SLOW FUSION TRIPLANAR CONVOLUTIONAL NEURAL NETWORKS FOR LIVER TUMOR SEGMENTATION

CHUNG SHENG HUNG

UNIVERSITI SAINS MALAYSIA

2022

SLOW FUSION TRIPLANAR CONVOLUTIONAL NEURAL NETWORKS FOR LIVER TUMOR SEGMENTATION

by

CHUNG SHENG HUNG

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

April 2022

ACKNOWLEDGEMENT

First and foremost, I would like to extend my heartiest thanks and appreciation to my supervisors, Dr Gan Keng Hoon, Dr Anusha Achuthan and Professor Mandava Rajeswari for their never-ending support and valuable guidance throughout my doctorate studies. I am grateful for their helpful advice and suggestions from the various perspective of this study. I would also like to express my gratitude to Associate Professor Dhanesh Ramachandram whom I contacted for his professional advice on deep learning approach. A token of appreciation also goes to the other experts and researchers for their tremendous positive feedback. Last but not least, an honourable mention goes to my beloved family members for their, ever-present support, love and tolerance with me in pursuing my PhD programme. Without their constant support and encouragement, the road to completing this study would have been desolate and hard.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

2D	2 Dimensional
3D	3 Dimensional
CT	Computed Tomography
ConvNet	Convolutional Neural Networks
CT	Computed Tomography
FCN	Fully Convolutional Network
fMRI	Functional Magnetic Resonance Imaging
GPU	Graphical Processing Unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LiTS	Liver Tumor Segmentation
MICCAI	Medical Image Computing and Computer Assisted Intervention
MRI	Magnetic Resonance Imaging
ReLU	Rectified Linear Unit
ROI	Region of Interest
SGD	Stochastics Gradient Descent
VOI	Voxel of Interest
WHO	World Health Organization

LIST OF APPENDICES

GABUNGAN PERLAHAN KONVOLUSI RANGKAIAN NEURAL TRIPLANAR UNTUK PENSEGMENAN TUMOR HATI

ABSTRAK

Menurut laporan World Health Organization (WHO), tumor dalam hati adalah salah satu punca utama kematian semua penyakit kanser yang dilaporkan di seluruh dunia, dengan kadar kematian pesakit 745,000 pada 2014, pesakit-pesakit seramai 788,000 pada 2015 dan 782,000 pesakit pada 2018 masing-masing. Tumor hati diagnosis dan pembedahan perancangan biasanya dilaksanakan dengan Computed Tomography (CT) scan untuk membantu doktor dalam penilaian tumor hati dan perancangan berkaitan rawatan kepada pesakit-pesakit yang memerlukan. Walau bagaimanapun, cabaran-cabaran yang dihadapi dalam segmen tumor hati adalah; (i) keamatan sama antara tumor hati dan tisu-tisu hati, (ii) kecil dan hati tumor yang sukar ditentukan dan (iii) hati tumor dengan bentuk yang tidak teratur dan sempadan. Oleh itu, pengesanan tumor hati yang tepat dan pengsegmenan merupakan prasyarat penting untuk tumor diagnosis hati, pembedahan dan rawatan perancangan. Dalam kajian ini, kami menunjukkan penggunaan beberapa pandangan termasuk imej bersama paksi, sagittal dan ejeksi sebagai input untuk Convolutional Neural Networks bagi segmen tumor hati, dinamakan sebagai Triplanar Convolutional Neural Networks. Model rangkaian kami, Triplanar Convolutional Neural Networks menggunakan pandangan yang berbeza daripada CT imej hati untuk mengekstrak ciri-ciri diskriminasi dari Voxel of Interest (VOI) untuk mengklasifikasikan tumor hati dari rantau hati yang sihat dalam dataset CT hati yang diperolehi daripada MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge. Dalam kajian ini, kami mereka Triplanar Convolutional Neural Network dengan menggunakan tiga aliran input (input 1:

bersama paksi; 2 input: Sagittal; 3 input: Coronal), yang akan dilatih dengan menggunakan tiga model Convolutional Neural Networks untuk mengetahui ciri-ciri diskriminatif untuk mengumpul tumor dari CT imej hati secara automatik. Model cadangan rangkaian yang dilatih menggunakan Stochastics Gradient Descent (SGD) dan Rectified Linear Unit (ReLU) sebagai fungsi pengaktifan dan tindak balas daripada neuron bercantum di lapisan yang digabungkan bagi mengklasifikasikan tumor hati dan hati ketumbuhan di lapisan output terakhir. Keputusan eksperimen menunjukkan bahawa penggunaan teknik diregularisasi, 25% kadar keciciran yang digunakan pada lapisan tersembunyi berjaya dikurangkan overfitting dalam rangkaian. Kajian ini juga mengamalkan pendekatan penambahan data dengan menggunakan skala, mendatar flip, putaran dan transformasi zoom untuk data latihan dalam meningkatkan keberkesanan model rangkaian. Hasil kajian juga menunjukkan bahawa pendekatan Triplanar Convolutional Neural Networks menyediakan peningkatan yang ketara daripada pendekatan Single-view Convolutional Neural Networks dalam segmen tumor hati. Tesis ini seterusnya mencadangkan model ConvNet Fusion: Slow-Fusion ConvNet (Axial+Sagittal→ Coronal), Slow-Fusion ConvNet (Sagittal+Coronal→ Axial) dan Slow-Fusion ConvNet (Axial+Coronal→ Sagittal), yang merupakan model ConvNet Fusion baru untuk mengsegmenkan rantau hati dan rantau tumor hati dengan berkesanya.

