
VHL Gene Mutation Prediction of Clear Cell Renal Cell Carcinoma Based on CT Images

UNIVERSITY OF TURKU

Department of Future Technologies

Master of Science in Technology Thesis

Computer Systems

October 2019

Wenshuai Zhao

Supervisors:

Prof. Juha Plosila

Prof. Yuanyuan Wang

Dr. Mohammad Hashem Haghbayan

UNIVERSITY OF TURKU
Department of Future Technologies

ZHAO WENSHUAI: VHL Gene Mutation Prediction of Clear Cell Renal Cell Carcinoma Based on CT Images

Master of Science in Technology Thesis, 69 p.

Computer Systems

October 2019

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cell carcinoma (RCC) with the highest mortality rate among genitourinary cancers. Even though it is complicated to find the exact causes for such disease, some research has shown that patients' genetic statuses are important risk factors, such as the prevalent von Hippel-Lindau tumor suppressor (VHL) gene mutation. Therefore, identification of corresponding gene mutations has huge potential to help determine the precise treatment and predict prognosis. However, it is difficult to promote this diagnostic technique because current gene sequencing relies on the invasive sampling of tumors, which is expensive and will bring huge physical impact on patients.

Radiomics and radiogenomics are new proposed concepts and treat medical images as minable data source to explore underlying pathophysiology information and genetic causes, which innovatively provides a perspective to develop noninvasive methods to identify the gene mutation status. Inspired by this concept, in this dissertation we focused on ccRCC and attempted to develop an accurate prediction model to detect the VHL gene mutation status based on patients' computer tomography (CT) images. Three major works have been done to deploy this research: semi-automatic segmentation of kidney tumor in CT images, fully automatic segmentation of kidneys and kidney tumor and prediction of VHL gene mutation.

Segmentation of kidney tumor is an important step for the next radiomic analysis. As a wise compromise, we proposed a 3D U-Net based semi-automatic segmentation method. This network employed the manually anchored region of interests (ROIs) as its input, and comprised only four layers rationally. To mitigate the risk brought by the limited dataset, a variety of data augmentation was adopted both in the training and testing. In addition, the Soft Dice was used to replace the traditional Dice loss to enhance the performance. Experiment on the independent testing dataset showed the Dice up to 0.87.

In order to further reduce the workload of radiologists, a fully automatic segmentation method was proposed. Based on the naïve 3D U-Net, we devised multi-scale supervision to encourage the network to make correct prediction from deeper low-resolution layers, which were then up sampled to form the final full-resolution prediction. At the same time, the exponential logarithmic loss was combined with Soft Dice to mitigate the severe sample imbalance. A connected-component based post processing method was also designed to utilize the prior knowledge to improve the segmentation accuracy further. This simple end-to-end network showed powerful performance in our independent testing dataset and also outperformed most algorithms in the KiTS19 challenge.

A classical prediction model was proposed to detect the VHL gene mutation. First, we employed a pre-trained multi-scale supervised 3D U-Net to segment the tumor area in CT images. Then 460-dimension radiomic features were extracted from tumor slices. Also, a sparse-representation-based method was adopted to select discriminating features. At last, the experiment on testing dataset showed the accuracy of 0.8 by using an ensemble subspace k-nearest neighbors (KNN) as the classifier.

In general, this dissertation progressively researched and developed the prediction method of VHL gene mutation based on CT images. It took advantage of deep learning and proposed useful and practical techniques to segment the kidney and kidney tumor. Based on the accurate segmentation, a typical machine learning method was employed to construct the prediction model.

Keywords: ccRCC; Segmentation; Classification; 3D U-Net

Contents

Chapter 1 Introduction	1
1.1 Background of the Research	1
1.2 Characteristics of CT Images	3
1.3 Challenges for CT Images Analysis of ccRCC	4
1.4 Related Researches.....	7
1.4.1 Current Situation of Radiomics and Radiogenomics	8
1.4.2 Current Situation of Medical Image Segmentation.....	10
1.5 Major Works and Highlights of the Research	12
1.6 Data and Experiment Conditions	13
Chapter 2 Semi-automatic Segmentation of Kidney Tumor in CT Images.	15
2.1 Introduction	15
2.2 Segmentation based on 3D U-Net.....	18
2.2.1 3D U-Net Architecture	18
2.2.2 Training Settings	19
2.2.3 Inference.....	20
2.3 Experiment and Results	21
2.3.1 Data Preparation.....	21
2.3.2 Results and Analysis	22
2.4 Summary	25
Chapter 3 Automatic Segmentation of Kidney and Kidney Tumor in CT Images	27
3.1 Introduction	27
3.2 Multi-Scale Supervised 3D U-Net	28
3.2.1 Preprocessing and Data Augmentation	30
3.2.2 Architecture.....	32
3.2.3 Training Procedure	34
3.2.4 Inference and Post Processing.....	36
3.3 Experiments and Results	37
3.3.1 Experiment Data and Set Up.....	37

