to minimize the false positive extracted HNC volumes.

However, there are still limitations. Firstly, the locating process involves many empirical setting and parameters, which is expected to be improved by statistical approach such as machine learning or deep learning. The 3D LSM has a tendency to maintain balloon shape (the curvature factors in internal force), which means it can expand too much in middle slices but insufficient evolve at the start and end slices (two ends of tumour). The unsatisfactory segmentations can be subsequently manually modified via the software called RTP Aid Tool we developed, this software can modify 2D auto-segmentation contours and reconstruct modified 3D tumour volume. More details of the software are provided in Appendix A. In the next chapter, we will introduce our proposed deep learning methods for semantic segmentation of head and neck tumours from MRI data.

Chapter 7 Automated HNC Segmentation using Deep Learning

7.1 Introduction

The traditional approaches using classical methods (knowledge-based detection with LSM contour tracking) for HNC segmentation provide interpretability and 3D visualisation, does not require significant amount of annotated data or computation power (GPUs). However traditional approaches suffer from subjectiveness, and many parameters need to be empirically set. As introduced in Section 3.5, if more data and GPUs are available, deep learning should be considered for application on HNC segmentation because DL uses end-to-end structure to get rid of hand craft features and majority parameters setting. Also, the more data it uses the better performance DL model could have, this makes it have scalability and robustness to handle potential newcome large dataset. This section will introduce the work of DL on HNC segmentation from MRI data, which is a modified version of U-Net [296]. The remainder of this chapter is organised as follows. Section 7.2describes the architecture of proposed segmentation network, while Section 7.3 describes details of techniques in the proposed work. Section 7.4 demonstrates the segmentation results, and a conclusion Is provided in Section 7.5.

7.2 Proposed Modified U-Net for improving performance of head and neck cancer segmentation

To eliminate the subjective knowledge and achieve end-to-end accurate segmentation of HNC from MRI data, a novel CNN architecture is proposed, whose idea is based on from classical U-Net [296]. Modifications are made to best suit the HNC segmentation problem. Only 2D MRI slices with annotations were available for training. The 2D slices can be reconstructed to a synthetic 3D MRI volume as we described in Section 3.2.1.6. However, to train the deep learning U-Net we only use the annotated 2D MRI image data set. The proposed modified U-Net model is shown as Fig. 7.1.

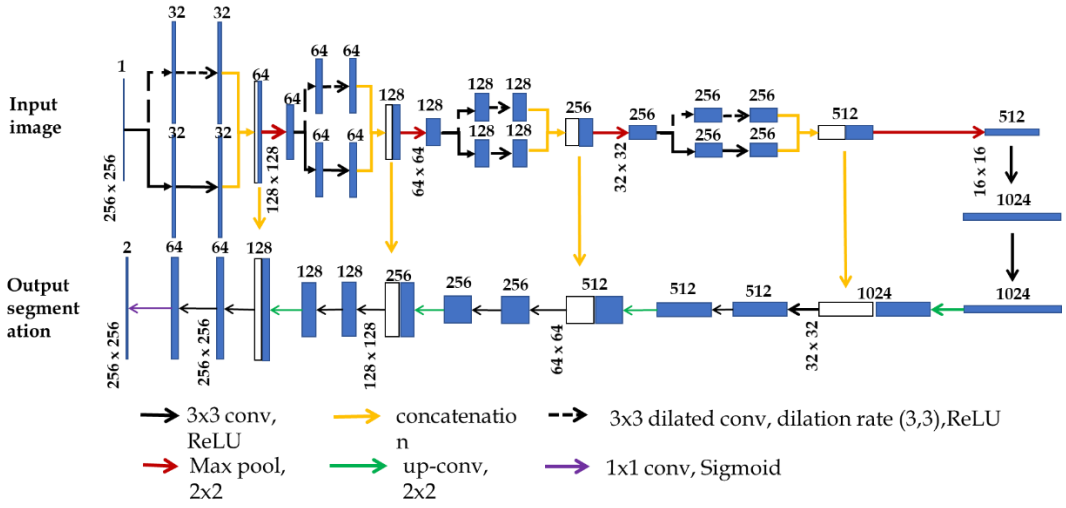


Fig. 7.1 The architecture of modified U-Net. The horizontal numbers on the top of boxes show the depth (number of channels) of feature maps, the vertical numbers on the left show the xy size of image/feature maps. Other operations are marked with arrows in different colours.

The modified U-Net has similar structure with classical U-Net. There are also two major paths inside DCNN architecture, one contracting path also is called as encoder (top row of the network in Fig. 7.1) and one expansive path is also called as decoder (bottom row of the network in Fig. 7.1), and several skip connections between contracting and expansive path.

In the contracting path, the input image is passed through a series of CNN module to extract the features and patterns. The feature extraction by CNN module will cause the reduction of xy size, and the increase of depth of feature maps. There are four stages of downsampling of image xy size, which make the image size reduce to half. With an input image 256 X 256, the deepest feature maps have 1024 channels, and 16 X 16 xy size. The expansive path has four stages of upsampling of xy size, which double the image xy size, and reduce the number of channels of feature maps. The final output of network is a prediction map whose channels equals to the classes need to be distinguished, and xy size equals to input image's, i.e., 256 X 256. The input size 256 X 256 is close to size of Raw MRI slices, besides this size can be divided by two in multiple downsampling operations.

Similar to U-Net, there are skip connections between contracting path and expansive path. The connection is on each stage and via concatenation, thus each skip connection will cause double of the numbers of channels of feature maps. Apart from this, in contracting path, a two-pathway convolution scheme is introduced into U-Net for better feature extraction. The two-pathway scheme use same numbers of filters, but one pathway is regular convolution, another one is dilated convolution. The feature maps from two pathways will be fused together via concatenation.

The major novelty of the modified U-Net comes from the two-pathway feature fusion of multi-resolution features. The two-pathway architecture come from the idea in [89], which aims to exploit both local features and global contextual features simultaneously. But different from the setting in [89], the proposed method uses dilated convolutions rather than classical convolutions, thus achieves larger receptive fields for global features but not involve additional computations. More details about the modified U-Net are in next section, includes several submodules of the network, and training setting.

7.3 Details of proposed DCNN model

This section includes the details of submodules in the proposed modified U-Net, these submodules are embedded into U-Net to improve the performance of segmentation of HNC from MRI data. These will include the two-pathway module, the dilated convolution, and the dice loss to build the model; also, the data augmentation and batch normalisation to improve the training process.

7.3.1 Dilated convolution

In deep convolutional neural network, the multi-stage feature extractions can find hierarchical representations of data, which have been introduced in Chapter 3. The shallow layers obtain the low-level features at local scale, while the deep layers acquire the high-level features at global scale of image. The determination of exact boundaries of target requires low-level features, while the detection and localization of target requires high-level non-local contexts. A classical way of accessing non-local context is to use larger convolution filter size so that get larger receptive field. As we introduced before, a 3 X 3 convolution filter has 3 X 3 receptive field. In [89], the two-

pathway network uses a 7 X 7 filter in one path, and a 13 X 13 filter in another path, which are significantly larger than the general 3 X 3 setting. This will introduce 7 X 7 and 13 X 13 receptive fields to extract non-local context.

However, one 3 X 3 filter has 9 parameters to train, one 7 X 7 filter has 49 parameters to train, the difference will be larger considering there are lots of filters inside a network. The drawback of using large filters can be in two aspects: 1) additional computation from additional number of parameters. 2) The network cannot be sufficiently trained if the number of training samples is small, but the network has too many parameters. Thus, the classical method is to use deeper structure rather than larger filter size, for example, a cascade of two 3 X 3 filters can have same receptive field with one 5 X 5 filter (as seen in Fig. 7.2 (a), the red neuron in third layer can access the context of 5 X 5 neighbour neuron in first layer). 5 X 5 filter costs 25 parameters to train, while two 3 X 3 only need 18 parameters in total, thus the combination of small size filters is a computational economic way to access the non-local context. The pooling layers also help the DCNN to access the global context, but this comes with the decrease of resolution.

In [325], a novel dilated convolution method was proposed to enlarge the receptive field but not at the expense of additional computation cost. The dilated convolution is also called atrous convolution. A seen in Fig. 7.2 (c), the classical 3 X 3 convolutional filter is like a 3 X 3 sliding window, the output value of the centre position of filtering is the weighted sum of all points inside window (red points), and the output is a data representation of the information inside receptive field (shown as blue region). The dilated convolution inserts gaps between convolutional kernels, as shown in Fig. 7.2 (d), with dilation factor 3, there are 2 intervals between two kernels inside filter. The dilated 3 X 3 convolutional filter with dilation factor 3 still uses 9 points (in red) to calculate the filtering output, which does not increase the number of parameters. However, due to the dilation, the output is influenced by context of 7 X 7 neighbour, which contains non-local features. In addition, these non-local features can be obtained from early layers, which maintains the resolution and detailed low-level features. This is also illustrated in Fig. 7.2 (b), the red neuron in third layer connects to the black neurons in the first layer, which is 5 neighbour neurons' away. With the

demonstrated advantages, the dilated convolution was introduced to aggregate multi-scale contextual information without losing resolution or coverage, which are shown by the dashed black arrows in Fig. 7.1.

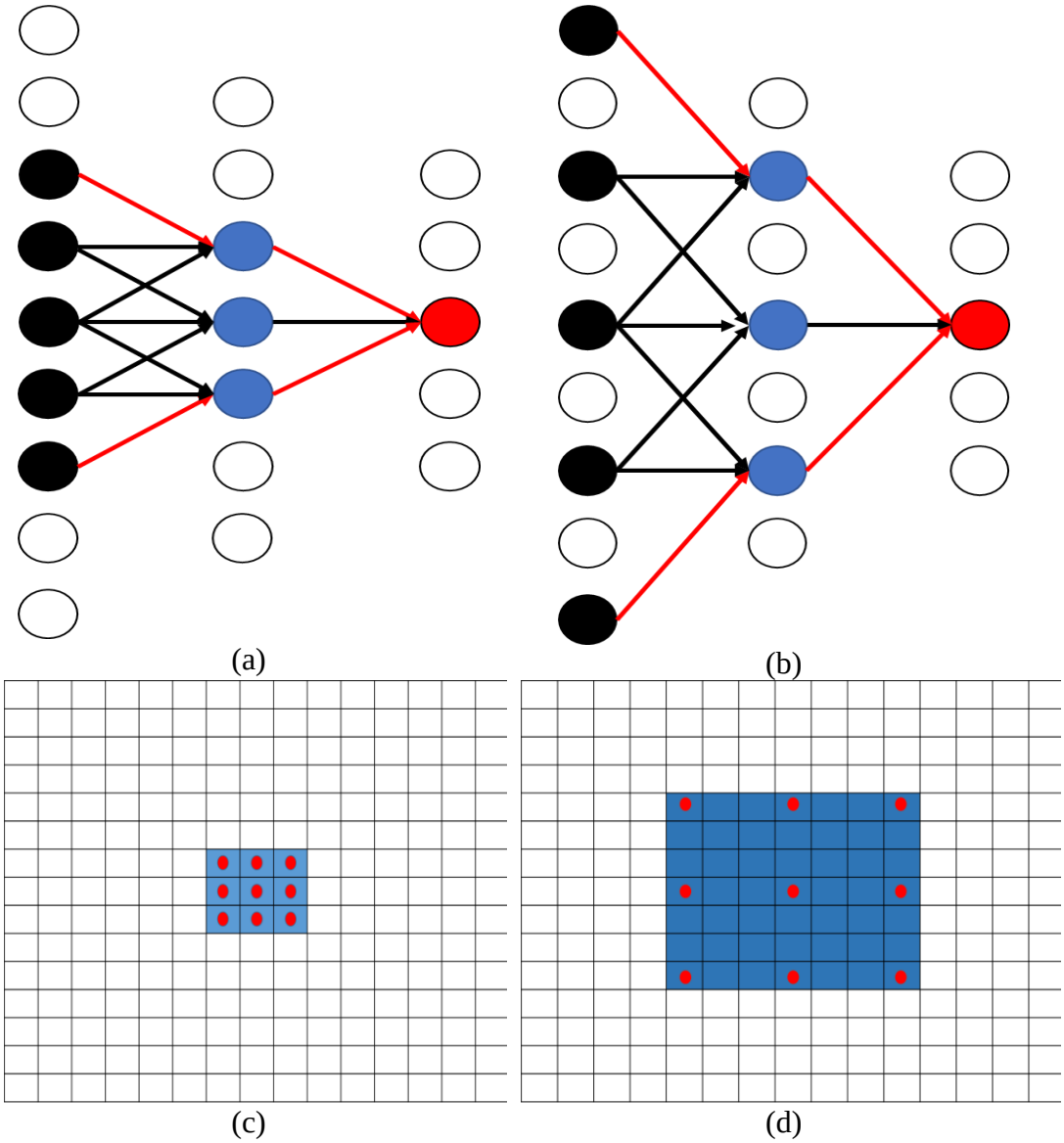


Fig. 7.2 Demonstration of differences between classical and dilated convolutions. (a) is the illustration of classical 3 X 3 convolution through multiple layers. (b) is the dilated 3 X 3 convolution with factor 2 through multiple layers. (c) is the classical 3 X 3 convolution on an image, the red points are the kernels of convolutional filter, the blue area is the receptive field. (d) is the dilated 3 X 3 convolution with factor 3 on an image.

7.3.2 Two pathway module

The classical U-Net aggregates low-level, local, high-resolution features in early layers and high-level, global, low-resolution feature in deep layers via the skip connections between encoder and decoder. This section will introduce the proposed novel DCNN, which involves a two-pathway module to aggregates more non-local features at early layers so that not lose the resolution. The whole network has been demonstrated in Fig. 7.1, here a two-pathway module example is illustrated as Fig. 7.3,

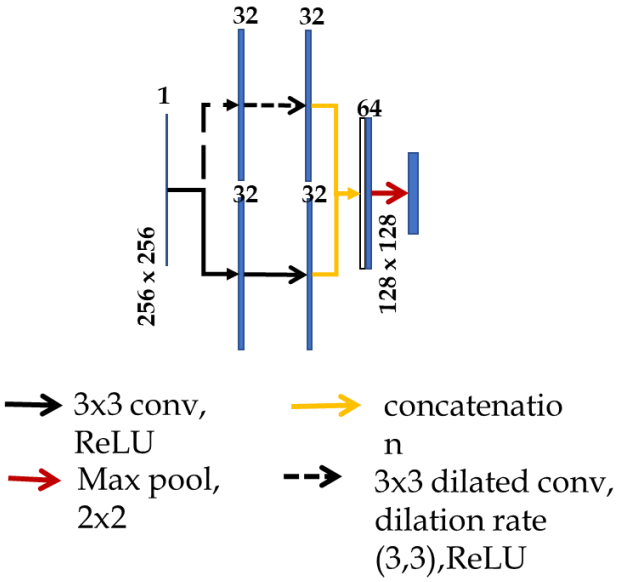


Fig. 7.3 The proposed two-pathway module. Horizontal texts are the depth of feature maps. Vertical texts are the xy size of feature maps.

As shown in the illustration in Fig. 7.3, the proposed module uses two path convolutions to extract the features. One pathway is using classical convolutions, shown with the black arrows. At each stage, the classical convolution pathway uses a cascade of two convolution layers, with ReLU as activation function, and zero paddings to keep the xy size. Both convolution layers use 32 3 X 3 classical convolutional layers with ReLU, so that given a 256 X 256 input this pathway produce a feature map whose size is 32 X 256 X 256. Another pathway is using dilated convolutions, shown with the dashed black arrows. At each stage, the dilated convolution pathway uses a cascade of two convolution layers, with ReLU as activation function, and zero padding to keep the xy size. Both convolution layers use

32 3 X 3 dilated convolutional filters, so that produce a 32 X 256 X 256 feature map from a 256 X 256 input. As introduced in last section, the classical convolution pathway results to 3X3 receptive field which extracts the local context, the dilated convolution pathway results to 7 X 7 receptive field (as shown in Fig. 7.2 (d)) which extracts the global context. The feature maps from two pathways are then fused via concatenation, the concatenation is to stack two feature maps together along the direction of depth and not change the xy size. After the concatenation, the fused feature map is 64 X 256 X 256. Finally, the fused feature map is downsampled by 2 X 2 max pooling, which causes the half of xy size, but not change the depth of feature maps.

There are four stages of two-pathway module (Fig. 7.3) in proposed network (Fig. 7.1), all of them use 3 X 3 filters in one pathway and 3 X 3 filters with dilation factor 3 in another pathway. From first stage to fourth, the numbers of filters are 32, 64, 128, and 256 respectively.

The dilated convolution has advantages for accessing non-local context, but it has an inherent problem called 'gridding' effect. As shown in Fig. 7.2 (c)(d) , only locations with non-zero kernel values (red points) are considered, this can cause losing of some neighbouring information, this will be worse when dilation rate increases [326]. There are different methods to solve the 'gridding' effect in dilation convolution from [326-328]. For example, in [326], a Hybrid Dilated Convolution (HDC) is proposed to use different dilation rates in subsequent layers rather than constant dilation rate, so that fully covers a square region without any holes or missing edges. In [327], a pyramid structure is used to concatenate feature maps extracted by dilated convolution with different dilation rates also image-level feature, so that combine multi-scale information without losing neighbour details. In this work, as shown in Fig. 7.3, the two-pathway module concatenate dilated feature maps with normal feature maps so that take the advantages of dilation convolutions but minimize the neighbour information lost in 'gridding' effect.