SLOW FUSION TRIPLANAR CONVOLUTIONAL NEURAL NETWORKS FOR LIVER TUMOR SEGMENTATION

ABSTRACT

According to the World Health Organization (WHO) report, liver tumor is one of the leading cause of death in all cancerous disease reported worldwide, with fatalities rate of 745,000 patients in 2014, 788,000 patients in 2015 and 782,000 patients in 2018 respectively. Liver tumor diagnosis and surgery planning are commonly performed with Computed Tomography (CT) scan to assist doctors in evaluating liver tumor and planning of the relevant treatment for the patients. However, there are challenges faced in liver tumor segmentation such as (i) similar intensities between liver tumor and liver tissues, (ii) small and indeterminate liver tumor which are difficult to characterize and (iii) liver tumor with irregular shapes and boundaries. Therefore, an accurate liver tumor detection and segmentation is a crucial prerequisite for liver tumor diagnosis, surgery and treatment planning. In this study, we demonstrate the use of multiple views including axial, sagittal and coronal images as the inputs for Convolutional Neural Networks for liver tumor segmentation, named as Triplanar Convolutional Neural Networks. Our designed network model, Triplanar Convolutional Neural Networks utilize different views of liver CT images to extract discriminative features from the Voxel of Interest (VOI) to classify liver tumor from a healthy liver region in the Liver CT dataset obtained from MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge. In this experimental study, we designed the Triplanar Convolutional Neural Networks by using three streams of inputs (1st input: Axial; 2nd input: Sagittal; 3rd input: Coronal), which are passed through three parallel Convolutional Neural Networks to automatically learn discriminative features for

segmenting tumor from liver CT images. The proposed network model is trained using Stochastics Gradient Descent (SGD) optimizer and Rectified Linear Unit (ReLU) as the activation function and the response of the neurons are concatenated in the merge layer in order to classify liver tumor and non-liver tumor in the final output layer. Experimental results show that the use of regularization technique, 25% dropout rate applied on the hidden layers successfully reduced overfitting in the network. This study also adopts data augmentation approach by applying scaling, horizontal flip, rotation and zoom transformation for the training data. The results also show that Triplanar Convolutional Neural Networks approach provides significant improvement than Single-view Convolutional Neural Network approach in liver tumor segmentation. This thesis further proposes ConvNet Fusion models: Slow-Fusion ConvNet (Axial+Sagittal→ Coronal), Slow-Fusion ConvNet (Sagittal+Coronal→ Axial) and Slow-Fusion ConvNet (Axial+Coronal→ Sagittal), which are new ConvNet Fusion models to effectively segment liver region and liver tumor region.

CHAPTER 1

INTRODUCTION

1.1 Liver Tumor

The role of the liver in the body is crucial. The liver eliminates toxins through the urine in the form of waste. According to Carneiro et al. (2016), liver cancer is among one of the most common types of cancerous diseases which was the cause of cancer death of 745,000 patients (World Health Organization, 2014), 788,000 patients worldwide (World Health Organization, 2015) and 782,000 patients in 2018 (World Health Organization, 2018) respectively. Liver cancer is one of the major cancer types that cause more than 600,000 deaths each year and there is increasing in the number of liver tumours diagnosed throughout the world (Frid-Adar et al., 2017). Hence, early treatment and diagnosis are needed for reducing number of cancer deaths.

As reported by WHO, liver cancer has been the second major causes of death in all cancerous diseases in 2014 and 2015; forth common causes of cancer death in 2018. Metastases are often spread from primary tumors to the liver during the course of a disease. Visible lesions in CT images and anomalies in the texture and shape of the liver are vital indicators for initial diagnosis and progression in hepatic tumor diseases (Heimann et al., 2009). Therefore, the liver and its tumor are routinely analysed in primary tumor staging.

1.2 Existing Approaches

Machine learning is the science of having a machine to perform without being programmed explicitly. It can be defined as computational methods to make accurate predictions or improve performance based on the experience. The experience is referred to as the past information available in the form of electronic data and made available for analysis.

Data is unable to be processed in the raw form using conventional machine learning techniques. According to Schwier et al. (2011) and Smeets et al. (2010), traditional segmentation methods depend heavily on hadcrafted features and prior knowledge using user refinement. The handcrafted features derived from the liver lesion detection work (Schwier et al., 2011) includes features such as eccentricity, local contrast, roundness, size and percentage of an object size overlaps. Smeets et al. (2010) required the use of input from the user to specify maximal radius surrounding the liver tissue for level set segmentation of liver tumor. Therefore, careful domain expertise and feature engineering are needed in the design of feature extractors in order to process raw data into feature vector as an input to a classifier (LeCun et al., 2015). Figure 1.1 illustrates the process of machine learning with handcrafted features using feature extraction.

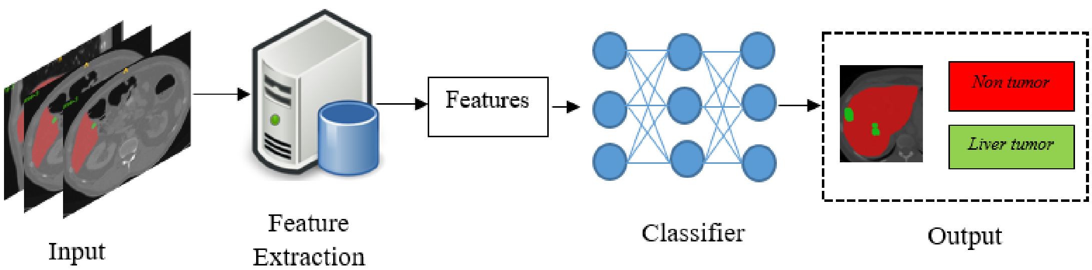


Figure 1.1 Machine Learning Algorithms

1.3 Deep Learning

In deep learning, computational models that consist of multiple processing layers are allowed to learn data representations with multiple levels of abstraction (Lin et al., 2018). Deep learning methods have improved dramatically in image recognition (Farabet et al., 2013; Krizhevsky et al., 2012; Le, 2013; Szegedy et al., 2014; Tompson et al., 2014) and speech recognition (G. Hinton et al., 2012; Mikolov et al., 2011;

Sainath et al., 2013) by sometimes more than 30% accuracy where the prior research work have been struggling to obtain 1-2%.

Deep learning is also known as representation learning, is considered as a set of techniques that automatically uncover the representations needed for classification and detection after feeding a machine with raw data. Deep learning uses multiple levels of representations obtained by the lower level into representation at a more abstract and higher level. The multi-level representations allow a deep architecture to learn hierarchical representations of the data without the need for dependence on handcrafted features as shown in Figure 1.2.