3.3.2	Results and Analysis	38
3.4	Summary	43
Chapter 4	VHL Gene Mutation Prediction of ccRCC in CT Images	45
4.1	Introduction	45
4.2	Radiomic Features Based Prediction Model	46
4.2.1	Introduction of Radiomic feature Sets	47
4.2.2	Features Selection	51
4.3	Experiments and Analysis	55
4.3.1	Data Preparation.....	55
4.3.2	Image Segmentation.....	56
4.3.3	Multi-Slice Based Experiment	58
4.4	Summary	61
Chapter 5	Conclusion and Discussion	62
5.1	Conclusion	62
5.2	Future Work.....	63
References	65
Acknowledgement	69

Chapter 1

Introduction

1.1 Background of the Research

Renal cell carcinoma (RCC) or kidney cancer, referring to cancerous renal masses, consists of a variety of cancers originated from renal tubular epithelial cells [1] and has the highest mortality rate of genitourinary cancers [2]. It accounts for about 2% of all cancer diagnoses and cancer deaths worldwide, and there are 295,000 new cases diagnosed and about 134,000 recorded every year [3], [4], [5]. Based on pathology and genetic abnormalities, RCC can be categorized as kinds of subtypes [6]. Major subtypes [7] occupying more than 5% incidence are clear cell RCC (ccRCC), papillary RCC (pRCC) and chromophobe RCC (chRCC) [8], [9], [10], among which ccRCC lead to the majority kidney cancer deaths.

In order to investigate the causes of RCC, scientists conducted a lot of research and concluded genetic and acquired risk factors [11]. Acquired risk factors usually indicate bad habits such as smoking, hypertension, obesity, chronic analgesic use, and diabetes [13]. They always vary from person to person. Relatively, the genetic risks draw more attention of researches due to their universality. The most common gene mutations involved in ccRCC can be attributed as von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase (VHL) gene located on the short arm of chromosome 3 (3p25), SET domain containing 2 (SETD2), polybromo 1 (PBRM1), lysine (K)-specific demethylase 5C

(KDM5C), and BRCA1-associated protein 1 (ubiquitin carboxy-terminal hydrolase) (BAP1) [12]. Due to the predominance of ccRCC and its main genetic factor VHL gene, they are chosen as the focus of this dissertation.

Identification of genetic mutations is clinically important because they are correlated with advanced stage, advanced grade, and patients' possible survival [14], [15]. Therefore, it has great prognostic value to get known of the genetic composition of a patient's ccRCC. Traditionally, mutations are identified through genome sequencing of ccRCC from kidney samples. In this way, it is necessary to obtain invasive nephrectomy or kidney biopsy [14]. However, this invasive detection method costs a lot and will bring great physical impact on the patient, and thus obstacles the clinical application of genomics.

Recently, some studies have shown that there are correlations between gene mutations and different phenotype features shown in patient CT images [12], [16]. For example, the VHL gene status is significantly associated with certain characteristics of ccRCC in contrast-enhanced CT images, such as nodular tumor enhancement ($P=0.021$), well-defined tumor margins ($P=0.13$) and gross appearance of intra-tumoral vascularity ($P=0.018$).

Similar with radiomics, the term of radiogenomics established in recent years is to build an integrated approach between radiology and genomics [17], [18]. In this concept, medical images are not only used to interpret a tumor's morphology, but also treated as the valuable data source to explore the underlying molecular and genomic alterations.

As more digital data including medical images are easily available and the huge improvement of computing has been achieved, machine learning [19], especially deep learning [20] have shown their powerful capability to mine the complex patterns hiding in the big data. Meantime, successful practices regarding computer vision [21] and natural language processing [22] have drawn researchers' attention.

With such background, this dissertation aims to study the relationship between phenotypical features extracted from CT images and the underlying VHL gene mutation. A machine learning based prediction model was developed to help clinicians learn the gene mutation status without using traditional invasive gene sequencing.