7.3.3 Upsampling: interpolation and deconvolution

The encoder part of the proposed network has four stages of two-pathway module, the decoder part (expansive path) correspondingly has four stages of skip-connections. The skip connection parts are shown as Fig. 7.4,

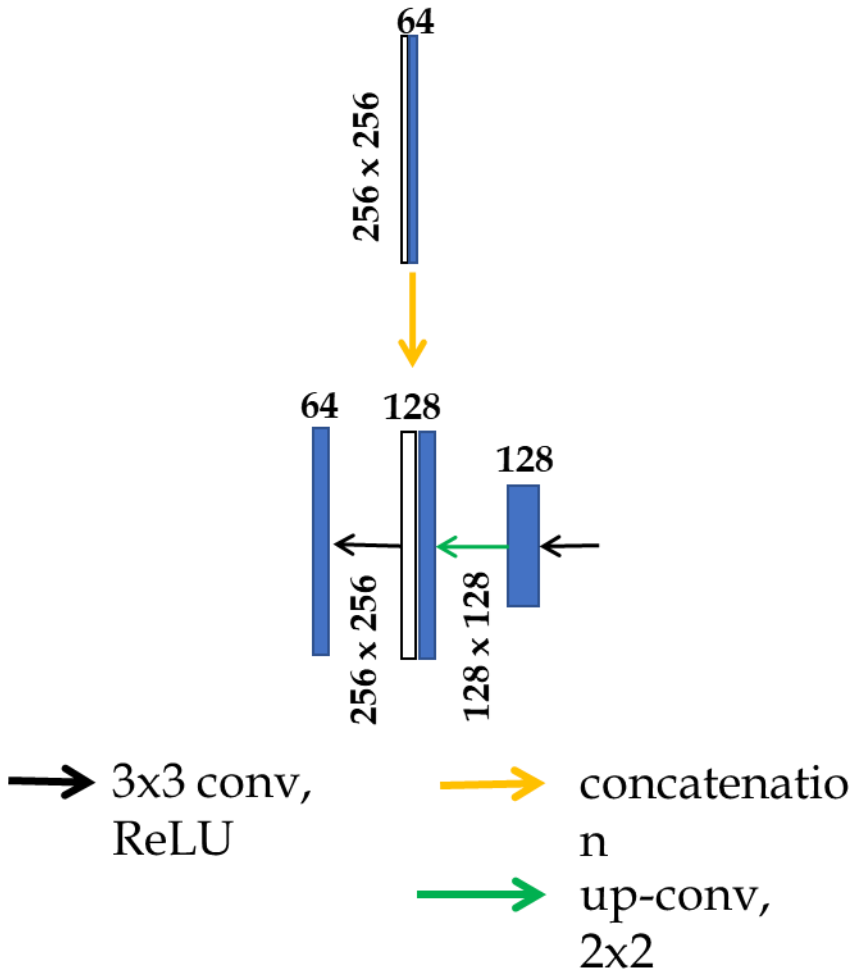


Fig. 7.4 Skip-connections, horizontal texts show channels of feature maps, vertical texts show xy size of feature maps.

In each stage, the skip-connection combines the feature maps in encoder and decoder (shown in Fig. 7.4). The combination is via concatenation (yellow arrow), which is stacking feature maps through the direction of depth of feature maps. The feature maps from encoder are the output of two-pathway module (shown in Fig. 7.3). The feature maps from decoder are the upsampled deep features. For example, the feature maps in encoder have 64 X 256 X 256 size (concatenated from two 32 X 256 X 256 as shown in Fig. 7.3). Correspondingly, the feature maps from decoder for concatenation also have 64 X 256 X 256 size, this feature maps are from the upsampling of deeper stage, shown with the green arrow in Fig. 7.4. The feature maps from deeper stage for example have 128 X 128 X 128 size, the upsampling operation changes it to 64 X 256

X 256, then concatenates with feature maps from encoder, so that form a 128 X 256 X 256 fused feature map.

There are two major ways of upsampling the feature maps in a decoder structure, deconvolution or interpolation followed by convolution. The deconvolution is also called as transposed convolution, given an input matrix, a convolutional filter, and the matrix, the convolution can be regarded as matrix multiplication, where the input matrix can be stretched as a 1-D vector V_{in}, the sliding of convolution filter can be regarded as a sparse matrix M_{conv}, and the output is then a 1-D vector V_{out}. So, the convolution of X is,

$$V_{out} = M_{conv} * {V_{in}}^{T} \tag{7.1}$$

where $*$ means the matrix multiplication, the Eq. 7.1 can be illustrated with Fig.7.5,

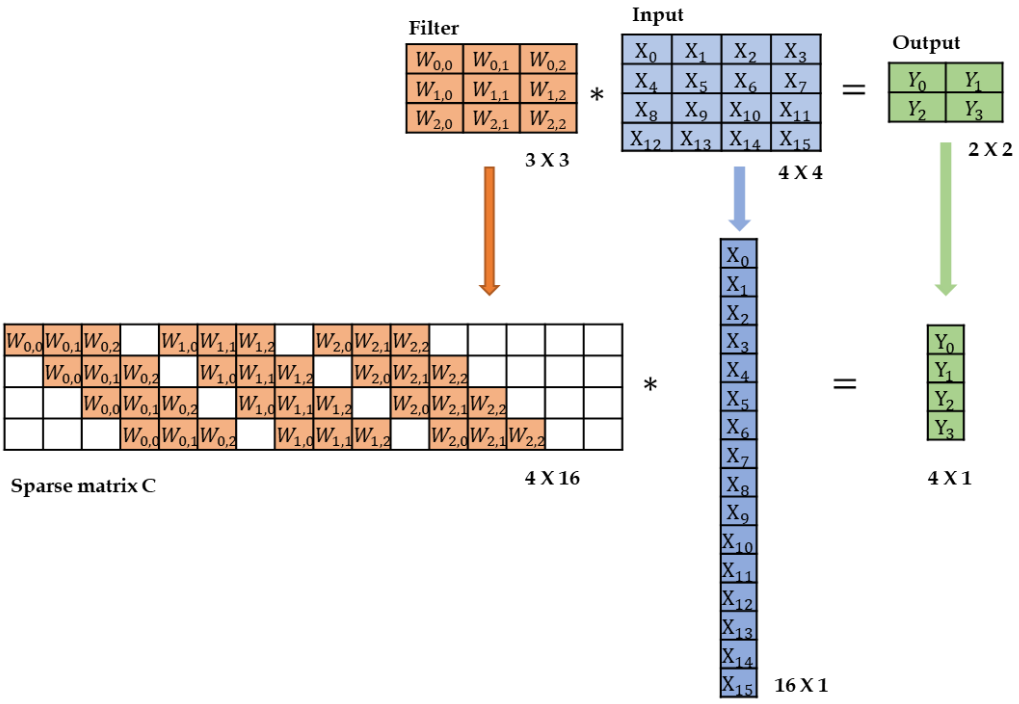


Fig.7.5 Convolution of a 4 X 4 input using 3 X 3 filter The white blanks stand for zero values.

The shows a convolution of filter size 3 X 3 on input size 4 X 4 with output size 2X2. The deconvolution is to apply the transposed matrix M_{conv} on V_{out} to recover the dimension of V_{in}, which is as,

$$V_{in} = {M_{conv}}^{T} * V_{out} \tag{7.2}$$

And the Eq. 7.2 can be illustrated as Fig. 7.6,

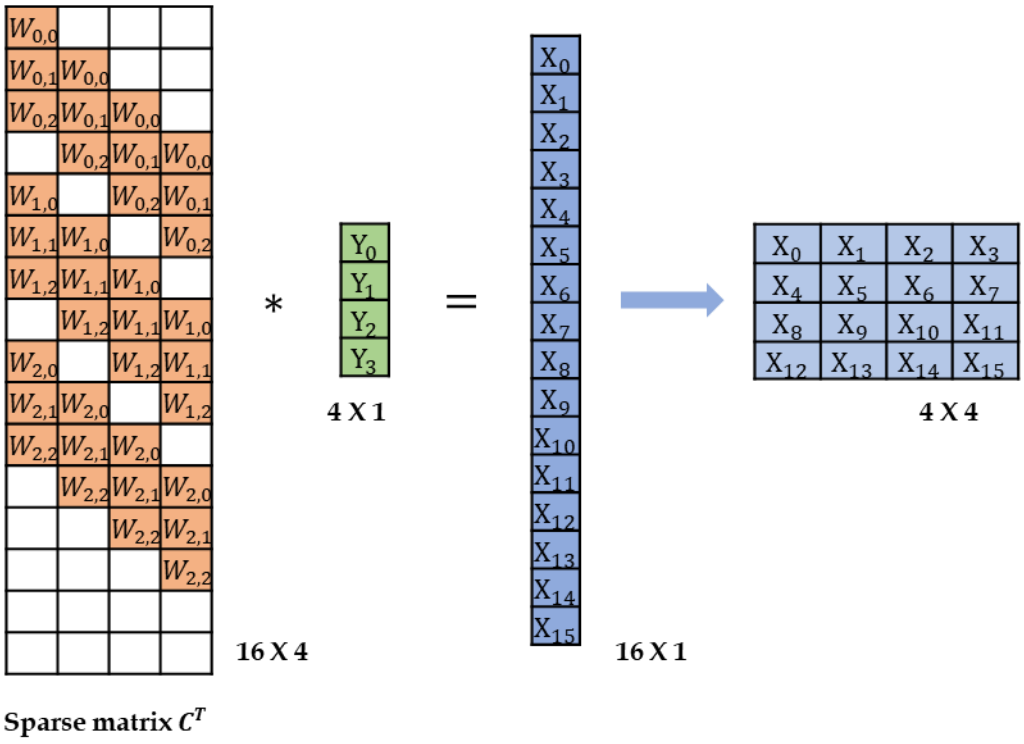


Fig. 7.6 Deconvolution (transposed convolution) of a 2 X 2 (stretched to 4 X 1) matrix to 16 X 1 vector and rearrange to 4 X 4.

Fig. 7.6 shows the recovery of xy dimensions of feature maps via deconvolution. The deconvolution uses the form of transposed sparse matrix M_{conv} and inverse the V_{in} and V_{out}, but it does not mean the values of input V_{in} are exactly recovered, the deconvolution is only used for the upsampling of the feature maps' size, the values of the filter are also trainable. However, the deconvolution can cause checkboard artifacts, which is shown as Fig. 7.8 (a). The artifacts are from the 'uneven overlap', which are in the upsampling putting more of the metaphorical paint in some places than others [329]. The uneven overlap occurs when size of deconvolutional filter cannot be divided by stride. An example is shown in Fig. 7.7,

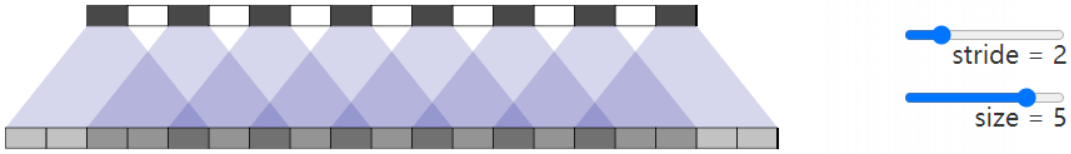
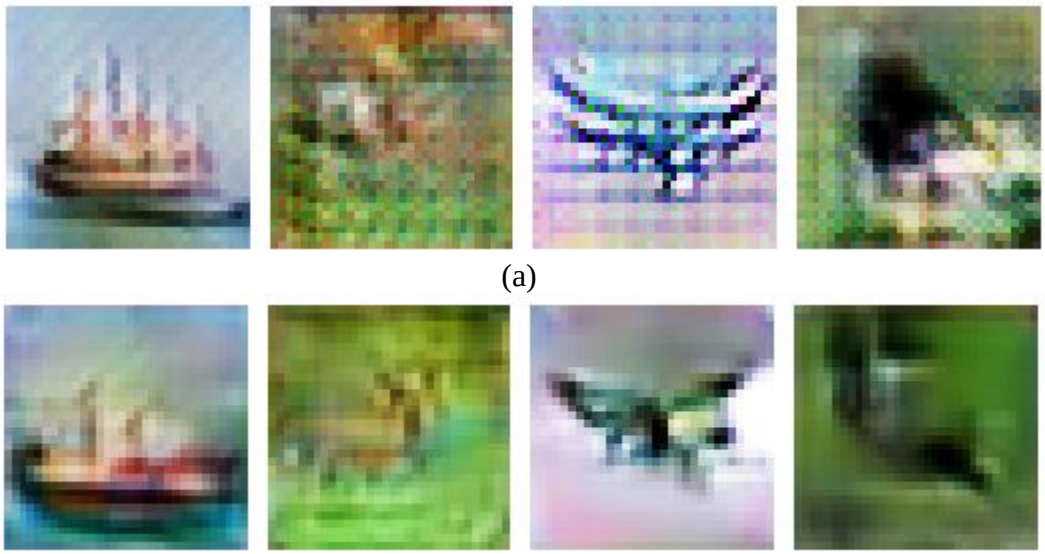


Fig. 7.7 The uneven overlap in deconvolution, with filter size 5 X 5, stride 2 [330]. The upper row is the input vector, the lower row is the upsampled vector.

It can be seen from Fig. 7.7 that when the filter size of deconvolution cannot be divided by the stride, the output vector has uneven overlaps in some upsampled places, some of the places have three times of overlaps during the deconvolution while some only have two or one, this is how the uneven occurs.

The artifacts in upsampling via deconvolution can be minimized by using strides that can be divided by filter size, but nowadays generally the artifacts are solved via replacing the devolution by resizing of feature maps to increase resolutions and followed by normal convolution. The commonly used resizing methods include nearest-neighbour interpolation and bilinear interpolation, and in Fig. 7.8 (b) the resize-convolution upsampling results to no artifacts.



(a)

(b)

Fig. 7.8 Last layers of Generative Adversarial Networks to visualise the checkboard artefacts. (a) Results from deconvolution. (b) Results from resize + convolution [330].

Based on the principles and experimental examples, it can be concluded that the resize and convolution is more suitable for the upsampling of low-resolution feature maps in decoder stages. Thus, in the skip-connection modules of this work (shown in Fig. 7.4), the resize and convolution operations are used instead of deconvolution.

7.3.4 1 X 1 convolution

In image-level classification tasks, convolutional filters with RELU activation are used for extraction of features along with the reduction of xy image size, then the extracted features are flattened and processed by fully connected layers with Sigmoid activation to output the probabilities of classes (as shown in Fig. 7.9). While the proposed network (shown in Fig. 7.1) requires an encoder-decoder structure to maintain the xy size in output prediction, thus here a different kind of classifier is used, which is 1 X 1 convolution.

The 1 X 1 convolution was proposed [331], and were widely used in many networks [332, 333] to manipulate the depth of feather maps. The 1 X 1 convolution uses 1 X 1 filter size so that it is designed for fusion of feature maps between channel without change xy dimension. As shown in Fig. 7.9 (b), the multi-channel feature maps will be fused by a 1 X 1 convolution along the depth direction, this can be understood as a fully connection layer, but does not require the flatten of feature map so that keep the spatial relationships among xy direction. The 1 X 1 convolution is followed by Sigmoid activation to produce dense predictions with probability between 0 – 1, so that achieve the pixel-level segmentation of input image.

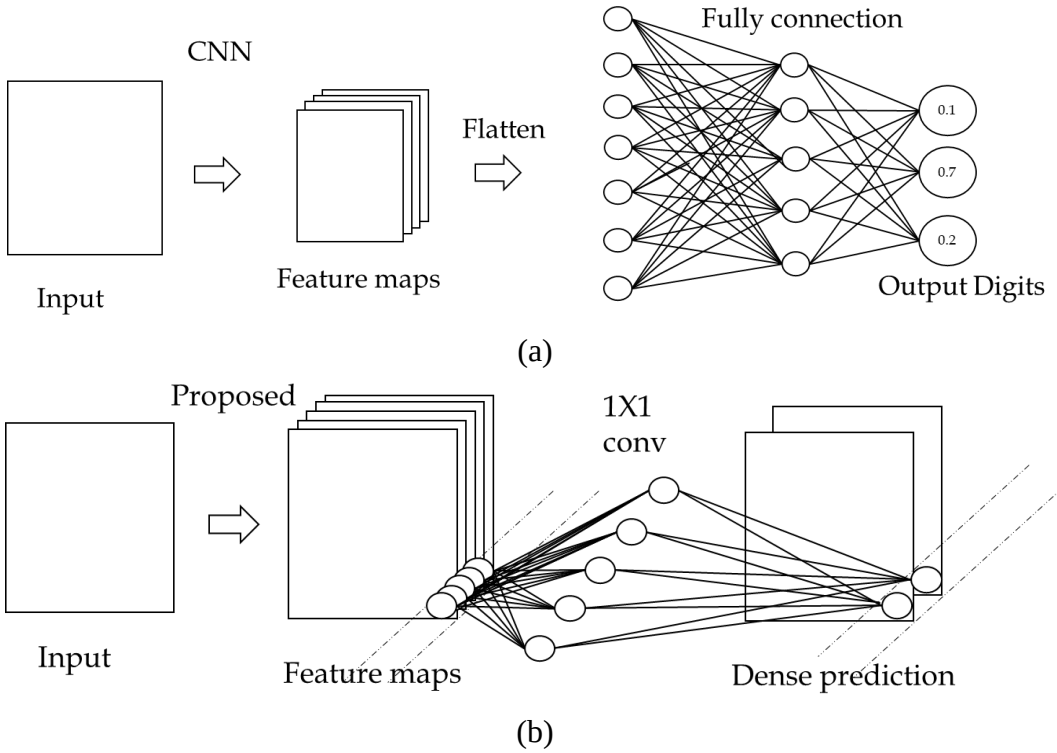


Fig. 7.9 Difference between classifier in normal CNN and proposed modified U-Net. (a) A classical CNN using fully connected layers to output classification. (b) The proposed network uses 1 X 1 convolution to output dense prediction.

7.3.5 Batch normalisation

Similar to many machine learning algorithms, the DCNN also is influenced by the initialization of weights in network, thus there were some researches on the initialization methods of DCNN such as HE, Xavier, and so on (see review in [334]). Also, with the increase of numbers of layers of DCNN, the network is harder to train, problems occur such gradient vanish and gradient explosion [335]. In [336], Internal Covariate Shift (ICS) as the change in the distribution of network activations due to the change in network parameters during training, which constrains the training speed. To solve the ICS, the batch normalisation (BN) is proposed to accelerate the DCNN training.

The BN works along with mini-batch training, which is dividing train data into several mini-batches and update the parameters of DCNN by using the mean gradients inside the mini-batch. The BN is to add an additional process before the output of a layer be activated and fed into next layer. The BN is given as

$$\mathrm{BN}_{\gamma_B,\beta_B}(x_i) = \gamma_B \hat{x}_i + \beta_B \tag{7.3}$$

where γ_B and β_B are trainable parameters to scale and shift the distribution of $\hat{x}_i$, where $\hat{x}_i$ is normalization of values in the mini batch, given a mini-batch $\mathcal{B} = \{x_{1\ldots m}\}$, the $\hat{x}_i$ is calculated as,

$$\begin{aligned} \mu_{\mathcal{B}} &\leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \\ \sigma_{\mathcal{B}}^2 &\leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \\ \hat{x}_i &\leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \end{aligned} \tag{7.4}$$

Where m is the number of data samples in the mini-batch $\mathcal{B}$, $\mu_{\mathcal{B}}$ is the mean of this mini-batch, $\sigma_{\mathcal{B}}^2$ is the variance, thus $\hat{x}_i$ is the normalized version original values x_i in the mini-batch. The $\hat{x}_i$ is scaled and shifted via Eq. 7.3, finally the $\mathrm{BN}_{\gamma_B,\beta_B}(x_i)$ is passed to activation function to next layer rather than x_i.