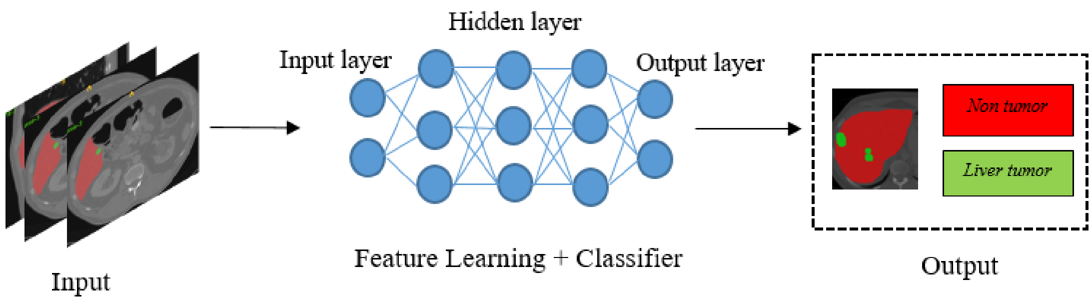


Figure 1.2 Deep Learning Algorithms

Deep learning focused on learning multi-level or hierarchical structure (Bengio, 2009) and it learns hierarchies of relevant features from the raw inputs directly (Bengio, 2009). Deep learning can be defined as learning multiple levels of abstraction and representation to gain insight from text, graphics and audio data. Many studies (Lee et al., 2008; Ranzato et al., 2006) have revealed that deep architectures with application to object recognition can be used to learn hierarchical structures in images.

For example, an image consists of an array of a pixel with intensity values. The features derived from the first layer of intensities representation may represent edges at a particular location and orientation of certain objects in an image (LeCun et al., 2015). The second layer detects motifs or patterns within an image by looking at the arrangements of edges. In the subsequent layer, motifs will be assembled into bigger combinations, followed by deeper layers which detect objects formed by combinations of the objects. Therefore, deep learning techniques target learning feature hierarchies with features from the highest level of hierarchies. The highest level of hierarchies is formed by combinations of lower level features (Bengio, 2009). The multi-level representations allow a deep architecture to learn feature hierarchies without the need for dependence on handcrafted features.

1.4 Problem Statement

Liver tumor segmentation is a difficult task due to variations of the intensity distributions. Researchers have been facing a variety of difficulties in liver tumor segmentation such as similar intensities, low contrast, fuzzy boundaries, ambiguity of boundaries between tumor and surrounding normal tissues (W. Li et al., 2015), heterogeneous densities and high variability in the shape and size (small in dimension) of liver tumor (Figure 1.3). In many cases, the variable size, shape and location of the organ tissue, medical image segmentation is considered as one of the challenging tasks in medical image analysis (Zhou et al., 2019).

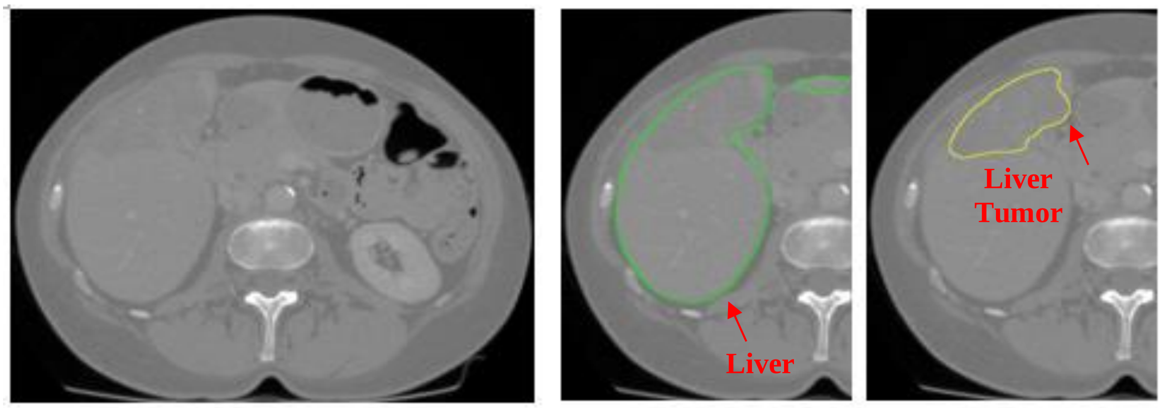


Figure 1.3 An Example of Liver Tumor CT Images Which Shows Overlapped Intensities

Manual segmentation or delineation of liver tumor is a laborious and time-consuming process, prone to interobserver variability. One of the main problems faced by liver tumor segmentation from CT images which are primarily related to the low contrast level between liver tumor and healthy liver intensities (Mandava et al., 2011; Sun et al., 2017). The characteristics of tumors are often found similar to those of the surrounding normal tissues. According to Shimizu et al. (2008), such challenge can also be attributed to high similarity of intensities (e.g. the CT value of liver metastases is found identical to the intensities of the gallbladder; the tumor necrosis CT values is similar to the fat tissue thus resulting in false positives).

Costa et al. (2011) highlighted that liver tumor characteristics such as size, margin definition, shape or enhancement pattern in different contrast phases are contributing to the liver tumor diagnosis and patients' treatment. The size of the liver are found to be varied with respect to the gender, age and body shape of a liver tumor patient (Sun et al., 2017). In identifying the liver tumor volume, different attenuation values are shown in the different contrast-enhancement phases for primary and secondary liver tumors (Choudhary et al., 2008). Nugroho et al. (2008) reported that the liver tumor size is small in dimension and lead to difficulties during detection. Besides, the challenges arise for identification of liver tumor includes similar-looking lesions may arise from different pathologies for abnormalities detection; benign or malignant; small and indeterminate liver lesions are difficult to characterize; liver lesions with different appearances (examples: intensities, contrast) may have similar characteristics (Costa et al., 2011). Hence, the difficulties of this task highlighted the needs for an accurate and robust liver tumor segmentation approach for detection and evaluation of liver tumor in CT images.

1.4.1 Convolutional Neural Networks (ConvNet) Fusion

With above challenges in liver tumor segmentation, automatic segmentation using a variety of methods such as region growing, graph cuts, watershed and methods based on machine learning such as decision tree, support vector machine, AdaBoost, K-means clustering are limited in their ability to process image data in their raw form and require the computation of hand-crafted features for high accuracy in liver tumor detection.

Recently, deep learning algorithms, especially ConvNet has been outperformed various algorithms in medical imageing (LeCun et al., 2015). In order to tackle the limitations faced in various methods described above, deep learning methods have been proposed, including ConvNet (Frid-Adar et al., 2017; W. Li et al., 2015; Vivanti et al., 2015), 3D CNN (Dou et al., 2016), Fully Convolutional Network (FCN) (Ben-Cohen et al., 2016; Christ et al., 2016; Sun et al., 2017; Vorontsov et al., 2018; Z. Wang et al., 2018) and Hybrid Densely Connected UNet (X. Li et al., 2018), Cascaded Deep Residual Networks (ResNet) (Bi et al., 2017). In addition, deep learning researchers mentioned above have applied approaches such as cascaded FCN (Christ et al., 2016; Z. Wang et al., 2018), multiphase contrast-enhanced CT liver tumor segmentation based on FCN (Sun et al., 2017), joint liver segmentation using FCN (Vorontsov et al., 2018) for liver tumor segmentation problem.

The above four cascaded methods (Bi et al., 2017; Christ et al., 2016; X. Li et al., 2018; Z. Wang et al., 2018) employed deep learning strategy which jointly segment the liver and the tumor in order to obtain high accuracy in liver tumor segmentation. The cascaded approach showed that two consecutive ConvNet are used to form the network architecture whereby the first network performs liver segmentation, while the second one incorporates the output of the first network and segment the tumor. In the cascaded approach performed by Christ et al. (2016), 3D CNN proposed by Dou et al. (2016) and Cascaded Deep Residual Networks designed by Bi et al. (2017), Conditional Random Fields (CRF) is used as post-process model on the output of the segmented liver region as contour refinement to segment the tumor region.

To the best of our knowledge, fusion methods in ConvNet have not been previously employed for liver tumor segmentation in CT liver images. A ConvNet fusion model utilizing different views of liver CT images would be benefitial to extract discriminative features from the Voxel of Interest (VOI) and extract useful data representation automatically with high accuracy and detection rate for the problem of liver tumor segmentation.

In this thesis, the following research questions are proposed for liver tumor segmentation:

(i) **How to develop a deep learning – Convolutional Neural Networks (ConvNet) model using fusion methods incorporating axial, sagittal and coronal images for liver tumor segmentation?**

(ii) **How to evaluate the performance of the proposed ConvNet model using liver CT dataset?**

1.5 Objectives

In this thesis, we propose the use of ConvNet approach to classify and segment liver tumor. This thesis attempts to achieve the main objectives through the following:

i) To propose modified ConvNet fusion methods namely:

- Slow-Fusion "Axial+Sagittal→Coronal" ConvNet
- Slow-Fusion "Sagittal+Coronal→Axial" ConvNet

- Slow-Fusion “Axial+Coronal→Sagittal” ConvNet

by integrating three planes of liver CT images including axial, sagittal and coronal images for liver tumor segmentation.

1.6 Scope

The scope of this thesis is limited to the following constraints as defined as follows:

- Contrast-enhanced abdominal CT scans which include the entire abdomen and thorax.

- The dataset in this work was acquired from MICCAI 2017 LiTS Challenge which comprises of six different clinical sites, with slice spacing from 0.45mm to 6.0mm and in-plane resolution from 0.55mm to 1.0mm. Each volume contains tumor/tumors in the liver. MICCAI 2017 LiTS Challenge dataset contains 131 training and 70 testing dataset of abdominal CT scans. The dataset contains the reference annotations of the liver tumor and liver region performed by trained radiologists. The LiTS liver CT dataset contains a total of 908 liver tumor. The images specification is as follows: 512 x 512 in-plane resolution, 8-bit and 512K resolution.

1.7 Significance of the Study

The need for an accurate liver tumor detection and segmentation is an essential prerequisite for computer-aided liver tumor diagnosis, planning of treatment and therapy. The liver tumor segmentation issues described as above has driven the motivation in this thesis for the use of deep learning approach for liver tumor segmentation from CT images. This work extends the use of Single-view Convolutional Neural Networks to Triplanar Convolutional Neural Networks by incorporating different views of liver tissues. The use of the Triplanar view aims to provide a better understanding by enriching the visualisation of important aspects of liver anatomy in order to contribute towards liver tumor segmentation problem. We proposed new ConvNet Fusion models namely Slow-Fusion ConvNet (Axial+Sagittal→Coronal), (Sagittal+Coronal→Axial) and (Axial+Coronal→Sagittal) and achieved high performance in liver tumor segmentation. We further address the challenges arose from class imbalance found in liver tumor CT images due to limited amount of dataset for liver tumor cases. We proposed the use of data augmentation, dropout regularization in building our network models.

1.8 Contributions

The thesis has lead to two main contributions as follows:

1. Proposing new ConvNet fusion models incorporating using three different orthogonal views, namely axial, coronal and sagittal views for liver tumor segmentation. The proposed Slow-Fusion ConvNet models are: "Axial+Sagittal→Coronal" ConvNet, "Sagittal+Coronal→Axial" ConvNet and "Axial+Coronal→Sagittal" ConvNet.
2. Proposing the use of Slow-Fusion ConvNet models for evaluation of liver tumor segmentation and achieves high performance in liver tumor segmentation accuraccies compared with related works. Evaluation metrics used as the performace measure include:
   a. Dice score comparison (Dice per case)

b. Accuracy
c. Optimizer
d. Learning rate
e. Training time

1.9 Outline

In **Chapter 2**, a review of the various methods done in the area of liver tumor segmentation is presented. The literature review based on Convolutional Neural Networks approaches, various methods developed for liver tumor segmentation, review of Deep Learning, the use of ConvNet for liver tumor segmentation, ConvNet Fusion approaches and other related work is presented.

Chapter 3 presents the proposed methodology of Slow-Fusion Triplanar Convolutional Neural Networks for liver tumor segmentation approach. This chapter covers the detailed explanations of the proposed ConvNet Fusion models, “Axial+Sagittal→Coronal” ConvNet, “Sagittal+Coronal→Axial” ConvNet and “Axial+Coronal→ Sagittal” ConvNet.

In **Chapter 4**, the proposed evaluation of Slow-Fusion Triplanar Convolutional Neural Networks, Single-View ConvNet, Early-Fusion and Late-Fusion ConvNet evaluations using liver CT dataset - MICCAI 2017 LiTS Challenge are presented.

Chapter 5 provides the conclusion of this thesis and discuss the potential future research.

CHAPTER 2

LITERATURE REVIEW

This chapter presents the literature study for the various methods of liver tumor segmentation. Firstly, Section 2.1 introduces Computed Tomography and its application in liver tumor diagnosis. Section 2.1 also explains different orthogonal views of liver CT images and overview of Convolutionl Neural Neworks algorithms. Section 2.2 describes different liver tumor characteristics. Section 2.3 presents various segmentation methods for liver tumor segmentation and reviews of ConvNet related work for liver tumor segmentation. We describes the basic of ConvNet and its components in Section 2.4. Section 2.5 provides a critical review of the state-of-the-art methods in liver tumor segmentation and various liver tumor segmentation methods. Section 2.6 discusses different ConvNet fusion approaches. Finally, the critical discussion for the use of existings liver tumor segmentation methods and the proposed ConvNet Fusion and Triplanar views are presented in Section 2.7.