The BN not only enables higher learning rates of DCNN, but also regularizes the model [336], this makes it can be an alternative approaches of Drop-Out. The U-Net [296] used Drop-out to minimize the overfitting, in the proposed network batch normalization is used which can also help regularization. The mini-batch size is set as 2, as the data is limited, the BN will work better when data and model size increase but requires larger GPU memory.

7.3.6 Definition of Dice loss

As introduced in Chapter 3, the Cross Entropy (CE) is a general used Loss function to train a DCNN which aims to classification, this include image-level classification, and pixel-level classification, which is semantic segmentation. The cross entropy between two N sized discrete distributions $\mathbf{p} \in [0,1]^N$ and $\mathbf{q} \in [0,1]^N$ is given as,

$$H(\mathrm{p},\mathrm{q}) = -\sum_{i=1}^{N} \mathrm{p}_i \log\, \mathrm{q}_i \tag{7.5}$$

where $\mathbf{p}$ is the classification target, $\mathbf{q}$ is the predicted likelihood per class. The $H(\mathbf{p},\mathbf{q})$ can be translated as a loss function, where N represents the number of classes, $\mathbf{p}_i$ is

the ground truth (label) of i^{th} class, and $\mathbf{q}_i$ is the model prediction of this class. CE loss of the HNC segmentation problems can be given as,

$$L_{ce} = -\sum_{i=1}^{2} \mathrm{p}_i \log\ \mathrm{q}_i \tag{7.6}$$

$$= -[p\log(q) + (1-p)\log(1-q)]$$

Because here are 2 classes in this work: HNC and background. In the binary classification, CE loss L_{ce} treat the foreground and background in a symmetric way (shown in Eq. 7.6). However, in many medical segmentation problems include the HNC extraction from MRI data, the foreground (tumours, vessels and so on) and background are imbalance, normally the foreground targets are small compared to the background and whole image, which makes the CE loss cannot solely solve the segmentation well because the background can dominate the gradients in training.

Some methods were proposed to solve the segmentation of imbalance data, such as Weighted Cross Entropy and Focal Loss [303]. In U-Net [296], an edge map is introduced to CE to improve the segmentation of cells. In this work, the Dice Loss from [305] is introduced to balance the weights between HNC and background in MRI slices. The Dice Loss comes from Dice coefficient, and in DCNN training, they are given as,

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{7.7}$$

$$L_{dice} = 1 - D$$

Where p_i is prediction (probability between [0,1]) of i^{th} pixel and g_i is its ground truth (binary digits). The definition Dice Loss L_{dice} mainly focus on the foreground, so that minimize the influence of large proportion of background. While in practical usage, the L_{dice} is not steady and sometimes not converge well, thus finally the proposed network uses combined loss given as,

$$L = L_{dice} + L_{ce} \tag{7.8}$$

Which is the combination of L_{ce} to keep the training steady and L_{dice} to minimize the inter-classes imbalances existed in HNC MRI slices.

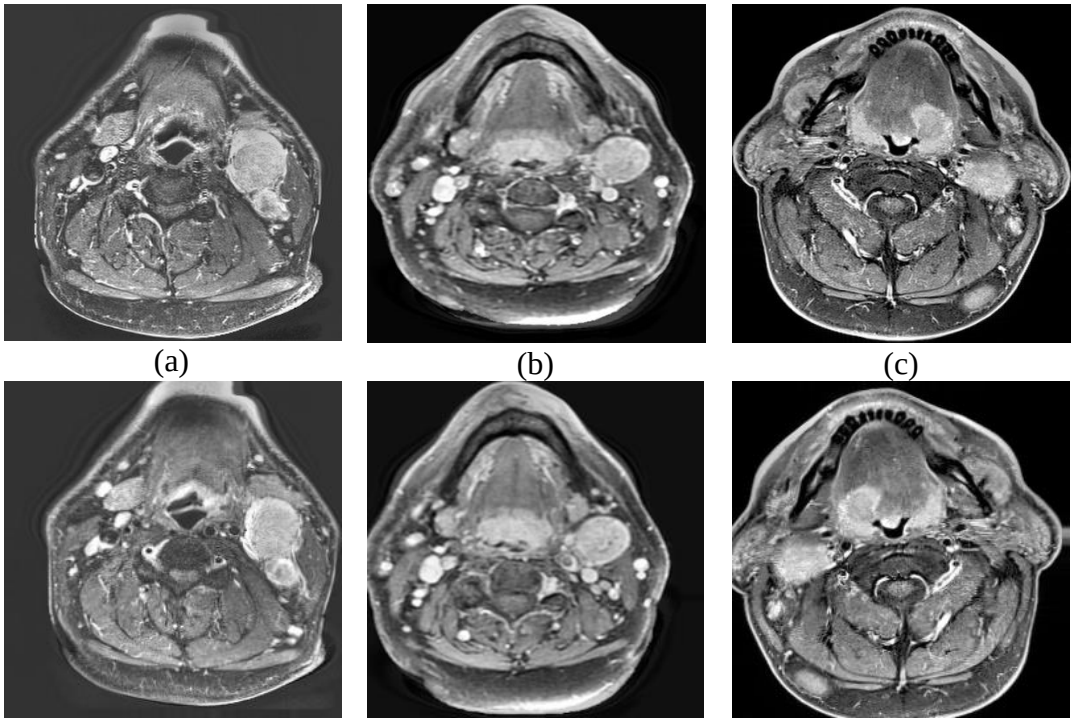
7.3.7 Data augmentation

The convolution and pooling operation gives DCNN abilities of invariant to small shift and scale transformation [289]. While, there are still gaps to work on to improve invariant properties of feature extraction in DCNN, such as works in [337-340]. In U-Net [296], the data augmentation is used to teach the network invariance and robustness properties, this will be more important when number of training samples are limited. Data augmentation is also introduced into this work, the setting of the augmentation is given as,

Table 7.1 Setting of data augmentation of proposed network

Rotation (Degree)	**Horizontal Shift**	**Vertical Shift**	**Shearing**	**Zoom**	**Horizontal Flip**
0.2	±0.05	±0.05	±0.05	±0.05	True

As shown in Table 7.1, there are six types of augmentation in total. During the training, additional training samples will be generated (along with mask) to feed into the training network. The generated samples are the transform. The bar graph shows that in half above datasets the proposed method performs bettered versions of original train data, and the transformations are the random combination from Table 7.1. The numbers in Table 7.1 are smaller than 1, represent the fraction, for example, the horizontal shift is between [-0.05, +0.05] of total width of the image. In shift, shearing, and zoom (shrink), there will be spaces (non-value points) occurs, and they will be filled with values of nearest neighbour.



(a) (b) (c)

(d) (e) (f)

Fig. 7.10 Examples of data augmentation. (a-c) are original images. (d-f) are generated images.

As shown in Fig. 7.10, the generated samples are transformed versions of original training samples. Some transformations can be clearly found in these examples, such as shift in Fig. 7.10 (a)(b), and horizontal flip in Fig. 7.10 (c). The data augmentation generates additional training samples along with the training epochs forwarding, hundreds of original samples run by dozens of epochs can generated thousands of augmentations, these additional data will improve the robustness and invariant properties of proposed segmentation network.

7.4 Experimental results

The proposed modified U-Net was implemented with Keras library of Python, running on PC with 16G RAM, 3.2GHz Intel(R) Core (TM) i7-8700 CPU, and a NVIDIA GTX 1070 GPU. Experiments were on real MRI data form Beatson west Scotland cancer centre. A total of 163 images (2D slices) from 17 patients were used in this work. The ground truth labels for training and validation are consensus manual delineation provided by clinicians from Beatson. Weights of proposed network are

optimized using Adam [341] with 1e-4 learning rate without weight decay, the Loss for optimization is given in Eq. 7.6 – Eq. 7.8.

The limited numbers of data used in training of DCNN may lead to overfitting (see Chapter 3), data augmentation can partially reduce the occurrence of this problem. In this work, the cross validations are used to avoid the overfitting, and in some extent show the ability of generalization of the proposed deep neural network.

The k-fold cross validation is to equally (or nearly equally) partition all data into k groups (folds), then conduct k times of training and validation. In data mining, ten-folds cross validation is most common [342], while it also depends on how many data are available. In this work, 3-fold cross validation is used, an illustration is given as Fig. 7.11,

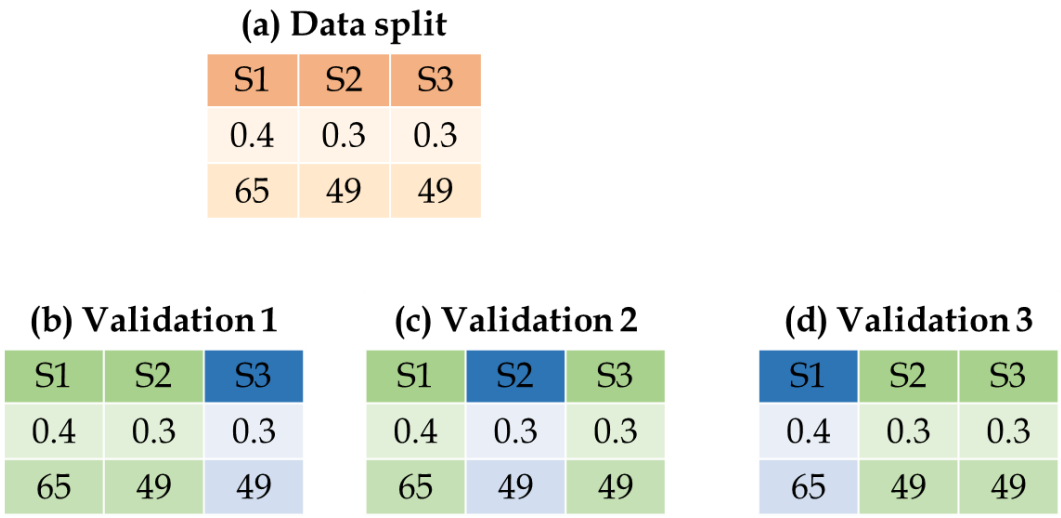


Fig. 7.11 3-fold cross validation of HNC MRI slices. (a) shows the split of data into segments. (b-d) show three times of validations, where greens represent segments used for training, blues represents the segments used for testing.

As shown in Fig. 7.11, the 163 HNC MRI slices are grouped into 3 segments S1, S2, and S3 with the proportions of 0.4, 0.3, 0.3, which are 65, 49, 49 slices respectively. The 3-fold cross validations are then finished by 3 times training and testing, first time S1, S2 for training, S3 for testing; second time S1, S3 for training, S2 for testing; third time S2, S3 for training, S1 for testing.

With this 3-fold cross validation setting, the proposed modified U-Net is evaluated by the performance compared to U-Net, the comparison between them measured by Dice Score [314] is given as,

Table 7.2 HNC segmentation performance comparison

Methods	**3-fold cross-validation**		
	Validation 1	**Validation 2**	**Validation 3**
U-Net	0.6230	0.5576	0.5954
Proposed	0.6735	0.6076	0.6510

Table 7.2 shows that the proposed modified U-Net has 0.644 mean Dice Score across three times cross validations, which is about 0.05 higher than the Dice Score of classical U-Net through 3-fold cross validation. Here are more visualisation comparisons between two methods shown as Fig. 7.12,

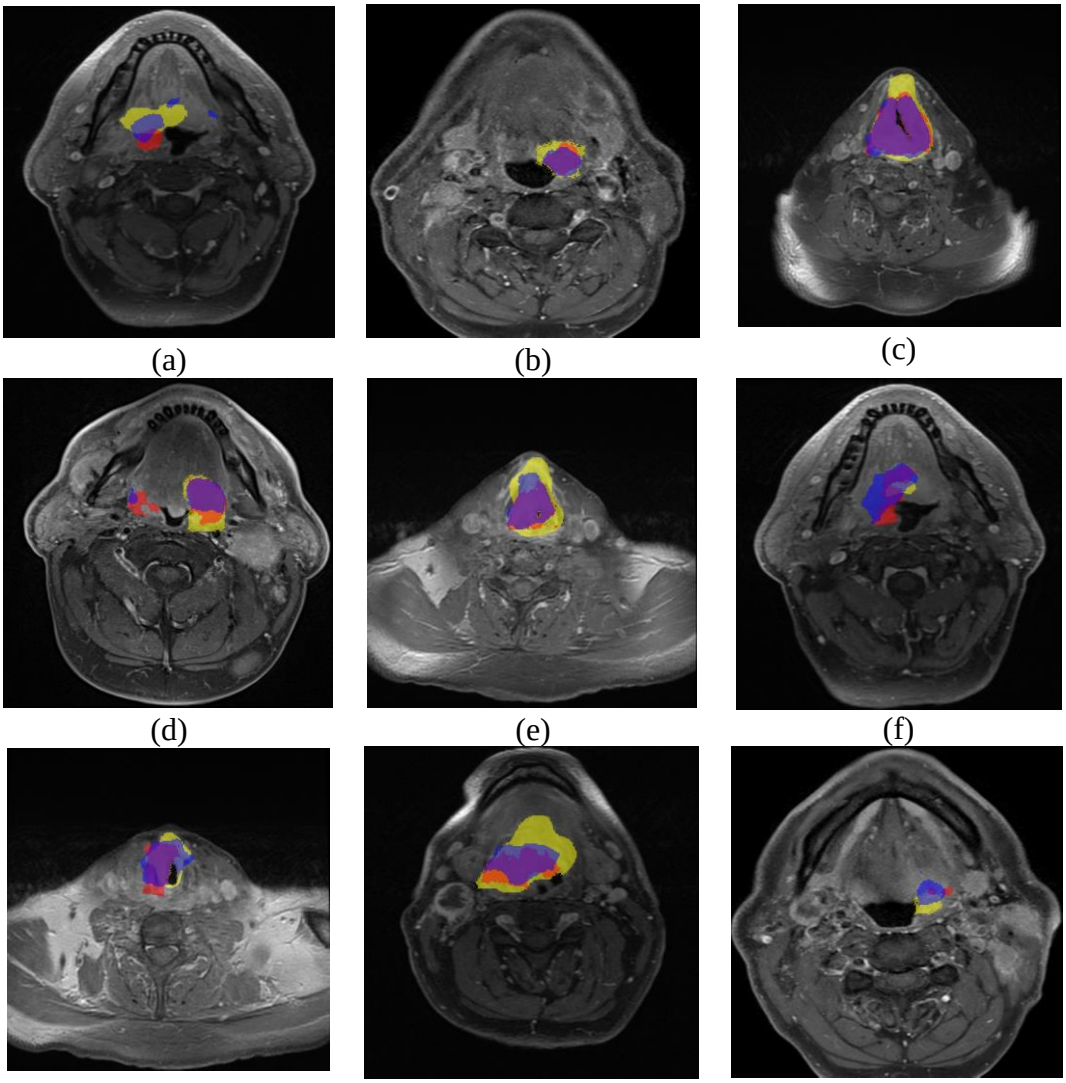


Fig. 7.12 Visualisation of segmentations of HNC from consensus manual delineations (yellow area), U-Net (red area), and proposed method (blue area). The images are from various patients.

Fig. 7.12 demonstrates several examples of the comparison of segmentation results between original U-Net and proposed modified U-Net. From (b)(c)(e)(g), the proposed network has segmentations better overlapped with gold standard. In some slices, U-Net or proposed network, or both have outliers' segmentations. Such as in (d), the tumour area (yellow) is on the right side of throat, but here are 'skipped' segmentation (red and blue) on the left side of throat. This may disappear when the segmentation algorithm considers inter-slices information, but the network is segmenting targets

only based on single slice input. There are more visual examples of results in Appendix D.

7.5. Conclusion

This chapter introduced a novel deep learning based automatic segmentation algorithm for HNC extraction from MRI slices. The proposed algorithm is a modified version of classical DCNN called U-Net. The modification is based on the motivation of better feature fusion between low-level and high-level features, but also local and non-local features. The performances of the U-Net and proposed network are cross validated on real MRI data from Beatson West Scotland Centre. The data is 163 slices in 2D form collected from 17 patients. The cross validation shows the proposed network can have about 0.05 higher dice score compared to original U-Net.

Through the development of the modified U-Net, several pros and cons of deep learning-based approach can be concluded. The first advantage of DL methods is that it is end-to-end structure combining feature extraction and classifier together and get rid of handcraft feature extractor designing. Another advantage of DL is that the algorithms can obtain robustness with the increase of data, rather than introducing significant modifications. While this reminds of first drawback of DL models, as they are data hungry, insufficient data cannot train a deep neural network well, but the deeper is better has been widely accepted in DL models' design. Also, the training of DL models requires computation power and time cost. The proposed network is 2D based for two reasons: 1) The raw MRI data and labels are 2D, where the data can be interpolated but the label cannot be accurately reconstructed to 3D. 2) The training of 3D models requires much higher memory of GPU, which is not available during the development the algorithm. Another challenge of DL is the difficulties of labelling medical data which highly relies on clinicians' expertise. The results of DL model cannot be exactly interpreted, and cannot be interacted midway, which may add its difficulties to be used in practical applications shortly.

The proposed network utilises single slices for prediction of tumour contours, the potential future work can be at the involving of 3D or inter-slices information. Also,

more data is awaiting to strengthen the model. Other methods such as transfer learning could be considered for the improvement of DL model performance with limited data.

Chapter 8 Conclusion and Future Work

8.1 Conclusion

This thesis explored various image processing approaches to tackle the several challenges accurate RTP. The major work of this thesis focuses on the automatic detection and segmentation of cancerous tumours and abnormal lymph nodes in head and neck region to provide objective contours to radiotherapist.

In the RTP process, there are existing challenges in the manual delineations of GTV. The challenges include the inter- and intra- subjectiveness from manual segmentation of GTV, and the demands of clinical expertise and high time cost.

In this thesis, firstly a novel knowledge-based abnormal lymph nodes detection and segmentation algorithm is presented. With real 2D T1-weighted MRI slices as input, the proposed algorithm goes through a pipeline of pre-processing, detection, segmentation, and post-processing. In pre-processing stages, the contrast of image is enhanced, then bias field is removed, an intensity standardisation is applied between slices to adjust the histogram distribution, finally Fourier interpolation is used to reconstruct 2D slices into 3D head and neck MRI volume. The detection algorithm is built on the knowledge of anatomical structure of head and neck region on single slice, firstly throat is automatically detected to act as an important marker, then a modified fuzzy c-means is used to clustering pixels into groups based on intensities and distance to throat, a knowledge-based scheme is then proposed to refine the pixels to find the rough detection of abnormal lymph node. The detections on each slice are fused through a majority voting process, to finally determine the rough 3D localization of abnormal lymph node inside head and neck MRI volume. The segmentation uses 3D level set method algorithm which starts from automatic detected ALN and evolves under guidance of level set function to track the surface of ALN volume, finally the algorithm will converge so that get a segmentation of ALN. The segmentation results will be post processed using 3D morphological operation to remove outliers and smooth the segmented volume.

The second achievement in this thesis is developing a novel knowledge-based algorithm for the automatic 3D detection and segmentation of cancerous head and neck tumours. This proposed algorithm has similar pipeline with ALN work, the differences are in detection and segmentation phases. The detection is majorly conducted on central slices. Apart from the fuzzy c-mean, the watershed algorithms and localized region-based LSM are used to further supress the false positives in detection. In segmentation phase, a novel spatial-constrained 3D LSM is proposed to find the surface of tumour volumes, the constrains are from the spatial information deduced from detection phase.

The both knowledge-based 3D segmentation algorithms achieved several goals of head and neck images analyse and tackled some difficulties in RTP. Firstly, the proposed algorithms can reconstruct 3D MRI with enhanced image quality based on input 2D MRI slices. This can produce extra 3D information to clinicians as well as engineers when the scanners can only provide limited spatial resolution. Secondly, the knowledge-based automated the process of clinicians identifying the ALNs and tumours from MRI slices and provided intermediate as well as final clinical targets to clinicians include locations, contours, and volumes of throat, abnormal lymph nodes, and tumours. In addition, the proposed algorithms can provide 3D visualisations of ALNs and tumours with clear and smooth boundaries, and thus combined with scanner parameters the quantifications of ALNs and tumours can be provided. 3D visualisation and quantifications of these clinical targets can help the planning of radiation treatment, also track the efficiency of treatment so that adjust the adaptive planning.

This thesis also explored the deep learning methods for end-to-end supervised semantic segmentation of head and neck cancerous tumours from real MRI slices. A novel deep neural network which is modified on the classical U-Net structure. The modifications are majorly on introducing a two-pathway structure and dilated convolution to enlarge the view of context fusion. The novel achievement of this work is the HNC segmentation with improved accuracy compared to the original U-Net structure. This was completed with limited data and GPUs for training. The proposed deep learning approach releases the burden of designing handcraft features and subjective knowledges, also provide end-to-end structure to generate final pixel-wise

prediction with direct input. This deep learning method has challenges in data and computational power demanding, also it cannot be as well interpretated to clinicians as knowledge-based approaches. However, the deep learning method has potentials in coping with larger amounts of data and improving the performance rather than degrading, this will be an advantage compared to convention method. Also, this network will be easier to expand to multi-modality input and multi-class output applications compared to conventional methods. Thus, there are potential works and benefits to be expected in deep learning methods.

8.2 Future work

The possible future works of this thesis are listed as follows:

1) **Train networks with more data:** the ability of prediction of a neural network will be improved by using deeper structures, which involves much more data and needs more data to train sufficiently. As shown in Chapter 7, currently less than 200 images are used for training and testing, the future work can be conducted with more HNC MRI data to train a deeper network to improve the accuracy of tumour segmentation.
2) **Evaluate the proposed algorithms on larger dataset:** currently the ALN segmentation algorithm is validated on 5 patients (Chapter 5), HNC segmentation algorithm is validated on 10 patients (Chapter 6), the DNN method uses 170 images for cross-validation (Chapter 7). It would be beneficial if more HN MRI data in different TNM stages and modalities can be used to re-evaluate the proposed algorithms and test the algorithms in clinical environments.
3) **Improve the usage of inter-slices information:** when clinicians meet challenging cases of contouring on one slice, a solution is to check former and latter slices to find clues. Currently, the proposed works (Chapter 5, 6) uses 2D detections and fuses them as a 3D detection. This concept could be used in Chapter 6 where more slices can be involved for detection rather than use central slice solely.
4) **Combine the deep learning and conventional approaches:** the advantages in deep learning such as abilities of detection, feature extraction can be fused in the initial and evolution phases of deformable model to give more guidance to model,

and the deformable model has advantage in better boundary tracking, thus the combination should be helpful in future work.

Appendix

A: Windows based software for automatic 3D segmentation of head and neck tumour

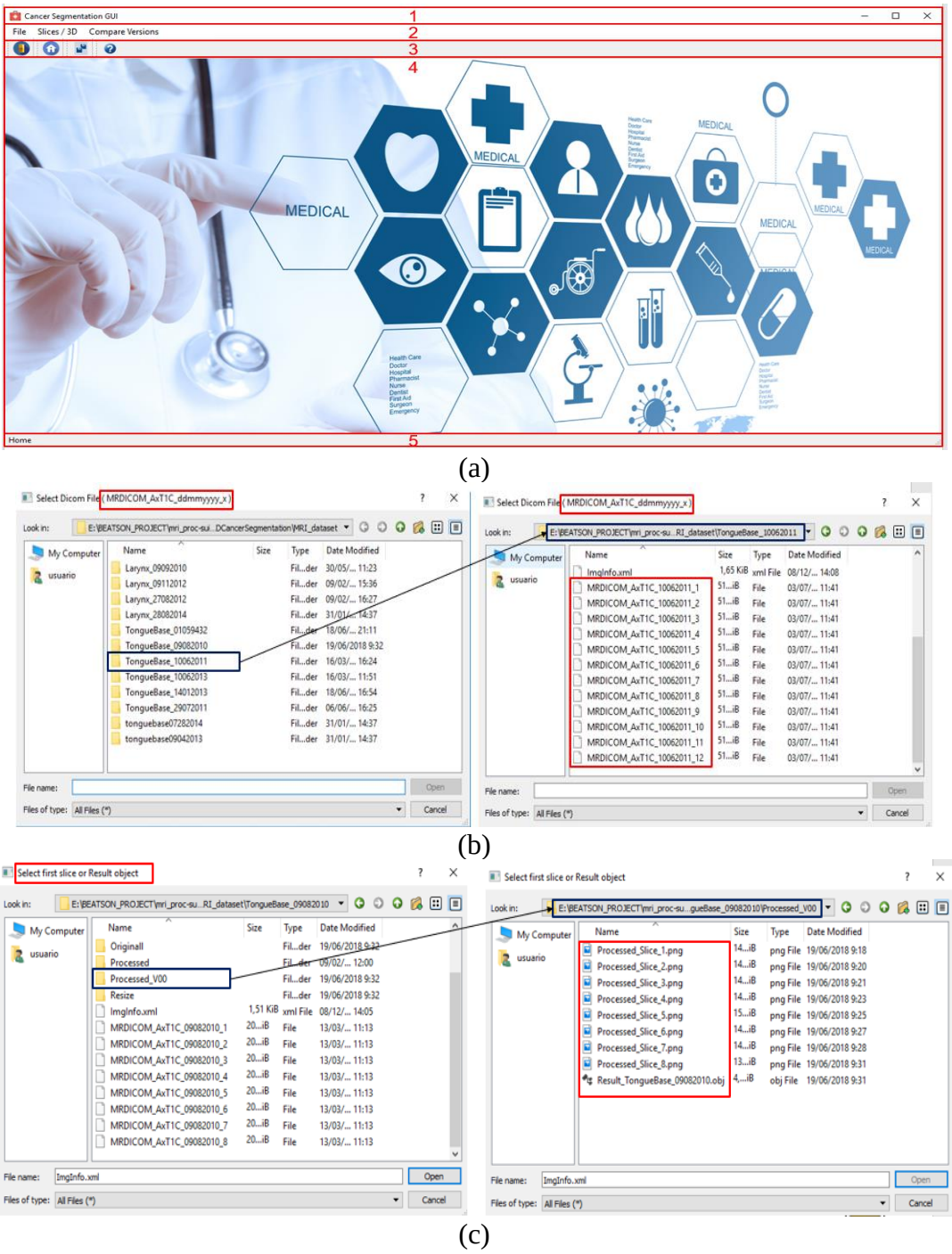
As discussed in Section 6.7, we have further 3D automatic segmentation algorithm to the development of windows-based software for automatic 3D segmentation of HNC from T1 axial MRI data.

The software has function in:

1) Load and save MRI data as well as delineations
2) Automatic and manually delineate HNC contours
3) Modify unsatisfied delineations
4) Reconstruction of 3D HNC volume
5) Quantitative measure distance in image and volume of HNC
6) Visualisation of 2D and 3D HNC

These functions will be demonstrated with screenshots of the software. In Fig A.1, the main menu (Fig. A.1 (a)), data loading menu (Fig. A.1 (b)) and delineations loading menu (Fig. A.1 (c)) is shown. In main menu, the five red rectangular area represent Title (1), Collapse Menus (2), Toolbar Menu (3), Main Window (4), and Status Bar Window (5). The Collapse Menus include file management system (load and save), slices functions (View and Edit processed data), and compare 3D volumes. The Toolbar Menu includes function of save project, open help window, go to home window. The Main Window display the being used window. The Status Bar gives information of the Main Window.

Fig A.1 (b) shows the loading of data, where the black box in left window marks the selected folder and the files (DICOM MRI) in this folder is shown in right window. Fig A.1 (c) shows the loading of processed data, where the black box in left window marks the selected folder and the files (delineations) in this folder is shown in right window.



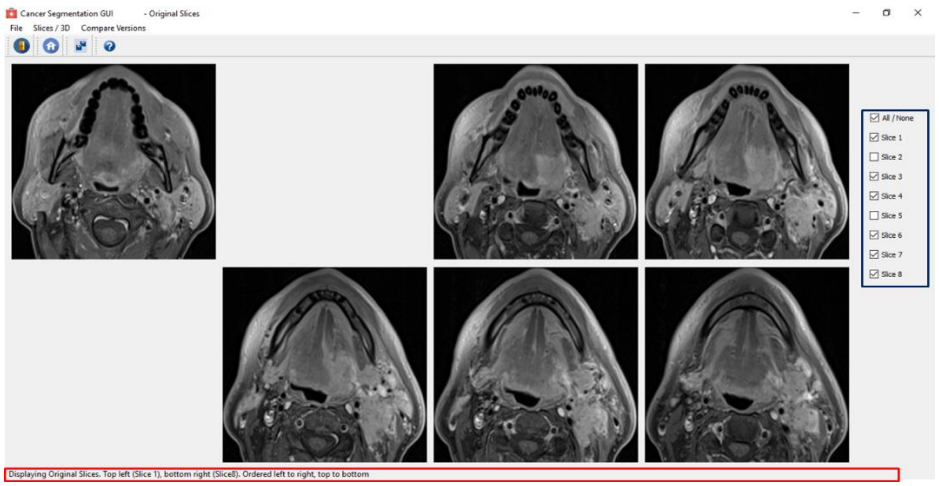
(a)

(b)

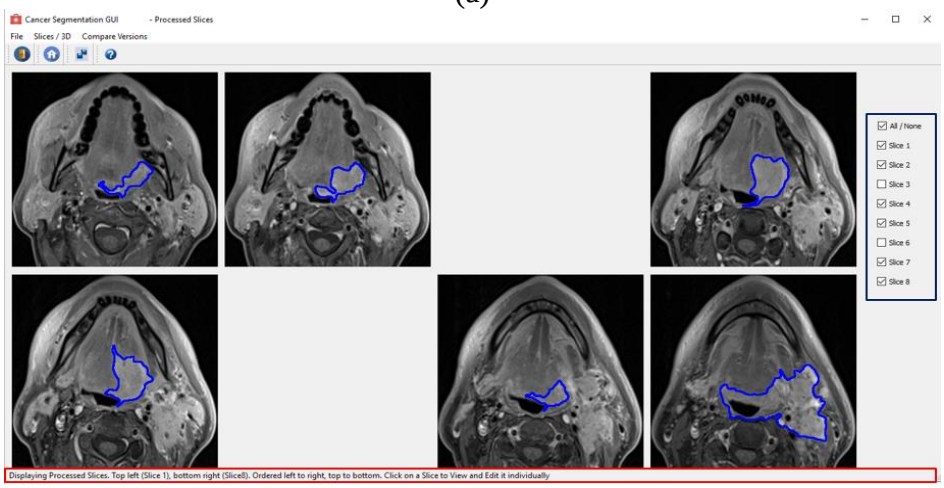
(c)

Figure A.1 RTP Aid auto-segmentation software. (a) Home window of software. (b) Window of loading data. (c) Window of loading processed delineations.

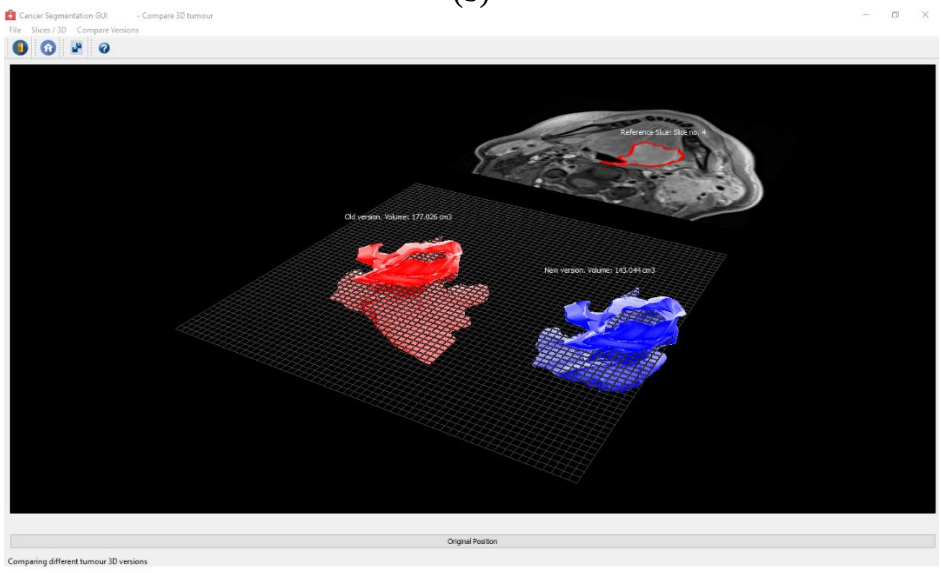
In Fig. A.2, the visualisation functions of this software are shown.



(a)



(b)



(c)

Fig. A.2 Visualisation functions of RTP Aid tool. (a) View original slices. (b) View 2D processed slices. (c) View 3D reconstruction from 2D slices.

Fig. A.2 shows the visualisation functions of RTP Aid Tool, where 2D slices before (Fig. A.2 (a)) and after (Fig. A.2) processing can be selected and viewed, and 3D volume (Fig. A.2 (c)) comparation of reconstructions from different 2D delineations can be visualised.

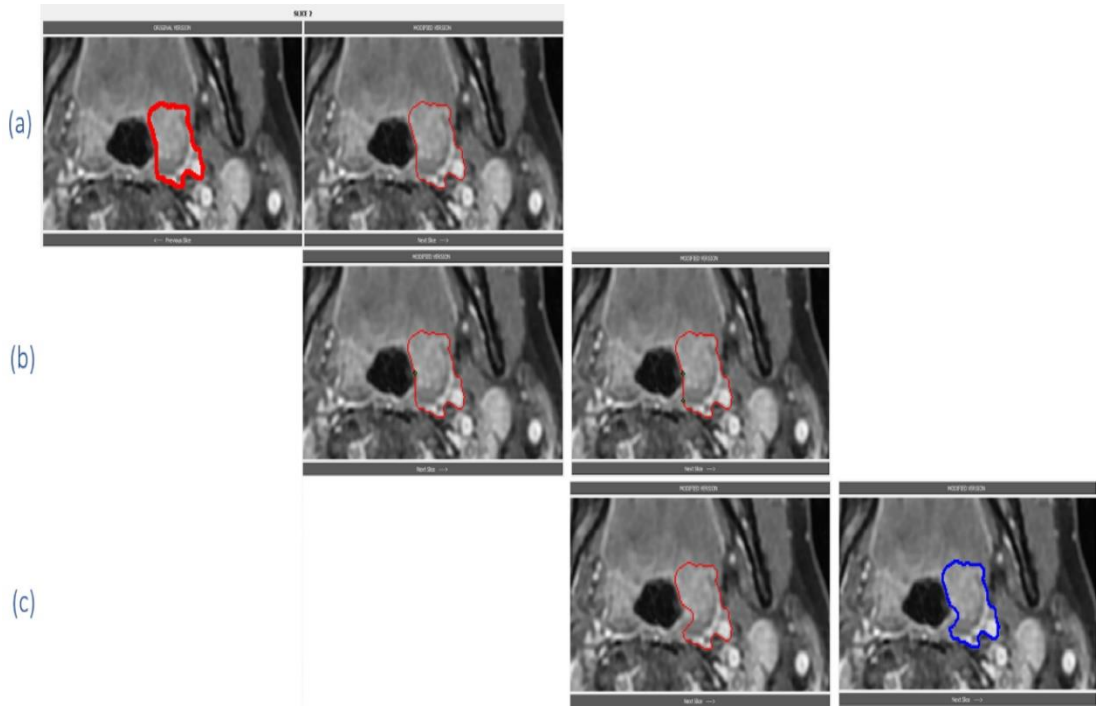


Fig. A.3 Manual modification of existed delineations. (a) Existed contour. (b) Modifying by dragging. (c) Modified contour

In Fig. A.3 the modification of existed delineation (thick red in Fig. A.3 (a)) by dragging of curve (in Fig. A.3 (b)) and achieve a modified delineation (thick blue in Fig. A.3 (c)).

B: MRI dataset with head and neck cancer

In Chapter 4, examples of MRI data and delineations from COs have been demonstrated, more examples will be given in this section.