2.1 Background

2.1.1 Computed Tomography

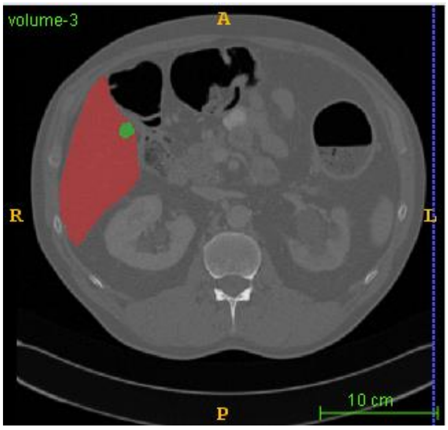
Computed Tomography (CT) images are one of the most common imaging modalities used for diagnosis of hepatic disease as it can provide relatively high-resolution images and accurate anatomical information (W. Li et al., 2015). According to the literature study, CT images are used for detection, diagnosis and follow-up of liver lesions and is the most commonly used modality for diagnosis of liver metastases, Hepatocellular Carcinoma (HCC) and Cholangiocarcinoma (CC). Contrast-enhanced abdominal CT scans are particularly useful in confirming the size and the location of the liver lesions (Sun et al., 2017).

Contrast can be defined as the separation of the darkest and lightest parts of an image. Highlights will be lightened and shadows will be darkened by increasing contrast. Contrast increasing will make objects in an image more distinguishable.

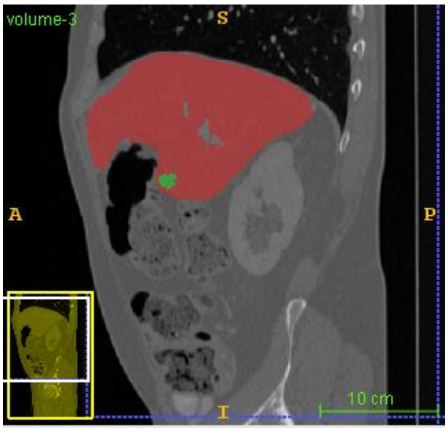
2.1.2 Views of CT Images

Liver tumor segmentation from abdominal CT images is an essential tasks for dignosis of liver tumor patients and is manually performed by doctors and radiologists using slice by slice on stack of CT images. In liver CT images, the liver and liver tumor exhibit different shapes and positions in different orthogonal projections. In this study, we explore the use of different orthogonal views of liver CT images, namely axial, sagittal and coronal views. The liver CT dataset was acquired in NIfTI format which three views were extacted and saved in data prep-processing. Axial view, also known as transverse view, divides the body into upper and lower segments. Sagittal view divides the body into left and right sections and coronal view (frontal) divides the body into front and back sections.

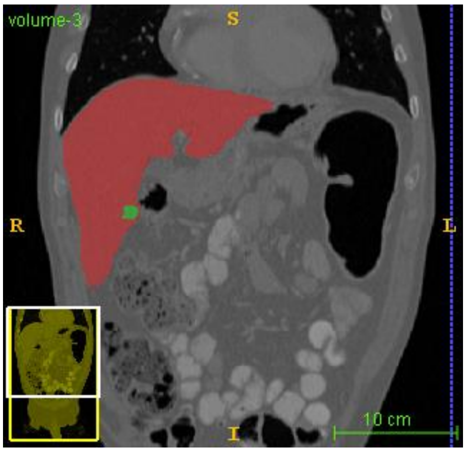
The three different views shown in Figure 2.1 are examples of axial, sagittal and coronal views respectively. The use of triplanar views provide more sufficient volumetric contexts and increase the representation abilities of a liver tumor in the abdominal view. The CT images dataset consists of enhanced abdominal CT scans which include the entire abdomen and thorax acquired from different medical centers.



Axial view (transverse)
Divides the body into upper and lower segments



Sagittal view
Divides the body into left and right sections



Coronal view (Frontal)
Divides the body into front and back sections

Figure 2.1 Orthogonal Views of Liver CT Images (LiTS 2017 Liver CT Dataset)

2.1.3 Overview of Convolutional Neural Networks Algorithms

The following section provides an overview of deep learning algorithms, specifically in the area of ConvNet. Among all these approaches, LeNet-5, designed by LeCun et al. (1998) is considered as the base of ConvNet network architecture, starting from the improved designed of AlexNet (Krizhevsky et al., 2012), to the more accurate and complex VGGNet (Simonyan et al., 2014) and the inception architecture of GoogLeNet (Szegedy et al., 2014). ConvNet has successfully improved classification and detection as demonstrated in the work carried out by Krizhevsky et al. (2012). Krizhevsky et al. (2012) achieved the top place in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC-2012) in almost halving the error rate of the image classification task.

Other recent success in image classification research works includes VGGNet, deeper neural networks, proposed by Simonyan et al. (2014) with the increased of the depth of ConvNet to 16-19 weight layers. VGGNet managed to secure the first and second places in localisation and image classification tracks respectively in ILSVRC 2014. GoogLeNet (Szegedy et al., 2014), also known as Inception Module, consists of 22 layers in deep, managed to achieve the winner for ILSVRC 2014. In He et al. (2016), ResNet-152 layers network was designed with residual blocks and skip connections and outperformed other methods in ImageNet 2015 Challenge.

2.1.3(a) LeNet-5

LeNet-5 is a five layers network architecture, consists of convolutional layers and pooling layers that were alternately connected. LeNet-5 architecture is being used effectively in handwritten character recognition (LeCun et al., 1990).

2.1.3(b) AlexNet

AlexNet (Krizhevsky et al., 2012) was introduced over a decade later after the creation of LeNet-5 (LeCun et al., 1998). AlexNet employed kernels with small receptive fields in layers closer to the output and larger kernels closer to the input. In their classification task, AlexNet incorporates Rectified Linear Units as the activation function. The summary of AlexNet's network configuration is as follow:

- 8 layers (5 Conv, 2 subsampling, 1 fully connected)

- Layer 1 Output
- Stride – 4
- Rectified Linear Units (ReLUs)
- Number of filters – 96
- Input Image size is – 224 x 224 x 3
- Filter size – 11 x 11 x 3
- Split across 2 GPUs – 55 x 55 x 48 for each GPU
- 224/4 x 224/4 x 96 = 55 x 55 x 96 (because of stride 4)

2.1.3(c) VGGNet

VGGNet (Simonyan et al., 2014) was trained by Visual Geometry Group from Oxford University which achieved very good performance in ImageNet dataset as first in localisation category and second places in the classification of an image in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) held in 2014.

ImageNet (J. Deng et al., 2009) consists of a large public images repositories which serve as a benchmark and resource for visual recognition applications like object recognition, object localization and image classification.

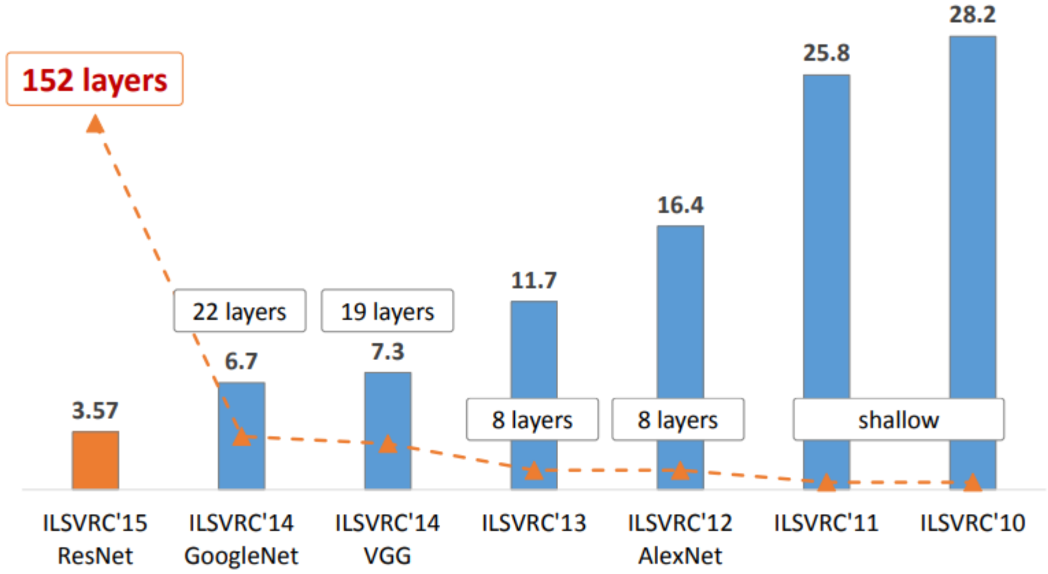


Figure 2.2 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) Winners

VGGNet's main contribution has shown that the depth of the network plays an important role and it is proven that deeper networks provide better performance results. There are only 3x3 convolution and 2x2 pooling used throughout the whole network. One of the drawbacks of VGGNet is that the network is it contains around 140~160M parameters which consumed a lot of memory and involved expensive evaluations. Most of the parameters are consumed in the FC layers. The summary of VGGNet's network configuration is illustrated in Table 2.1.

2.1.3(d) GoogLeNet

In GoogLeNet (Szegedy et al., 2014), a deep convolutional neural network known as Inception is proposed for ImageNet 2014 Challenge. GoogLeNet design utilizes increasing depth and width of the network model with 22 layers. GoogLeNet (Szegedy et al., 2014) was the winner for the ImageNet 2014 image classification challenge and GoogLeNet architecture design is able to maintain computational resources and provides significant decrease in error rate compared with AlexNet (Krizhevsky et al., 2012).

2.1.3(e) ResNet

In Deep Residual Network (He et al., 2016), the authors proposed the use of residual block and shortcut connections known as "identity shortcut connection" to deal with vanishing gradient problem and increasing the network depth. In ResNet architecture, the authors demonstrated five different variants of ResNet with different number of layers. ResNet 18 –layers, ResNet 34-layers, ResNet 50-layers, ResNet 101-layers and ResNet 152-layers. ResNet-152 is capable to overcome the vanishing gradient in neural networks through the skip connections network and flow towards he initial filters. ResNet-152 layers network was capable to breakthrough in ImageNet 2015 Challenge with increasing network depth and minimize the error rate.

Table 2.2 summarizes several main ConvNet architectures and summary of contributions in achiving better accuracy and reducing computational costs. The strenghts and the disadvantages of the ConvNet architectures are also presented in Table 2.1.

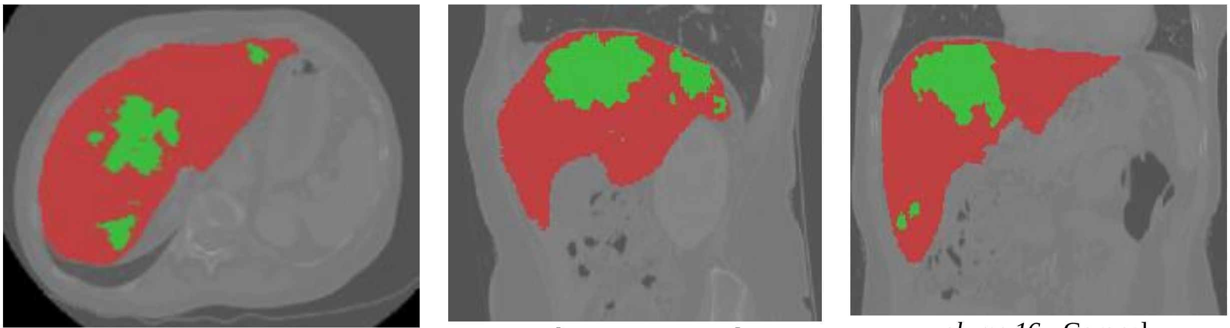
Table 2.1 Summary ConvNet Architectures

Research Work	**Pros and Cons**	**Summary of Contributions**
LeNet-5 LeCun et al. (1998) Consists of six layers 3 conv layers 2 subsampling layers 1 fully-connected layer	• Pros – Classify digits, recognize hand-written numbers with greyscale input images • Cons – Limited by computing resources	The work highlighted the use of ConvNet in end-to-end training with gradient-based optimization. LeNet-5 is considered as the first ConvNet performed on handwritten or machine-printed character recognition task.
AlexNet Krizhevsky et al. (2012) Consists of eight layers 5 conv layerrs 3 fully-connected	• Pros – Deeper with 8 layers. The network is with more filters and stacked conv layers • Cons - The network comprises of many hyperparameters	The work reported the breakthrough that employed ConvNet in ImageNet Large Scale Visual Recognition Challenge with top-5 error rate of 15.3%, compared to 26.2% achieved by second-best entry.
GoogLeNet Inception Szegedy et al. (2014) Consists of 22 layers	• Pros – Used inception module by stacking multiple kernels at the same level which further increases the accuracy and decreasing computational costs • Cons - The network comprises of many hyperparameters	The Inception module takes convolutions at multiple layers and provides a good receptive field as well as reducing the overall number of parameters