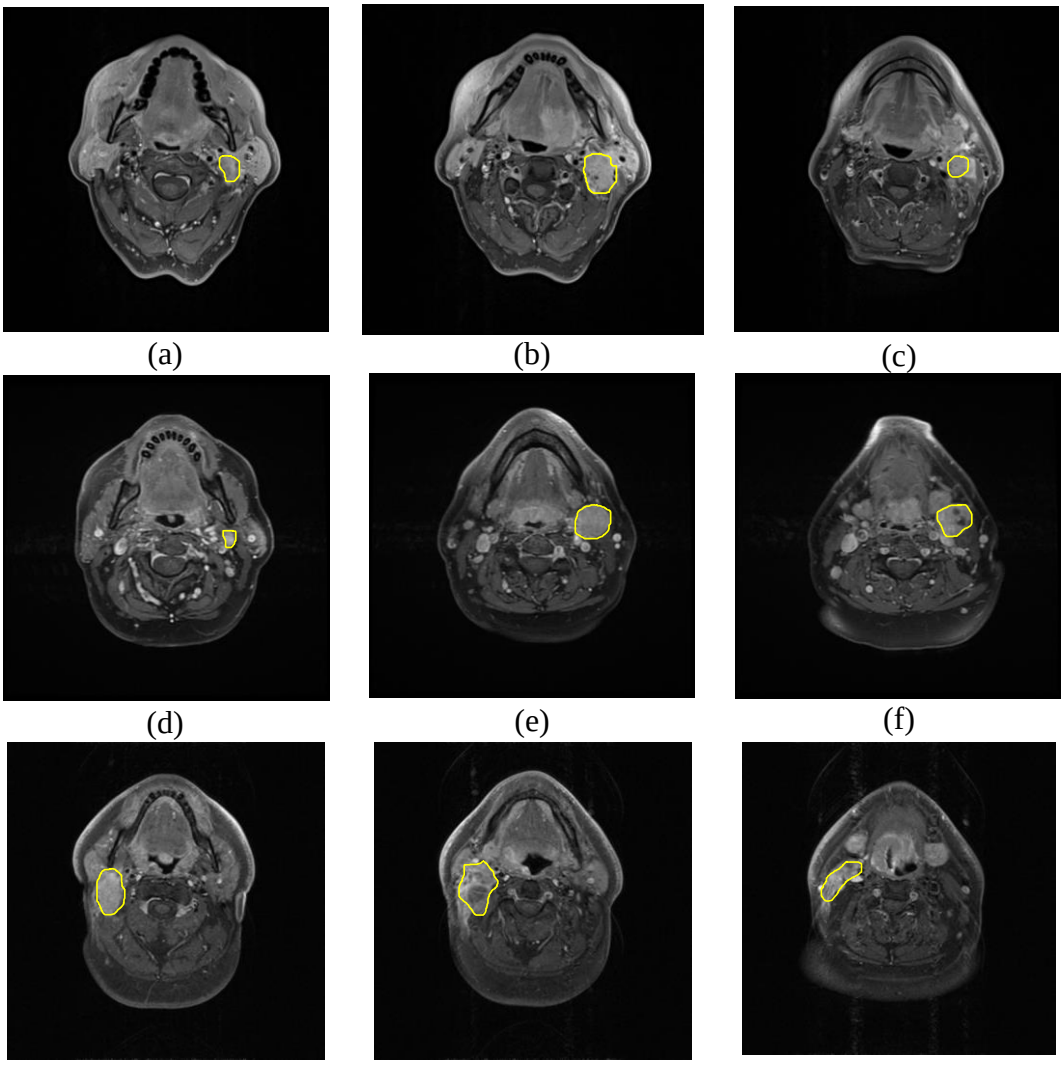


Fig. A.4 Examples of T1 axial MRI slices with ALNs marked by yellow. (a-c) are slices from top to bottom order of first patient. (d-f) are from second patient. (g-i) are from third patient.

Fig. A.4 shows examples of T1 axial MRI slices with labelled ALNs, Fig. A.5 displays MRI slices with labelled HNC.

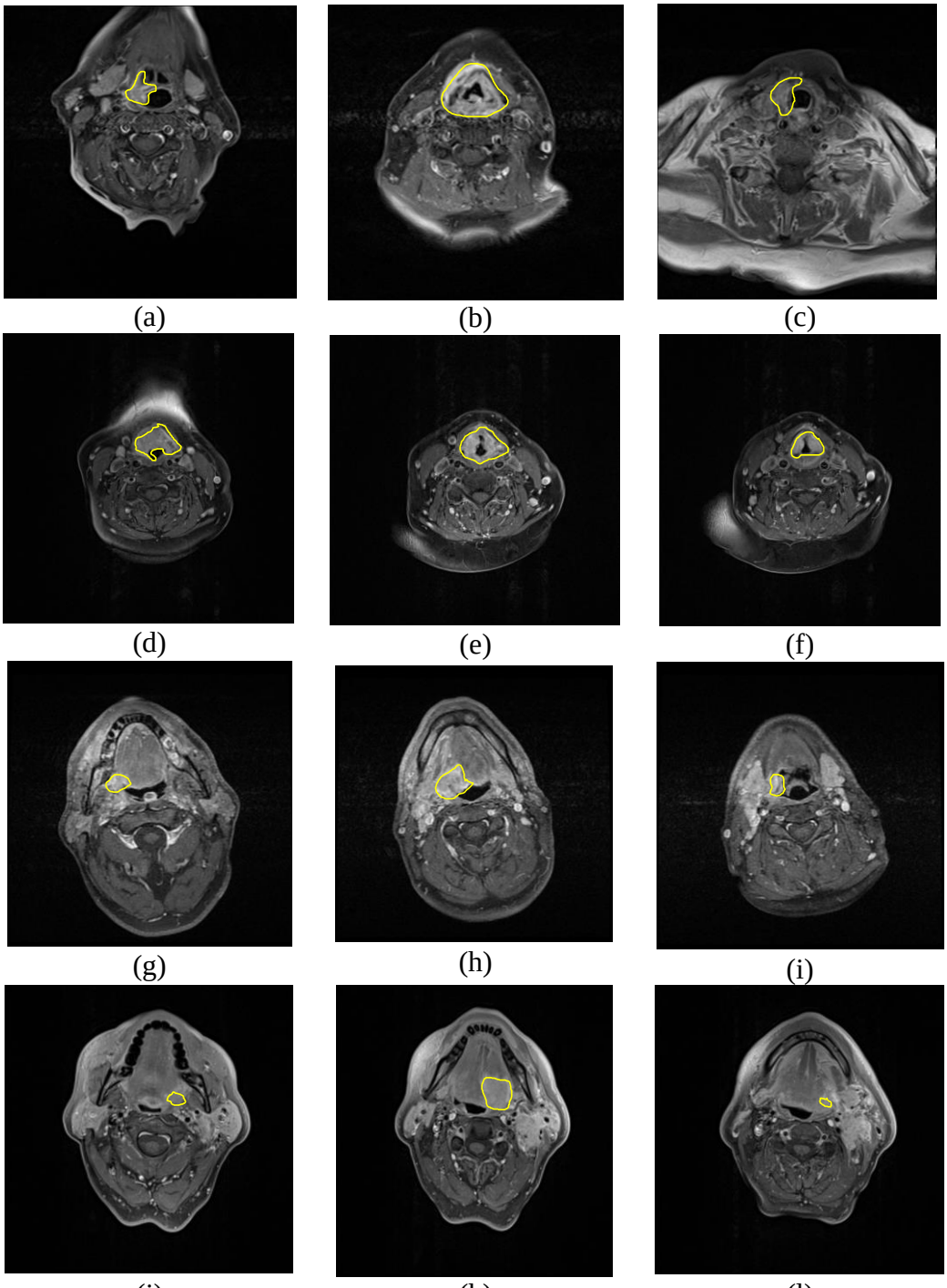


Fig. A.5 Examples of T1 axial MRI slices with HNC marked by yellow. (a-c) are slices from top to bottom order of first patient. (d-f) are from second patient. (g-i) are from third patient. (j-l) are from fourth patient.

C: Automatic 3D segmentation results of ALNs and HNC

The ALNs and HNC segmentation results have been visualized and quantitative analysed in Section 5.5 and Section 5.6. More examples of results are shown in Fig. A.6 and Fig. A.7.

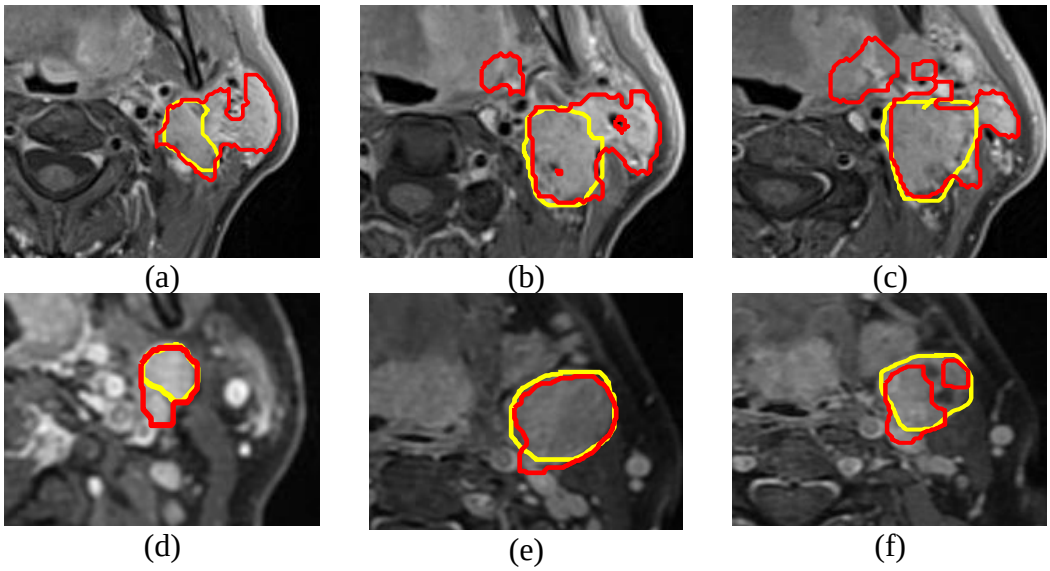


Fig. A.6 Examples of T1 axial MRI slices with ALNs drawn by manual (yellow) and automatic algorithm (red). (a-c) are from first patient. (d-f) are from second patient

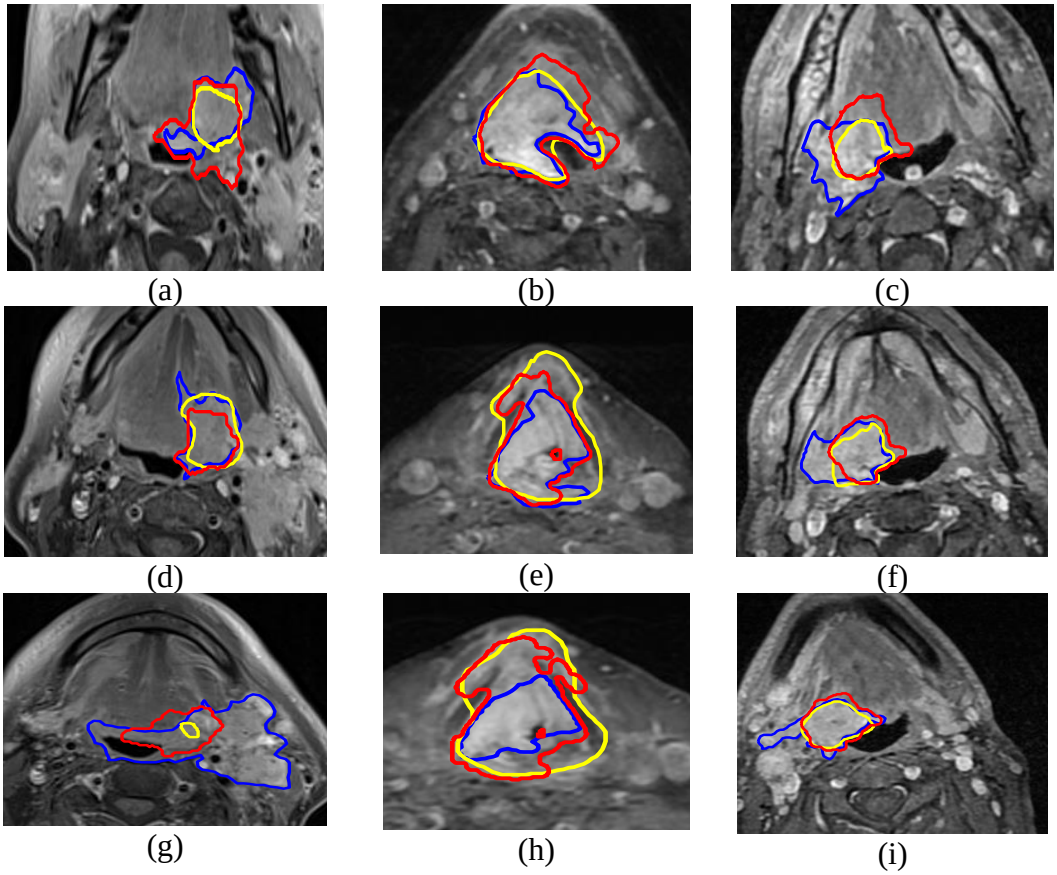


Fig. A.7 Examples of T1 axial MRI slices with HNC drawn by manual (yellow), methods in [193] (blue), and proposed automatic algorithm (red). (a-c) are from first patient. (d-f) are from second patient

D: Modified U-Net segmentation results

The quantitative analysis and some visual examples of proposed modified U-Net segmentation on HNC from MRI slice have been shown in Section 7.4, more visual examples are shown in

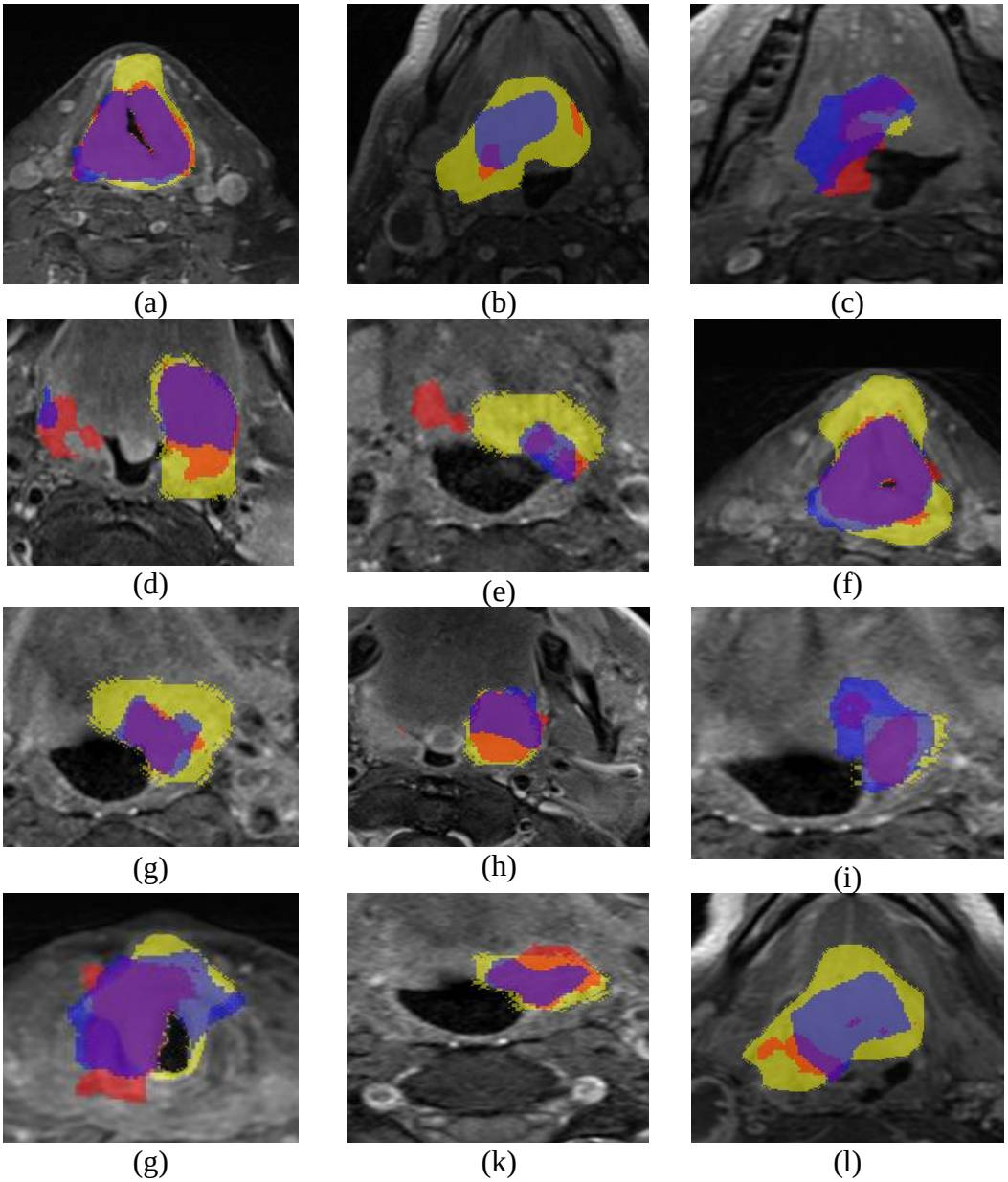


Fig. A.8 Visualisation of HNC segmentation. Yellow areas are from consensus manual outline, red areas are from U-Net, blue areas are from proposed DCNN.

Reference

[1] *Head and neck cancer*. Available: https://www.cancerresearchuk.org/about-cancer/head-neck-cancer

[2] L. Siegel Rebecca and M. K. D. J. Ahmedin, "Cancer statistics, 2020.[J]," *CA. Cancer J Clin*, vol. 70, pp. 7-30, 2020.

[3] S. Marur and A. A. Forastiere, "Head and neck cancer: changing epidemiology, diagnosis, and treatment," in *Mayo Clinic Proceedings*, 2008, pp. 489-501.

[4] P. C. Doyle, *Clinical Care and Rehabilitation in Head and Neck Cancer*: Springer, 2019.

[5] X. Geets, J.-F. Daisne, S. Arcangeli, E. Coche, M. De Poel, T. Duprez, *et al.*, "Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI," *Radiotherapy and oncology*, vol. 77, pp. 25-31, 2005.

[6] A. R. Gordon, L. A. Loevner, A. Shukla-Dave, R. O. Redfern, A. I. Sonners, A. M. Kilger, *et al.*, "Intraobserver variability in the MR determination of tumor volume in squamous cell carcinoma of the pharynx," *American journal of neuroradiology*, vol. 25, pp. 1092-1098, 2004.

[7] C. Perez and L. Brady, "Principles and practice of radiation oncology," *Journal of Pediatric Hematology/Oncology*, vol. 21, p. 560, 1999.

[8] D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," *Physics in Medicine & Biology*, vol. 42, p. 123, 1997.

[9] E. E. Cohen, S. J. LaMonte, N. L. Erb, K. L. Beckman, N. Sadeghi, K. A. Hutcheson, *et al.*, "American Cancer Society head and neck cancer survivorship care guideline," *CA: a cancer journal for clinicians*, vol. 66, pp. 203-239, 2016.

[10] D. Adelstein, M. L. Gillison, D. G. Pfister, S. Spencer, D. Adkins, D. M. Brizel, *et al.*, "NCCN guidelines insights: head and neck cancers, version 2.2017," *Journal of the National Comprehensive Cancer Network*, vol. 15, pp. 761-770, 2017.

[11] A. A. Forastiere, Q. Zhang, R. S. Weber, M. H. Maor, H. Goepfert, T. F. Pajak, *et al.*, "Long-term results of RTOG 91-11: a comparison of three nonsurgical treatment strategies to preserve the larynx in patients with locally advanced larynx cancer," *Journal of clinical oncology*, vol. 31, p. 845, 2013.

[12] R. Bourne, *Fundamentals of digital imaging in medicine*: Springer Science & Business Media, 2010.

[13] N. Dimakis, "Introduction to Medical Imaging–Physics, Engineering and Clinical Applications, by Nicholas Dimakis: Scope: medical imaging. Level: advanced undergraduates and graduate students in medical imaging and related disciplines," ed: Taylor & Francis, 2012.

[14] H. Singh, J. A. Neutze, and J. R. Enterline, *Radiology fundamentals: Introduction to imaging & technology*: Springer, 2014.

[15] N. Dey, V. Bhateja, and A. E. Hassanien, "Medical Imaging in Clinical Applications," *Springer International Publishing*, vol. 10, pp. 978-3, 2016.

[16] R. Liang, *Biomedical optical imaging technologies: design and applications*: Springer Science & Business Media, 2012.

[17] J. A. Radosevich, "Head & neck cancer: Current perspectives, advances, and challenges," 2013.

[18] My_MS.rog. (7/7). *Basic Plane Mathematics of MRI*.

[19] J. A. Ridge, B. S. Glisson, M. N. Lango, S. Feigenberg, and E. Horwitz, "Head and neck tumors," *Cancer management: a multidisciplinary approach*, vol. 11, p. 369, 2008.

[20] R. Hermans, *Head and neck cancer imaging*: Springer, 2006.

[21] J. L. Hiatt, *Textbook of head and neck anatomy*: Jones & Bartlett Publishers, 2020.
[22] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," *CA: a cancer journal for clinicians,* vol. 71, pp. 7-33, 2021.
[23] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians,* vol. 71, pp. 209-249, 2021.
[24] W. J. Blot, J. K. McLaughlin, D. M. Winn, D. F. Austin, R. S. Greenberg, S. Preston-Martin, *et al.*, "Smoking and drinking in relation to oral and pharyngeal cancer," *Cancer research,* vol. 48, pp. 3282-3287, 1988.
[25] D. I. Rosenthal, T. R. Mendoza, M. S. Chambers, J. A. Asper, I. Gning, M. S. Kies, *et al.*, "Measuring head and neck cancer symptom burden: the development and validation of the MD Anderson symptom inventory, head and neck module," *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck,* vol. 29, pp. 923-931, 2007.
[26] E. Maghami and A. S. Ho, *Multidisciplinary Care of the Head and Neck Cancer Patient* vol. 174: Springer, 2018.
[27] P. Denoix, "Tumor, node and metastasis (TNM)," *Bull Inst Nat Hyg (Paris),* vol. 1, pp. 1-69, 1944.
[28] C. C. Compton, D. R. Byrd, J. Garcia-Aguilar, S. H. Kurtzman, A. Olawaiye, and M. K. Washington, *AJCC cancer staging atlas: a companion to the seventh editions of the AJCC cancer staging manual and handbook*: Springer Science & Business Media, 2012.
[29] M. C. Jäckel, A. Martin, and W. Steiner, "Twenty-five years experience with laser surgery for head and neck tumors," *European archives of oto-rhino-laryngology,* vol. 264, pp. 577-585, 2007.
[30] S. A. Dowthwaite, J. H. Franklin, D. A. Palma, K. Fung, J. Yoo, and A. C. Nichols, "The role of transoral robotic surgery in the management of oropharyngeal cancer: a review of the literature," *International Scholarly Research Notices,* vol. 2012, 2012.
[31] A. A. Bui and R. K. Taira, *Medical imaging informatics*: Springer Science & Business Media, 2009.
[32] A. G. Webb, *Introduction to biomedical imaging*: John Wiley & Sons, 2017.
[33] I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, *et al.*, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning a," *Medical physics,* vol. 34, pp. 4738-4749, 2007.
[34] J. Marx, R. Hockberger, and R. Walls, *Rosen's Emergency Medicine-Concepts and Clinical Practice E-Book: 2-Volume Set*: Elsevier Health Sciences, 2013.
[35] M. A. de Souza, H. R. Gamba, and H. Pedrini, *Multi-Modality Imaging: Applications and Computational Techniques*: Springer, 2018.
[36] R. Ansorge and M. J. Graves, *The Physics and Mathematics of MRI*: Morgan & Claypool Publishers San Rafael, 2016.
[37] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, *et al.*, "MR pulse sequences: what every radiologist wants to know but is afraid to ask," *Radiographics,* vol. 26, pp. 513-537, 2006.
[38] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging,* vol. 34, pp. 1993-2024, 2014.
[39] M. J. Graves and D. G. Mitchell, "Body MRI artifacts in clinical practice: a physicist's and radiologist's perspective," *Journal of Magnetic Resonance Imaging,* vol. 38, pp. 269-287, 2013.
[40] N. A. R. C. Color Usage Research Lab. (19/6). *luminance contrast*.

[41] D. G. Pelli and P. Bex, "Measuring contrast sensitivity," *Vision research,* vol. 90, pp. 10-14, 2013.
[42] P. J. Bex and W. Makous, "Spatial frequency, phase, and the contrast of natural images," *JOSA A,* vol. 19, pp. 1096-1106, 2002.
[43] D. G. Pelli and B. Farell, "Why use noise?," *JOSA A,* vol. 16, pp. 647-653, 1999.
[44] L. Erasmus, D. Hurter, M. Naudé, H. Kritzinger, and S. Acho, "A short overview of MRI artefacts," *SA Journal of Radiology,* vol. 8, 2004.
[45] R. Duarte, A. Repetti, P. A. Gómez, M. Davies, and Y. Wiaux, "Greedy approximate projection for magnetic resonance fingerprinting with partial volumes," *Inverse Problems,* vol. 36, p. 035015, 2020.
[46] J. Veraart, E. Fieremans, I. O. Jelescu, F. Knoll, and D. S. Novikov, "Gibbs ringing in diffusion MRI," *Magnetic resonance in medicine,* vol. 76, pp. 301-314, 2016.
[47] P. J. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance, Series B,* vol. 103, pp. 247-254, 1994.
[48] J. Beutel, H. L. Kundel, and R. L. Van Metter, *Handbook of medical imaging* vol. 1: Spie Press, 2000.
[49] R. Mehta and J. Sivaswamy, "A hybrid approach to tissue-based intensity standardization of brain MRI images," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 95-98.
[50] A. Mansoor and M. G. Linguraru, "Generic method for intensity standardization of medical images using multiscale curvelet representation," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1320-1323.
[51] I. J. Cox, S. Roy, and S. L. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," in *Proceedings., International Conference on Image Processing*, 1995, pp. 366-369.
[52] L. G. Nyúl and J. K. Udupa, "On standardizing the MR image intensity scale," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine,* vol. 42, pp. 1072-1081, 1999.
[53] D. G. Pfister, K.-K. Ang, D. M. Brizel, B. A. Burtness, A. J. Cmelak, A. D. Colevas, *et al.*, "Head and neck cancers," *Journal of the National Comprehensive Cancer Network,* vol. 9, pp. 596-650, 2011.
[54] N. G. Burnet, S. J. Thomas, K. E. Burton, and S. J. Jefferies, "Defining the tumour and target volumes for radiotherapy," *Cancer imaging : the official publication of the International Cancer Imaging Society,* vol. 4, pp. 153-161, 2004.
[55] T. Doshi, "Segmentation and quantification of oropharynx and larynx tumours from MRI data," Thesis [PhD] --University of Strathclyde, 2016, 2016.
[56] C. Clark, E. Miles, M. G. Urbano, S. Bhide, A. Bidmead, K. Harrington, *et al.*, "Pre-trial quality assurance processes for an intensity-modulated radiation therapy (IMRT) trial: PARSPORT, a UK multicentre Phase III trial comparing conventional radiotherapy and parotid-sparing IMRT for locally advanced head and neck cancer," *The British journal of radiology,* vol. 82, pp. 585-594, 2009.
[57] M. G. Urbano, C. Clark, C. Kong, E. Miles, D. Dearnaley, K. Harrington, *et al.*, "Target volume definition for head and neck intensity modulated radiotherapy: pre-clinical evaluation of PARSPORT trial guidelines," *Clinical oncology,* vol. 19, pp. 604-613, 2007.
[58] J. Van der Veen, S. Willems, S. Deschuymer, D. Robben, W. Crijns, F. Maes, *et al.*, "Benefits of deep learning for delineation of organs at risk in head and neck cancer," *Radiotherapy and Oncology,* vol. 138, pp. 68-74, 2019.

[59] E. Street, L. Hadjiiski, B. Sahiner, S. Gujar, M. Ibrahim, S. K. Mukherji, *et al.*, "Automated volume analysis of head and neck lesions on CT scans using 3D level set segmentation," *Medical physics*, vol. 34, pp. 4399-4408, 2007.

[60] W. Deng, L. Luo, X. Lin, T. Fang, D. Liu, G. Dan, *et al.*, "Head and neck cancer tumor segmentation using support vector machine in dynamic contrast-enhanced MRI," *Contrast media & molecular imaging*, vol. 2017, 2017.

[61] A. Iantsen, D. Visvikis, and M. Hatt, "Squeeze-and-excitation normalization for automated delineation of head and neck primary tumors in combined PET and CT images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, 2020, pp. 37-43.

[62] J. Xie and Y. Peng, "The head and neck tumor segmentation using nnU-Net with spatial and channel 'squeeze & excitation'blocks," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, 2020, pp. 28-36.

[63] Y. Yuan, "Automatic head and neck tumor segmentation in PET/CT with scale attention network," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, 2020, pp. 44-52.

[64] J. Y. Lim and M. Leech, "Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck," *Acta Oncologica*, vol. 55, pp. 799-806, 2016.

[65] L. Bielak, N. Wiedenmann, A. Berlin, N. H. Nicolay, D. D. Gunashekar, L. Hägele, *et al.*, "Convolutional neural networks for head and neck tumor segmentation on 7-channel multiparametric MRI: a leave-one-out analysis," *Radiation Oncology*, vol. 15, pp. 1-9, 2020.

[66] J. Ma and X. Yang, "Combining cnn and hybrid active contours for head and neck tumor segmentation in ct and pet images," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, 2020, pp. 59-64.

[67] R. Sims, A. Isambert, V. Grégoire, F. Bidault, L. Fresco, J. Sage, *et al.*, "A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck," *Radiotherapy and Oncology*, vol. 93, pp. 474-478, 2009.

[68] D. P. Huyskens, P. Maingon, L. Vanuytsel, V. Remouchamps, T. Roques, B. Dubray, *et al.*, "A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer," *Radiotherapy and Oncology*, vol. 90, pp. 337-345, 2009.

[69] G. V. Walker, M. Awan, R. Tao, E. J. Koay, N. S. Boehling, J. D. Grant, *et al.*, "Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer," *Radiotherapy and Oncology*, vol. 112, pp. 321-325, 2014.

[70] G. Sharp, K. D. Fritscher, V. Pekar, M. Peroni, N. Shusharina, H. Veeraraghavan, *et al.*, "Vision 20/20: perspectives on automated image segmentation for radiotherapy," *Medical physics*, vol. 41, p. 050902, 2014.

[71] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magnetic resonance imaging*, vol. 30, pp. 1323-1341, 2012.

[72] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nature methods*, vol. 9, pp. 671-675, 2012.

[73] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, *et al.*, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, pp. 1116-1128, 2006.

[74] E. Gómez-de-Mariscal, C. García-López-de-Haro, W. Ouyang, L. Donati, E. Lundberg, M. Unser, *et al.*, "DeepImageJ: A user-friendly environment to run deep learning models in ImageJ," *Nature Methods*, vol. 18, pp. 1192-1195, 2021.

[75] A. Virzì, C. O. Muller, J.-B. Marret, E. Mille, L. Berteloot, D. Grévent, *et al.*, "Comprehensive review of 3D segmentation software tools for MRI usable for pelvic surgery planning," *Journal of digital imaging,* vol. 33, pp. 99-110, 2020.
[76] M. McAuliffe, "Medical image processing, analysis, and visualization (MIPAV)," *National Institutes of Health,* vol. 4, 2009.
[77] K. Doi, "Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology," *Physics in Medicine & Biology,* vol. 51, p. R5, 2006.
[78] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics,* vol. 31, pp. 198-211, 2007.
[79] U. Sinha, A. Bui, R. Taira, J. Dionisio, C. Morioka, D. Johnson, *et al.*, "A review of medical imaging informatics," *Annals of the New York Academy of Sciences,* vol. 980, pp. 168-197, 2002.
[80] C.-C. Teng, J. Mitchell, C. Walker, A. Swan, C. Davila, D. Howard, *et al.*, "A medical image archive solution in the cloud," in *2010 IEEE International Conference on Software Engineering and Service Sciences*, 2010, pp. 431-434.
[81] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing,* vol. 266, pp. 8-20, 2017.
[82] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging,* vol. 26, pp. 1025-1039, 2013.
[83] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE transactions on Information Technology in Biomedicine,* vol. 11, pp. 58-69, 2007.
[84] H.-P. Chan, B. Sahiner, L. Hadjiyski, C. Zhou, and N. Petrick, "Lung nodule detection and classification," ed: Google Patents, 2005.
[85] S.-C. Lo, S.-L. Lou, J.-S. Lin, M. T. Freedman, M. V. Chien, and S. K. Mun, "Artificial convolution neural network techniques and applications for lung nodule detection," *IEEE transactions on medical imaging,* vol. 14, pp. 711-718, 1995.
[86] S. G. Armato, F. Li, M. L. Giger, H. MacMahon, S. Sone, and K. Doi, "Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program," *Radiology,* vol. 225, pp. 685-692, 2002.
[87] X. Li, L. Shen, X. Xie, S. Huang, Z. Xie, X. Hong, *et al.*, "Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection," *Artificial intelligence in medicine,* vol. 103, p. 101744, 2020.
[88] T. Saba, "Automated lung nodule detection and classification based on multiple classifiers voting," *Microscopy research and technique,* vol. 82, pp. 1601-1609, 2019.
[89] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis,* vol. 35, pp. 18-31, 2017.
[90] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Medical image analysis,* vol. 8, pp. 275-283, 2004.
[91] G. Wang, W. Li, T. Vercauteren, and S. Ourselin, "Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation," *Frontiers in computational neuroscience,* vol. 13, p. 56, 2019.
[92] H. Chen, Z. Qin, Y. Ding, L. Tian, and Z. Qin, "Brain tumor segmentation with deep convolutional symmetric neural network," *Neurocomputing,* vol. 392, pp. 305-313, 2020.

[93] H. Li, A. Li, and M. Wang, "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks," *Computers in biology and medicine,* vol. 108, pp. 150-160, 2019.
[94] B. Łukaszewski, J. Nazar, M. Goch, M. Łukaszewska, A. Stępiński, and M. U. Jurczyk, "Diagnostic methods for detection of bone metastases," *Contemporary Oncology,* vol. 21, p. 98, 2017.
[95] E. Hausmann, K. Allen, L. Carpio, L. Christersson, and V. Clerehugh, "Computerized methodology for detection of alveolar crestal bone loss from serial intraoral radiographs," *Journal of periodontology,* vol. 63, pp. 657-662, 1992.
[96] A. Oliveira, S. Pereira, and C. A. Silva, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Systems with Applications,* vol. 112, pp. 229-242, 2018.
[97] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowledge-Based Systems,* vol. 178, pp. 149-162, 2019.
[98] N. Eladawi, M. M. Elmogy, M. Ghazal, O. Helmy, A. Aboelfetouh, A. Riad, *et al.*, "Classification of retinal diseases based on OCT images," *Front Biosci,* vol. 23, pp. 247-64, 2018.
[99] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images," *Ophthalmology Retina,* vol. 1, pp. 322-327, 2017.
[100] G. Samagaio, A. Estévez, J. de Moura, J. Novo, M. I. Fernández, and M. Ortega, "Automatic macular edema identification and characterization using OCT images," *Computer methods and programs in biomedicine,* vol. 163, pp. 47-63, 2018.
[101] R. J. van Klaveren, M. Oudkerk, M. Prokop, E. T. Scholten, K. Nackaerts, R. Vernhout, *et al.*, "Management of lung nodules detected by volume CT scanning," *New England Journal of Medicine,* vol. 361, pp. 2221-2229, 2009.
[102] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated detection of lung nodules in CT scans: preliminary results," *Medical physics,* vol. 28, pp. 1552-1561, 2001.
[103] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical physics,* vol. 38, pp. 915-931, 2011.
[104] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, *et al.*, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artificial intelligence in medicine,* vol. 83, pp. 58-66, 2017.
[105] E.-L. Chen, P.-C. Chung, C.-L. Chen, H.-M. Tsai, and C.-I. Chang, "An automatic diagnostic system for CT liver image classification," *IEEE transactions on biomedical engineering,* vol. 45, pp. 783-794, 1998.
[106] Y. Li, S. Hara, and K. Shimura, "A machine learning approach for locating boundaries of liver tumors in ct images," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 400-403.
[107] Y. Wang, L. Zhao, M. Wang, and Z. 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, pp. 144591-144602, 2019.
[108] Y. Lei, J. Harms, X. Dong, T. Wang, X. Tang, S. Y. David, *et al.*, "Organ-at-Risk (OAR) segmentation in head and neck CT using U-RCNN," in *Medical Imaging 2020: Computer-Aided Diagnosis*, 2020, p. 1131444.