VGG net Simonyan et al. (2014) Consists of 16/19 layers 11 conv layers 5 max-pooling layers 3 fully-connected layers	• Pros – managed to solve the issue of large amount of hyperparaments in AlexNet. • Cons – longer training time, vanishing gradient	The main contributions of VGG net is the evaluation of networks by increasing depth which shows that significant of increase in accuracy for localisation and classification tasks by increasing the depth to 16-19.

2.2 Liver Tumor Characteristics

Liver tumor appear in different shapes, sizes and locations in liver tumor patient. Figure 2.3 illustrates the characteristics of liver tumor for patient *volume-16.nii* which shows liver tumors with variations of shapes, sizes and with no clear boundaries in the three planes. The liver is indicated in the red region and the tumor is highlighted in green. The dataset consists of enhanced abdominal CT scans which include the entire abdomen and thorax.



volume 16 - Axial   *volume 16* - Sagittal   *volume 16* - Coronal

Figure 2.3 Variation of shapes, sizes and locations

As observed in Figure 2.4, X-axis represents patients of liver tumor and Y-axis represents their tumor size. There are multiple blue dots which indicate liver tumor with smaller size and the red dots represent the liver tumor which are large in size. The red dots the liver CT dataset has large variations of the size of the tumor, ranging from 0.018% to 31.41% with respect to the overall volume of the liver. In 2017 (Bilic et al., 2019) and 2008 (X. Deng et al., 2008), two Grand Challenges which focused on the benchmark dataset of liver and liver lesion segmentation have been conducted in conjunction with Medical Image Computing and Computer Assisted Intervention (MICCAI) conference.

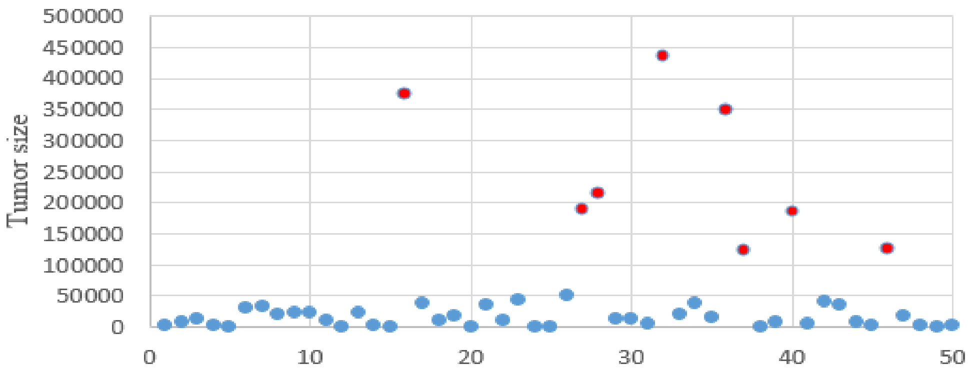


Figure 2.4 Liver Tumor Dataset (*volume-0.nii – volume-50.nii*)

Blue dots indicate tumor in small size and red dots indicate tumor in large size.

Figure 2.5 provides the frequency histogram for the liver tumor and non-liver tumor intensities. The histogram shows the distribution of pixels in tumor and non-tumor tissues found in the CT images. As we observed, the intensities are overlapped throughout the range of values from 0 to 255 for the two regions. The information depicted in the histograms provide an estimation of segmentation of pixels of an image into distinct classes based on their intensities.

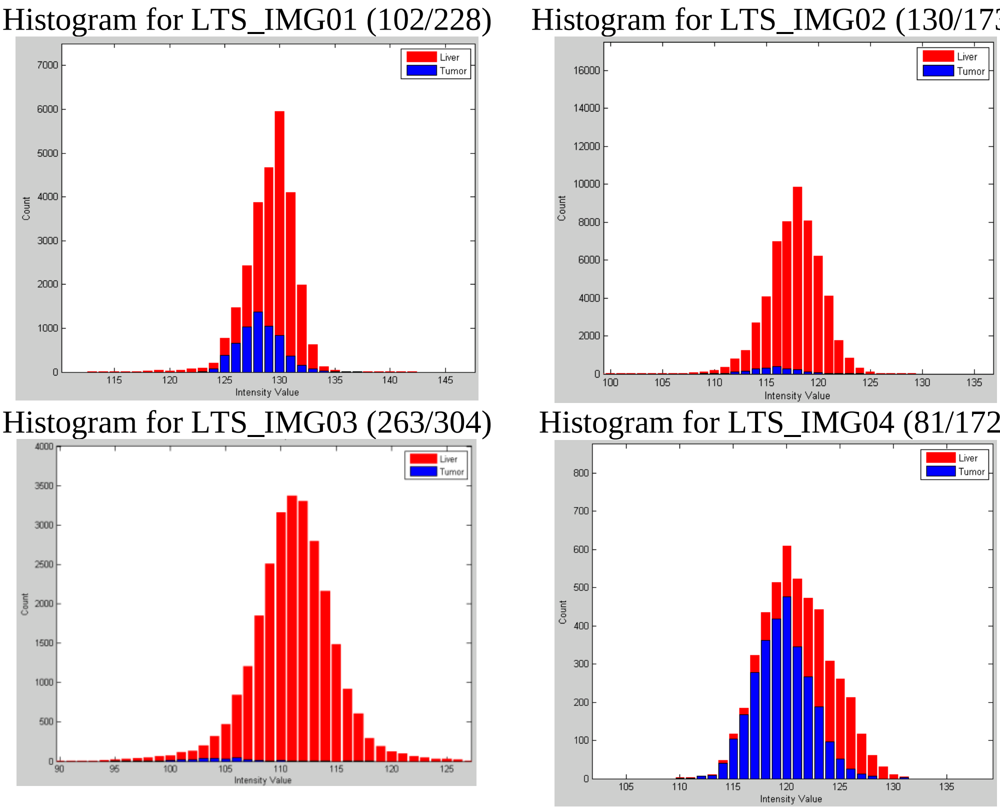


Figure 2.5 Frequency Histogram for the Liver Tumor and Non-liver Tumor Intensities (MICCAI Liver CT dataset)

2.3 Related Works

In Section 2.3.1, various methods developed for liver tumor segmentation is presented. Section 2.3.2 covers the review of liver tumor segmentation based on Deep Learning. Section 2.3.3 describes the use of ConvNet for liver tumor segmentation.