[109] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, *et al.*, "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, pp. 202-207, 2020.
[110] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z. A. Fayad, N. Zhang, *et al.*, "Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection," *Radiology*, p. 200463, 2020.
[111] Y. Li and L. Xia, "Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management," *American Journal of Roentgenology*, vol. 214, pp. 1280-1286, 2020.
[112] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, "Brain tumor segmentation based on a hybrid clustering technique," *Egyptian Informatics Journal*, vol. 16, pp. 71-81, 2015.
[113] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE transactions on medical imaging*, vol. 35, pp. 1240-1251, 2016.
[114] R. J. Van der Geest, V. G. Buller, E. Jansen, H. J. Lamb, L. H. Baur, E. E. van der Wall, *et al.*, "Comparison between manual and semiautomated analysis of left ventricular volume parameters from short-axis MR images," *Journal of computer assisted tomography*, vol. 21, pp. 756-765, 1997.
[115] R. Rehr, C. R. Malloy, N. Filipchuk, and R. M. Peshock, "Left ventricular volumes measured by MR imaging," *Radiology*, vol. 156, pp. 717-719, 1985.
[116] D. M. Vigneault, E. Yang, P. J. Jensen, M. W. Tee, H. Farhad, L. Chu, *et al.*, "Left ventricular strain is abnormal in preclinical and overt hypertrophic cardiomyopathy: cardiac MR feature tracking," *Radiology*, vol. 290, pp. 640-648, 2019.
[117] P. C. Tang, J. W. Haft, M. A. Romano, A. Bitar, R. Hasan, M. Palardy, *et al.*, "Right ventricular function and residual mitral regurgitation after left ventricular assist device implantation determines the incidence of right heart failure," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 159, pp. 897-905. e4, 2020.
[118] S. Tilborghs, T. Dresselears, P. Claus, J. Bogaert, and F. Maes, "3D Left Ventricular Segmentation from 2D Cardiac MR Images Using Spatial Context," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2019, pp. 90-99.
[119] I. Thomassin-Naggara, S. Lamrabet, A. Crestani, A. Bekhouche, C. A. Wahab, E. Kermarrec, *et al.*, "Magnetic resonance imaging classification of deep pelvic endometriosis: description and impact on surgical management," *Human Reproduction*, vol. 35, pp. 1589-1600, 2020.
[120] I. Nekooeimehr, S. K. Lai-Yuen, P. Bao, A. Weitzenfeld, and S. Hart, "Automated contour tracking and trajectory classification of pelvic organs on dynamic MRI," *Journal of Medical Imaging*, vol. 5, p. 014008, 2018.
[121] R. Gonzalez and R. Woods, "Digital Image Processing (4th, Illustr ed.)," ed: Pearson, 2018.
[122] C. Rother, V. Kolmogorov, and A. Blake, "" GrabCut" interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)*, vol. 23, pp. 309-314, 2004.
[123] D. Terzopoulos and K. Fleischer, "Deformable models," *The visual computer*, vol. 4, pp. 306-331, 1988.
[124] R. C. Gonzalez, *Digital Image Processing*: Pearson Education, 2009.
[125] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 43, pp. 1-8, 1997.

[126] S. Gupta and Y. Kaur, "Review of different local and global contrast enhancement techniques for a digital image," *International Journal of Computer Applications*, vol. 100, pp. 18-23, 2014.
[127] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE transactions on Consumer Electronics*, vol. 45, pp. 68-75, 1999.
[128] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, *et al.*, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, pp. 355-368, 1987.
[129] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, *et al.*, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital imaging*, vol. 11, pp. 193-200, 1998.
[130] K. Somasundaram and P. Kalavathi, "Medical image contrast enhancement based on gamma correction," *Int J Knowl Manag e-learning*, vol. 3, pp. 15-18, 2011.
[131] T. K. Agarwal, M. Tiwari, and S. S. Lamba, "Modified histogram based contrast enhancement using homomorphic filtering for medical images," in *2014 IEEE International Advance Computing Conference (IACC)*, 2014, pp. 964-968.
[132] U. K. Acharya and S. Kumar, "Genetic algorithm based adaptive histogram equalization (GAAHE) technique for medical image enhancement," *Optik*, vol. 230, p. 166273, 2021.
[133] Y. Ma, J. Liu, Y. Liu, H. Fu, Y. Hu, J. Cheng, *et al.*, "Structure and illumination constrained GAN for medical image enhancement," *IEEE Transactions on Medical Imaging*, vol. 40, pp. 3955-3967, 2021.
[134] S. U. Khan, N. Ullah, I. Ahmed, I. Ahmad, and M. I. Mahsud, "MRI imaging, comparison of MRI with other modalities, noise in MRI images and machine learning techniques for noise removal: a review," *Current Medical Imaging*, vol. 15, pp. 243-254, 2019.
[135] S. Rakshit, A. Ghosh, and B. U. Shankar, "Fast mean filtering technique (FMFT)," *Pattern Recognition*, vol. 40, pp. 890-897, 2007.
[136] E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?," *The Annals of Statistics*, vol. 37, pp. 1172-1206, 2009.
[137] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 60-65.
[138] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, pp. 629-639, 1990.
[139] A. Akl and C. Yaacoub, "A hybrid wavelet-spatial denoising filter," in *2013 18th International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1-5.
[140] P. Elahi, S. Beheshti, and M. Hashemi, "BM3D mridenoising equipped with noise invalidation technique," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6612-6616.
[141] S. Fadnavis, J. Batson, and E. Garyfallidis, "Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16293-16303, 2020.
[142] D. Hong, C. Huang, C. Yang, J. Li, Y. Qian, and C. Cai, "FFA-DMRI: A network based on feature fusion and attention mechanism for brain MRI denoising," *Frontiers in Neuroscience*, vol. 14, p. 934, 2020.
[143] S. Suhas and C. Venugopal, "MRI image preprocessing and noise removal technique using linear and nonlinear filters," in *2017 International Conference on Electrical,*

Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017, pp. 1-4.
[144] M. Goyal, "Morphological image processing," *IJCST*, vol. 2, 2011.
[145] *Morphological Image Processing*. Available: https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic4.htm
[146] J. Mehena, "Medical image edge detection using modified morphological edge detection approach," *International Journal of Computer Sciences and Engineering*, vol. 7, pp. 523-528, 2019.
[147] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Medical image fusion via convolutional sparsity based morphological component analysis," *IEEE Signal Processing Letters*, vol. 26, pp. 485-489, 2019.
[148] T. A. Mahmoud and S. Marshall, "Medical image enhancement using threshold decomposition driven adaptive morphological filter," in *2008 16th European Signal Processing Conference*, 2008, pp. 1-5.
[149] J. Juntu, J. Sijbers, D. Van Dyck, and J. Gielen, "Bias field correction for MRI images," in *Computer recognition systems*, ed: Springer, 2005, pp. 543-551.
[150] S. Song, Y. Zheng, and Y. He, "A review of methods for bias correction in medical images," *Biomedical Engineering Review*, vol. 1, 2017.
[151] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE transactions on medical imaging*, vol. 26, pp. 405-421, 2007.
[152] D. Yang, H. Gach, H. Li, and S. Mutic, "TU-H-206-04: An Effective Homomorphic Unsharp Mask Filtering Method to Correct Intensity Inhomogeneity in Daily Treatment MR Images," *Medical Physics*, vol. 43, pp. 3774-3774, 2016.
[153] L. Axel, J. Costantini, and J. Listerud, "Intensity correction in surface-coil MR imaging," *American Journal of Roentgenology*, vol. 148, pp. 418-420, 1987.
[154] B. H. Brinkmann, A. Manduca, and R. A. Robb, "Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction," *IEEE transactions on medical imaging*, vol. 17, pp. 161-171, 1998.
[155] W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE transactions on medical imaging*, vol. 15, pp. 429-442, 1996.
[156] D. L. Pham, "Spatial models for fuzzy clustering," *Computer vision and image understanding*, vol. 84, pp. 285-297, 2001.
[157] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE transactions on medical imaging*, vol. 17, pp. 87-97, 1998.
[158] O. Salvado, C. Hillenbrand, S. Zhang, and D. L. Wilson, "Method to correct intensity inhomogeneity in MR images for atherosclerosis characterization," *IEEE transactions on medical imaging*, vol. 25, pp. 539-552, 2006.
[159] L. G. Nyúl, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE transactions on medical imaging*, vol. 19, pp. 143-150, 2000.
[160] G. Collewet, M. Strzelecki, and F. Mariette, "Influence of MRI acquisition protocols and image intensity normalization methods on texture classification," *Magnetic resonance imaging*, vol. 22, pp. 81-91, 2004.
[161] A. Madabhushi and J. K. Udupa, "New methods of MR image intensity standardization via generalized scale," *Medical physics*, vol. 33, pp. 3426-3434, 2006.
[162] X. Sun, L. Shi, Y. Luo, W. Yang, H. Li, P. Liang, *et al.*, "Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions," *Biomedical engineering online*, vol. 14, pp. 1-17, 2015.

[163] G. De Nunzio, R. Cataldo, and A. Carlà, "Robust intensity standardization in brain magnetic resonance images," *Journal of digital imaging*, vol. 28, pp. 727-737, 2015.
[164] F. Jäger, Y. Deuerling-Zheng, B. Frericks, F. Wacker, and J. Hornegger, "A new method for MRI intensity standardization with application to lesion detection in the brain," in *Vision modeling and visualization*, 2006, pp. 296-276.
[165] F. Jager and J. Hornegger, "Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 28, pp. 137-150, 2008.
[166] O. Dzyubachyk, M. Staring, M. Reijnierse, B. P. Lelieveldt, and R. J. van der Geest, "Inter-station intensity standardization for whole-body MR data," *Magnetic resonance in medicine*, vol. 77, pp. 422-433, 2017.
[167] P. Hellier, "Consistent intensity correction of MR images," in *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, 2003, pp. I-1109.
[168] Y. Gao, Y. Liu, Y. Wang, Z. Shi, and J. Yu, "A universal intensity standardization method based on a many-to-one weak-paired cycle generative adversarial network for magnetic resonance images," *IEEE transactions on medical imaging*, vol. 38, pp. 2059-2069, 2019.
[169] C. Chu, A. Zhmoginov, and M. Sandler, "Cyclegan, a master of steganography," *arXiv preprint arXiv:1712.02950*, 2017.
[170] M. Pedder, "Interpolation and filtering of spatial observations using successive corrections and Gaussian filters," *Monthly Weather Review*, vol. 121, pp. 2889-2902, 1993.
[171] T. Moraes, P. Amorim, J. V. Da Silva, and H. Pedrini, "Medical image interpolation based on 3D Lanczos filtering," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, pp. 294-300, 2020.
[172] E. J. Kirkland, "Bilinear interpolation," in *Advanced Computing in Electron Microscopy*, ed: Springer, 2010, pp. 261-263.
[173] O. Rukundo and H. Cao, "Nearest neighbor value interpolation," *arXiv preprint arXiv:1211.1768*, 2012.
[174] S. McKinley and M. Levine, "Cubic spline interpolation," *College of the Redwoods*, vol. 45, pp. 1049-1060, 1998.
[175] J. Adams, "A subsequence approach to interpolation using the FFT," *IEEE transactions on circuits and systems*, vol. 34, pp. 568-570, 1987.
[176] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, pp. 295-307, 2015.
[177] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664-1673.
[178] M.-s. Pan, X.-l. Yang, and J.-t. Tang, "Research on interpolation methods in medical image processing," *Journal of Medical Systems*, vol. 36, pp. 777-807, 2012.
[179] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE transactions on medical imaging*, vol. 18, pp. 1049-1075, 1999.
[180] Y. Sato, C.-F. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, *et al.*, "Tissue classification based on 3D local intensity structures for volume rendering," *IEEE Transactions on visualization and computer graphics*, vol. 6, pp. 160-180, 2000.
[181] S. Campbell, T. Doshi, J. Soraghan, L. Petropoulakis, G. Di Caterina, D. Grose, *et al.*, "3-dimensional throat region segmentation from MRI data based on fourier interpolation and 3-dimensional level set methods," in *2015 37th Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 2419-2422.
[182] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, pp. 321-331, 1988/01/01 1988.
[183] E. N. Mortensen and W. A. Barrett, "Interactive segmentation with intelligent scissors," *Graphical models and image processing*, vol. 60, pp. 349-384, 1998.
[184] W. A. Barrett and E. N. Mortensen, "Interactive live-wire boundary extraction," *Medical image analysis*, vol. 1, pp. 331-341, 1997.
[185] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, 2001, pp. 105-112.
[186] L. Grady, "Random walks for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, pp. 1768-1783, 2006.
[187] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, and J. Ding, "Neutro-connectedness cut," *IEEE Transactions on Image Processing*, vol. 25, pp. 4691-4703, 2016.
[188] S. Goswami and L. K. P. Bhaiya, "Brain tumour detection using unsupervised learning based neural network," in *2013 International Conference on Communication Systems and Network Technologies*, 2013, pp. 573-577.
[189] L. Rundo, L. Beer, S. Ursprung, P. Martin-Gonzalez, F. Markowetz, J. D. Brenton, *et al.*, "Tissue-specific and interpretable sub-segmentation of whole tumour burden on CT images by unsupervised fuzzy clustering," *Computers in biology and medicine*, vol. 120, p. 103751, 2020.
[190] S. Chatterjee, A. Sciarra, M. Dünnwald, P. Tummala, S. K. Agrawal, A. Jauhari, *et al.*, "StRegA: Unsupervised Anomaly Detection in Brain MRIs using a Compact Context-encoding Variational Autoencoder," *arXiv preprint arXiv:2201.13271*, 2022.
[191] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 616-625.
[192] N. Xu, B. Price, S. Cohen, J. Yang, and T. S. Huang, "Deep interactive object selection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 373-381.
[193] T. Doshi, J. Soraghan, L. Petropoulakis, G. Di Caterina, D. Grose, K. MacKenzie, *et al.*, "Automatic pharynx and larynx cancer segmentation framework (PLCSF) on contrast enhanced MR images," *Biomedical Signal Processing and Control*, vol. 33, pp. 178-188, 2017.
[194] R. Lan, Y. Zhou, Z. Liu, and X. Luo, "Prior knowledge-based probabilistic collaborative representation for visual recognition," *IEEE transactions on cybernetics*, vol. 50, pp. 1498-1508, 2018.
[195] N. Archip, P.-J. Erard, M. Egmont-Petersen, J.-M. Haefliger, and J.-F. Germond, "A knowledge-based approach to automatic detection of the spinal cord in CT images," *IEEE Transactions on medical imaging*, vol. 21, pp. 1504-1516, 2002.
[196] Z. Li, Y. Liu, R. Hayward, J. Zhang, and J. Cai, "Knowledge-based power line detection for UAV surveillance and inspection systems," in *2008 23rd International Conference Image and Vision Computing New Zealand*, 2008, pp. 1-6.
[197] Y.-J. Liu, C.-C. Yu, M.-J. Yu, and Y. He, "Manifold SLIC: A fast method to compute content-sensitive superpixels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 651-659.
[198] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, pp. 2290-2297, 2009.

[199] D. Weikersdorfer, A. Schick, and D. Cremers, "Depth-adaptive supervoxels for RGB-D video segmentation," in *2013 IEEE International Conference on Image Processing*, 2013, pp. 2708-2712.
[200] S. Chuai-Aree, C. Lursinsap, P. Sophasathit, and S. Siripant, "Fuzzy c-mean: A statistical feature classification of text and image segmentation method," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, pp. 661-671, 2001.
[201] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, pp. 1124-1137, 2004.
[202] J. Long, X. Feng, X. Zhu, J. Zhang, and G. Gou, "Efficient superpixel-guided interactive image segmentation based on graph theory," *Symmetry*, vol. 10, p. 169, 2018.
[203] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, pp. 641-647, 1994.
[204] M. Fan and T. C. Lee, "Variants of seeded region growing," *IET image processing*, vol. 9, pp. 478-485, 2015.
[205] S. Manoharan, "Improved version of graph-cut algorithm for CT images of lung cancer with clinical property condition," *Journal of Artificial Intelligence*, vol. 2, pp. 201-206, 2020.
[206] Z. Liu, Y.-Q. Song, V. S. Sheng, L. Wang, R. Jiang, X. Zhang, *et al.*, "Liver CT sequence segmentation based with improved U-Net and graph cut," *Expert Systems with Applications*, vol. 126, pp. 54-63, 2019.
[207] L. Liu, D. Raber, D. Nopachai, P. Commean, D. Sinacore, F. Prior, *et al.*, "Interactive separation of segmented bones in CT volumes using graph cut," in *International conference on medical image computing and computer-assisted intervention*, 2008, pp. 296-304.
[208] A. Stefano, S. Vitabile, G. Russo, M. Ippolito, M. G. Sabini, D. Sardina, *et al.*, "An enhanced random walk algorithm for delineation of head and neck cancers in PET studies," *Medical & biological engineering & computing*, vol. 55, pp. 897-908, 2017.
[209] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan, "Graph cut segmentation with a statistical shape model in cardiac MRI," *Computer Vision and Image Understanding*, vol. 117, pp. 1027-1035, 2013.
[210] I. Njeh, L. Sallemi, I. B. Ayed, K. Chtourou, S. Lehericy, D. Galanaud, *et al.*, "3D multimodal MRI brain glioma tumor and edema segmentation: a graph cut distribution matching approach," *Computerized Medical Imaging and Graphics*, vol. 40, pp. 108-119, 2015.
[211] S. Ghose, J. Mitra, A. Oliver, R. Marti, X. Llado, J. Freixenet, *et al.*, "Graph cut energy minimization in a probabilistic learning framework for 3D prostate segmentation in MRI," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 125-128.
[212] S. Esneault, N. Hraiech, E. Delabrousse, and J.-L. Dillenseger, "Graph cut liver segmentation for interstitial ultrasound therapy," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 5247-5250.
[213] J.-w. Kuo, J. Mamou, Y. Wang, E. Saegusa-Beecroft, J. Machi, and E. J. Feleppa, "Segmentation of 3-d high-frequency ultrasound images of human lymph nodes using graph cut with energy functional adapted to local intensity distribution," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, pp. 1514-1525, 2017.