2.3.1 Liver Tumor Segmentation

In liver tumor segmentation, accurate measurement of liver tumor size from liver CT images is essential and useful liver tumor diagnosis. Due to different contrast enhancement behaviour of liver parenchyma and lesions, automatic detection and segmentation are considered time consuming and difficult tasks as it often depends on clinical knowledge and experience of the clinical radiologists. One of the main difficulties of liver tumor segmentation is the individual differences in scan time and perfusion which may cause low image contrast between healthy liver tissues and liver tumor tissues. Furthermore, liver tumor shape, size and texture may vary from patient to patient which contribute to the difficulties of liver tumor segmentation.

In this study, we review several liver tumor segmentation algorithms. Shimizu et al. (2008) trained a boosting algorithm, AdaBoost for ensemble segmentation and applied to the problem of liver tumor extraction. In the ensemble segmentation algorithm, (Shimizu et al., 2008), a sequence of weak hypotheses were obtained and found to be suitable to train the classification algorithm. The weak hypotheses consist of (i) CT value features computed from a local region, (ii) convergence index filter and (iii) sobel filter. These weak hypothese were combined by a weighted majority vote to construct a strong ensemble classifier for liver tumor segmentation.

Massoptier et al. (2008) applied a statistical model-based approach for liver lesions segmentation. It was combined with an Active Contour Algorithm for obtaining a smoother liver segmentation using gradient vector flow. In the clustering approach by Hame (2008), a rough segmentation is firstly obtained by thresholding and morphological operations and further refined the rough segmentation results using fuzzy clustering and geometric deformable model.

Choudhary et al. (2008) performed liver tumor segmentation using minimum cross-entropy multi-thresholding algorithm and liver segmentation from the organ structure using the watershed algorithm. The segmentation of the tumor is further developed with 3D level-set based surface smoothing and region growing techniques. The 3D level sets are applied to the bidimensional contours obtained.

In clustering approach, Mandava et al. (2011) demonstrated a kernel-based clustering algorithm that incorporate Tsallis entropy for resolveing long-range interactions between healthy and tumor tissue intensities in the tumor segmentation process. The Tsallis entropy principle is included by using Fuzzy C-Means (FCM) clustering algorithm to cater to long range interactions and maximize the separability between the two liver classes. Mandava et al. (2011) work focused on the intensity difference between healthy and tumor tissue and extends the generalized C-Means algorithms with kernelized distances for tumor delineation. The work involved delineating and localizing tumor regions based on differences of intensity.

Zhou et al. (2008) used Support Vector Machine (SVM) classifier for extracting liver tumor by demonstrating a semi-automatic segmentation technique. Freiman et al. (2011) used SVM classification engine to classify liver voxels into healthy tissue classes and tumor. The conjugate gradients method optimised a set of functional linear equations and produce a continuous segmentation map to output a binary segmentation.

Zhang et al. (2011) presented a liver tumor segmentation technique to extract tumor from liver parenchyma using Support Vector Machine (SVM) trained on user-selected seed points. The feature vector for prediction and training is calculated based on each small region produced by a watershed transform. The rough segmentation result of SVM classification is refined by performing morphological operations on the whole segmented binary volume.

Rodrigues et al. (2011) presented an interactive liver tumor segmentation evaluated on CT and MR data using an intelligent paintbrush to paint the object of interest. The approach applies the pseudo-watershed algorithm on an image gradient magnitude to partition homogeneous primitives regions. The best image partition based on the Minimum Description Length (MDL) principle is found by feeding the output of the initial segmentation into an efficient region merging process in reducing the over-segmented image.

An automatic liver initialization method based on a statistical model distribution of liver average intensity and its standard deviation was developed by Casciaro et al. (2012) to produce fully automatic segmentation based on two different segmentation techniques namely Graph Cut and gradient flow Active Contour Algorithm. The two different methods were evaluated using the evaluation framework and the initialization technique was applied for segmentation of hepatic tumors and liver tissue from abdominal CT images. Graph Cut algorithm has shown superior overall performance in terms of processing time and accuracy. The research works involved assessment of segmentation accuracies through the evaluation of Dice Similarity Coefficient (DSC), False Positive Ratio (FPR) and False Negative Ratio (FNR).

Table 2.2 summarizes several classic liver tumor segmentation research work, main methods employed and important key points undertaken by the authors.

Table 2.2 Liver Tumor Segmentation Research Work and Methods

Research Work	**Methods**	**Mode**
Shimizu et al. (2008)	Adaboost	• Automatic
Massoptier et al. (2008)	Statistical model-based approach	• Automatic

Hame (2008)	Fuzzy clustering & Geometric Deformable Model	• Interactive
Zhou et al. (2008)	Support Vector Machine	• Semi-automatic
Stawiaski et al. (2008)	Graph-cuts	• Interactive
Choudhary et al. (2008)	Minimum cross-entropy Multi-Thresholding	• Semi-automatic
Mandava et al. (2011)	Kernal-Based FGCM and PGCM	• Automatic
Freiman et al. (2011)	Support Vector Machine	• Interactive • User defined seeds
Zhang et al. (2011)	Support Vector Machine	• Interactive
Rodrigues et al. (2011)	Pseudo-watershed Algorithm	• Interactive
Casciaro et al. (2012)	Graph Cut & Gradient Flow Active Contour Algorithm	• Automatic

2.3.2 Deep Learning for Medical Images

The increasing interest in anatomical structures segmentation in medical imaging has seen a rapid growth of the use of deep learning approaches in analyzing medical images (Litjens et al., 2017). The use of deep learning is surveyed for the following area such as segmentation of biological neuron membrane (Ciresan, Giusti, et al., 2012), bone tissues (Cernazanu-Glavan et al., 2013), manifold learning of brain (Brosch et al., 2013) and detection of multiple organ (Shin et al., 2013).

In this section, we review and highlight the research works undertaken by different deep learning methods. Table 2.3 summarizes the research works done for