[214] Z. Pan and J. Lu, "A Bayes-based region-growing algorithm for medical image segmentation," *Computing in science & Engineering*, vol. 9, pp. 32-38, 2007.
[215] G. Mühlenbruch, M. Das, C. Hohl, J. E. Wildberger, D. Rinck, T. G. Flohr, *et al.*, "Global left ventricular function in cardiac CT. Evaluation of an automated 3D region-growing segmentation algorithm," *European radiology*, vol. 16, pp. 1117-1123, 2006.
[216] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic, and M. Smajlovic, "Automatic CT image segmentation of the lungs with region growing algorithm," in *18th international conference on systems, signals and image processing-IWSSIP*, 2011, pp. 395-400.
[217] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach," *IEEE transactions on medical imaging*, vol. 27, pp. 467-480, 2008.
[218] Y. Lu, T. Jiang, and Y. Zang, "Region growing method for the analysis of functional MRI data," *NeuroImage*, vol. 20, pp. 455-465, 2003.
[219] N. Raja, S. Fernandes, N. Dey, S. C. Satapathy, and V. Rajinikanth, "Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-12, 2018.
[220] S. Poonguzhali and G. Ravindran, "A complete automatic region growing method for segmentation of masses on ultrasound images," in *2006 International Conference on Biomedical and Pharmaceutical Engineering*, 2006, pp. 88-92.
[221] H. Fan, F. Meng, Y. Liu, F. Kong, J. Ma, and Z. Lv, "A novel breast ultrasound image automated segmentation algorithm based on seeded region growing integrating gradual equipartition threshold," *Multimedia Tools and Applications*, vol. 78, pp. 27915-27932, 2019.
[222] J. A. Sethian, "Level set methods: An act of violence," *American Scientist*, vol. 85, pp. 12-35, 1997.
[223] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, pp. 430-436.
[224] T. Chan and L. Vese, "An active contour model without edges," in *International Conference on Scale-Space Theories in Computer Vision*, 1999, pp. 141-151.
[225] J. Huang, F. Jian, H. Wu, and H. Li, "An improved level set method for vertebra CT image segmentation," *Biomedical engineering online*, vol. 12, pp. 1-16, 2013.
[226] H. Gao and O. Chae, "Individual tooth segmentation from CT images using level set method with shape and intensity prior," *Pattern Recognition*, vol. 43, pp. 2406-2417, 2010.
[227] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images," *Expert systems with applications*, vol. 39, pp. 9661-9668, 2012.
[228] B. Amarapur, "Cognition-based MRI brain tumor segmentation technique using modified level set method," *Cognition, Technology & Work*, vol. 21, pp. 357-369, 2019.
[229] J. Wu, T. R. Mazur, S. Ruan, C. Lian, N. Daniel, H. Lashmett, *et al.*, "A deep Boltzmann machine-driven level set method for heart motion tracking using cine MRI images," *Medical image analysis*, vol. 47, pp. 68-80, 2018.
[230] Q. Zhang, J. Zhang, M. Wang, L. He, Y. Men, J. Wei, *et al.*, "Head and Neck Tumor Segmentation Based on Augmented Gradient Level Set Method," *Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi*, vol. 32, pp. 887-91, 904, 2015.

[231] C. Hollensen, P. S. Jørgensen, L. Højgaard, L. Specht, and R. Larsen, "Auto-Segmentation of head and neck cancer using textural features," in *29th European Society for Therapeutic Radiology and Oncology*, 2010.
[232] A. Rodtook, K. Kirimasthong, W. Lohitvisate, and S. S. Makhanov, "Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities," *Pattern Recognition*, vol. 79, pp. 172-182, 2018.
[233] S. Fan, L. K. Voon, and N. W. Sing, "3D prostate surface detection from ultrasound images based on level set method," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 389-396.
[234] D. Jayadevappa, S. Srinivas Kumar, and D. Murty, "Medical image segmentation algorithms using deformable models: a review," *IETE Technical review*, vol. 28, pp. 248-255, 2011.
[235] L. D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-D and 3-D images," *IEEE Transactions on Pattern Analysis and machine intelligence*, vol. 15, pp. 1131-1147, 1993.
[236] C. S. Poon, M. Braun, R. Fahrig, A. Ginige, and A. Dorrell, "Segmentation of medical images using an active-contour model incorporating region-based image features," in *Visualization in Biomedical Computing 1994*, 1994, pp. 90-97.
[237] C. Xu and J. L. Prince, "Generalized gradient vector flow external forces for active contours," *Signal processing*, vol. 71, pp. 131-139, 1998.
[238] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape," *International journal of computer vision*, vol. 72, pp. 195-215, 2007.
[239] J. Deng and H.-T. Tsui, "A fast level set method for segmentation of low contrast noisy biomedical images," *Pattern Recognition Letters*, vol. 23, pp. 161-169, 2002.
[240] S. Jin-Ping and L. Shuai-Jie, "an improved Mumford-Shah model and its applications to image processing with the piecewise constant level set method," *Acta Automatica Sinica*, vol. 33, pp. 1259-1262, 2007.
[241] Y. Z. Ma and J. X. Chen, "A new medical image segmentation method based on chan-vese model," in *Applied Mechanics and Materials*, 2014, pp. 3750-3756.
[242] J. Zhao, F. Shao, Y. Xu, X. Zhang, and W. Huang, "An improved Chan-Vese model without reinitialization for medical image segmentation," in *2010 3rd International Congress on Image and Signal Processing*, 2010, pp. 1317-1321.
[243] N. Zhang, J. Zhang, and R. Shi, "An Improved Chan-Vese model for medical image segmentation," in *2008 International Conference on Computer Science and Software Engineering*, 2008, pp. 864-867.
[244] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *2008 fifth international conference biomedical visualization: information visualization in medical and biomedical informatics*, 2008, pp. 71-76.
[245] S. Ghose, L. Holloway, K. Lim, P. Chan, J. Veera, S. K. Vinod, *et al.*, "A review of segmentation and deformable registration methods applied to adaptive cervical cancer radiation therapy treatment planning," *Artificial intelligence in medicine*, vol. 64, pp. 75-87, 2015.
[246] A. Kuisma, I. Ranta, J. Keyriläinen, S. Suilamo, P. Wright, M. Pesola, *et al.*, "Validation of automated magnetic resonance image segmentation for radiation therapy planning in prostate cancer," *Physics and imaging in radiation oncology*, vol. 13, pp. 14-20, 2020.
[247] L. Rundo, C. Militello, G. Russo, S. Vitabile, M. C. Gilardi, and G. Mauri, "GTVcut for neuro-radiosurgery treatment planning: an MRI brain cancer seeded image

segmentation method based on a cellular automata model," *Natural Computing,* vol. 17, pp. 521-536, 2018.
[248] H. Zaidi and I. El Naqa, "PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques," *European journal of nuclear medicine and molecular imaging,* vol. 37, pp. 2165-2187, 2010.
[249] G. Dougherty, M. J. Johnson, and M. D. Wiers, "Measurement of retinal vascular tortuosity and its application to retinal pathologies," *Medical & biological engineering & computing,* vol. 48, pp. 87-95, 2010.
[250] G. Wang, M. Li, Z. Yun, Z. Duan, K. Ma, Z. Luo, *et al.*, "A novel multiple subdivision-based algorithm for quantitative assessment of retinal vascular tortuosity," *Experimental Biology and Medicine,* vol. 246, pp. 2222-2229, 2021.
[251] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence,* pp. 679-698, 1986.
[252] S. M. Smith and J. M. Brady, "SUSAN—a new approach to low level image processing," *International journal of computer vision,* vol. 23, pp. 45-78, 1997.
[253] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science,* vol. 2, pp. 559-572, 1901.
[254] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems,* vol. 2, pp. 37-52, 1987.
[255] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 374, p. 20150202, 2016.
[256] C. He, Q. Liu, H. Li, and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Procedia Engineering,* vol. 7, pp. 280-285, 2010.
[257] L. Shang, J. C. Lv, and Z. Yi, "Rigid medical image registration using PCA neural network," *Neurocomputing,* vol. 69, pp. 1717-1722, 2006.
[258] J. Katkar, T. Baraskar, and V. R. Mankar, "A novel approach for medical image segmentation using PCA and K-means clustering," in *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2015, pp. 430-435.
[259] S. Li, T. Fevens, A. Krzyżak, and S. Li, "Automatic clinical image segmentation using pathological modeling, PCA and SVM," *Engineering Applications of Artificial Intelligence,* vol. 19, pp. 403-410, 2006.
[260] S. Dambreville, Y. Rathi, and A. Tannen, "Shape-based approach to robust image segmentation using kernel PCA," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, pp. 977-984.
[261] T. Arulananth, L. Balaji, M. Baskar, V. Anbarasu, and K. S. Rao, "PCA based dimensional data reduction and segmentation for DICOM images," *Neural Processing Letters*, pp. 1-15, 2020.
[262] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 24, pp. 971-987, 2002.
[263] E. Badeka, C. I. Papadopoulou, and G. A. Papakostas, "Evaluation of LBP variants in retinal blood vessels segmentation using machine learning," in *2020 international conference on intelligent systems and computer vision (ISCV)*, 2020, pp. 1-7.
[264] R. Namías, M.-E. Bellemare, M. Rahim, and N. Pirró, "Uterus segmentation in dynamic MRI using lbp texture descriptors," in *Medical Imaging 2014: Image Processing*, 2014, pp. 1009-1017.
[265] D. Q. Zeebaree, H. Haron, A. M. Abdulazeez, and D. A. Zebari, "Trainable model based on new uniform LBP feature to identify the risk of the breast cancer," in *2019*

International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 106-111.

[266] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, pp. 610-621, 1973.

[267] M. Yamunadevi and S. S. Ranjani, "Efficient segmentation of the lung carcinoma by adaptive fuzzy–GLCM (AF-GLCM) with deep learning based classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 4715-4725, 2021.

[268] T. T. Htay and S. S. Maung, "Early stage breast cancer detection system using glcm feature extraction and k-nearest neighbor (k-NN) on mammography image," in *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, 2018, pp. 171-175.

[269] T. Yun and H. Shu, "Ultrasound image segmentation by spectral clustering algorithm based on the curvelet and GLCM features," in *2011 International Conference on Electrical and Control Engineering*, 2011, pp. 920-923.

[270] S. Sedai, P. K. Roy, and R. Garnavi, "Right ventricle landmark detection using multiscale HOG and random forest classifier," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 814-818.

[271] Z. Al Sadeque, T. I. Khan, Q. D. Hossain, and M. Y. Turaba, "Automated detection and classification of liver cancer from ct images using hog-svm model," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, 2019, pp. 21-26.

[272] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.

[273] Y. Xu, C. Xu, X. Kuang, H. Wang, E. I. C. Chang, W. Huang, *et al.*, "3D-SIFT-Flow for atlas-based CT liver image segmentation," *Medical physics*, vol. 43, pp. 2229-2241, 2016.

[274] M. Yang, Y. Yuan, X. Li, and P. Yan, "Medical Image Segmentation Using Descriptive Image Features," in *BMVC*, 2011, pp. 1-11.

[275] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, pp. 660-674, 1991.

[276] P. Rajendran and M. Madheswaran, "Hybrid medical image classification using association rule mining with decision tree algorithm," *arXiv preprint arXiv:1001.3503*, 2010.

[277] V. Tallapragada, D. M. Reddy, P. S. Kiran, and D. V. Reddy, "A novel medical image segmentation and classification using combined feature set and decision tree classifier," *International Journal of Research in Engineering and Technology*, vol. 4, pp. 83-86, 2016.

[278] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001.

[279] D. Mahapatra, "Analyzing training information from random forests for improved image segmentation," *IEEE Transactions on Image Processing*, vol. 23, pp. 1504-1512, 2014.

[280] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, "Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model," *IEEE transactions on medical imaging*, vol. 37, pp. 1081-1091, 2017.

[281] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144-152.

[282] X.-j. Chen and D. Li, "Medical image segmentation based on threshold SVM," in *2010 International Conference on Biomedical Engineering and Computer Science*, 2010, pp. 1-3.
[283] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting brain tumors with conditional random fields and support vector machines," in *International Workshop on Computer Vision for Biomedical Image Applications*, 2005, pp. 469-478.
[284] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, pp. 533-536, 1986.
[285] N. Torbati, A. Ayatollahi, and A. Kermani, "An efficient neural network based method for medical image segmentation," *Computers in biology and medicine*, vol. 44, pp. 76-87, 2014.
[286] G. Margi and B. Talati, "Automated medical image segmentation using RBF ANN," in *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*, 2017, pp. 1-5.
[287] L. E. Raileanu and K. Stoffel, "Theoretical comparison between the gini index and information gain criteria," *Annals of Mathematics and Artificial Intelligence*, vol. 41, pp. 77-93, 2004.
[288] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278-2324, 1998.
[289] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*: MIT press, 2016.
[290] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[291] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.
[292] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[293] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214-224, 2015.
[294] C. Cernazanu-Glavan and S. Holban, "Segmentation of bone structure in X-ray images using convolutional neural network," *Adv. Electr. Comput. Eng*, vol. 13, pp. 87-94, 2013.
[295] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.
[296] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241.
[297] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, pp. 2481-2495, 2017.
[298] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," *Advances in neural information processing systems*, vol. 27, 2014.
[299] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012-10022.

[300] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[301] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE journal of biomedical and health informatics*, vol. 25, pp. 121-130, 2020.
[302] C. Kaul, S. Manandhar, and N. Pears, "Focusnet: An attention-based fully convolutional network for medical image segmentation," in *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, 2019, pp. 455-458.
[303] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.
[304] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, ed: Springer, 2017, pp. 240-248.
[305] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565-571.
[306] Z. Guo, N. Guo, K. Gong, and Q. Li, "Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network," *Physics in Medicine & Biology*, vol. 64, p. 205015, 2019.
[307] Y. Chen, B. Shi, Z. Wang, P. Zhang, C. D. Smith, and J. Liu, "Hippocampus segmentation through multi-view ensemble ConvNets," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 192-196.
[308] K. Men, J. Dai, and Y. Li, "Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks," *Medical physics*, vol. 44, pp. 6377-6389, 2017.
[309] Z. Jiang, C. Ding, M. Liu, and D. Tao, "Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task," in *International MICCAI brainlesion workshop*, 2019, pp. 231-241.
[310] X. Guo, C. Chen, Y. Lu, K. Meng, H. Chen, K. Zhou, *et al.*, "Retinal vessel segmentation combined with generative adversarial networks and dense U-Net," *IEEE Access*, vol. 8, pp. 194551-194560, 2020.
[311] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, "U-Net transformer: Self and cross attention for medical image segmentation," in *International Workshop on Machine Learning in Medical Imaging*, 2021, pp. 267-276.
[312] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: theory and applications," *arXiv preprint arXiv:2011.01118*, 2020.
[313] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *Journal of Imaging Science and Technology*, vol. 64, pp. 20508-1-20508-12, 2020.
[314] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297-302, 1945.
[315] M.-P. Dubuisson and A. K. Jain, "A modified Hausdorff distance for object matching," in *Proceedings of 12th international conference on pattern recognition*, 1994, pp. 566-568.
[316] B. Zhao, J. Soraghan, G. Di-caterina, L. Petropoulakis, D. Grose, and T. Doshi, "Automatic 3D segmentation of MRI data for detection of head and neck cancerous

lymph nodes," in *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2018, pp. 298-303.
[317] J. D. Mason, K. Hlaváčková, and K. Warwick, "Approximation using cubic B-splines with improved training speed and accuracy," in *Computer Intensive Methods in Control and Signal Processing*, ed: Springer, 1997, pp. 295-303.
[318] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, pp. 62-66, 1979.
[319] B. Zhao, J. Soraghan, D. Grose, T. Doshi, and G. Di-Caterina, "Automatic 3D detection and segmentation of head and neck cancer from MRI data," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018, pp. 1-6.
[320] A. E. Lefohn, J. M. Kniss, C. D. Hansen, and R. T. Whitaker, "A streaming narrow-band algorithm: interactive computation and visualization of level sets," in *ACM SIGGRAPH 2005 Courses*, ed, 2005, pp. 243-es.
[321] T. Doshi, J. Soraghan, D. Grose, K. MacKenzie, and L. Petropoulakis, "Modified fuzzy c-means clustering for automatic tongue base tumour extraction from MRI data," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 2460-2464.
[322] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE transactions on medical imaging*, vol. 23, pp. 447-458, 2004.
[323] S. Lankton and A. Tannenbaum, "Localizing region-based active contours," *IEEE transactions on image processing*, vol. 17, pp. 2029-2039, 2008.
[324] A. E. Lefohn, J. M. Kniss, C. D. Hansen, and R. T. Whitaker, *Interactive deformation and visualization of level set surfaces using graphics hardware*: IEEE, 2003.
[325] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
[326] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, *et al.*, "Understanding convolution for semantic segmentation," in *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, pp. 1451-1460.
[327] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.
[328] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 472-480.
[329] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester*, vol. 2014, p. 2, 2014.
[330] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, p. e3, 2016.
[331] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.
[332] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
[333] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.
[334] S. K. Kumar, "On weight initialization in deep neural networks," *arXiv preprint arXiv:1704.08863*, 2017.
[335] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," ed: A field guide to dynamical recurrent neural networks. IEEE Press, 2001.

[336] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015, pp. 448-456.
[337] R. Zhang, "Making convolutional networks shift-invariant again," in *International conference on machine learning*, 2019, pp. 7324-7334.
[338] Y. Luo, H. Cheng, and L. Yang, "Size-invariant fully convolutional neural network for vessel segmentation of digital retinal images," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016, pp. 1-7.
[339] Y. Xu, T. Xiao, J. Zhang, K. Yang, and Z. Zhang, "Scale-invariant convolutional neural networks," *arXiv preprint arXiv:1411.6369*, 2014.
[340] V. G. Mahesh, A. N. J. Raj, and Z. Fan, "Invariant moments based convolutional neural networks for image analysis," *International Journal of Computational Intelligence Systems*, vol. 10, pp. 936-950, 2017.
[341] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[342] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," *Encyclopedia of database systems*, vol. 5, pp. 532-538, 2009.