1.2 Characteristics of CT Images

Computed tomography (CT) [23] plays a major role in the evaluation of a patient with a suspected renal mass due to its easy accessibility and high quality. It not only helps to detect the mass but also could reflect its characteristics. Hence CT images could provide essential accordance for doctors to stage and plan surgery of the patient.

CT images acquired through different protocols could show distinct characteristics, which would lead to different intensity distribution and bring obstacles in training machine learning models. In common cases, renal CT images can be divided into two kinds, unenhanced scans and contrast material-enhanced scans. Usually there are four different phases of renal enhancement. They can be imaged based on the acquisition time after the injection of non-ionic contrast in a large antecubital vein [24]: (1) the arterial phase, (2) the corticomedullary phase (also called angionephrographic or late arterial phase; CMP/AP), (3) the nephrographic phase (also called the parenchymal venous phase, NP/VP), and (4) the excretory phase (EP). The different enhancement phases are shown in Fig. 1.1.

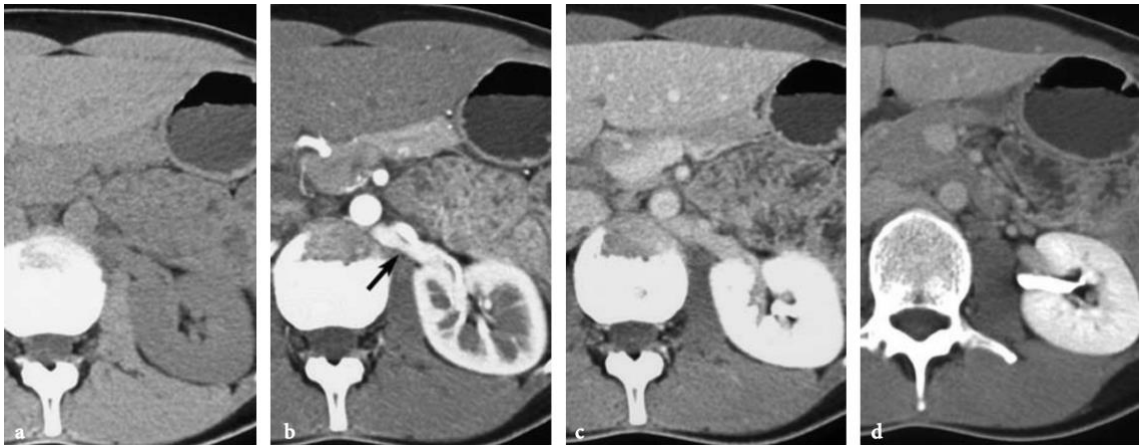


Fig. 1.1: Different phases of renal enhancement at CT images of the same tissue. (a) Precontrast CT scan. (b) AP scan. The arrow indicates its good opacification of the retro-aortic renal vein. (c) VP scan. (d) EP scan.

Clinically, in order to fully evaluate renal masses, at least three acquisition sequences are required to detect and characterize renal lesions, and at the same time it is not encouraged conduct more scans to minimize the radiation to patients. The unenhanced scans provide baselines to assess the enhancement within the lesion after the injection of intravenous contrast. Other two different chosen enhanced scans would be performed to distinguish hyperdense cysts from solid tumors.

1.3 Challenges for CT Images Analysis of ccRCC

The workflow of prediction of VHL gene mutation follows the basic process of radiomics, which can be summarized as two categories, i.e. traditional radiomics and deep-learning based methods, shown as Fig. 1.2.

Currently, the basic flow for traditional CT images analysis includes three steps, as shown in Fig.1.2 (a): first, as the raw images acquired, it is necessary to perform segmentation of the kidney tumor; second, manually designed or radiomic features should be extracted from the lesion or targeted tissue area and then select the most significant features containing the distinguishing characteristics to represent diseases or clinical status; last, from selected features, experts would employ classifiers to diagnose or predict some clinical information, such as the VHL gene mutation status in this dissertation.

In contrast to the traditional radiomics, deep learning doesn't need specially defined features, but learns to extract features automatically by a neural network, especially the

powerful convolutional network. As we input images or ROIs, through back propagation, deep learning adapts parameters or weights to decrease the loss function, which enable the network to extract self-designed features and perform best prediction, shown as Fig. 1.2 (b). The shortcomings of deep learning based prediction methods are their data thirsty natural and poor interpretability.

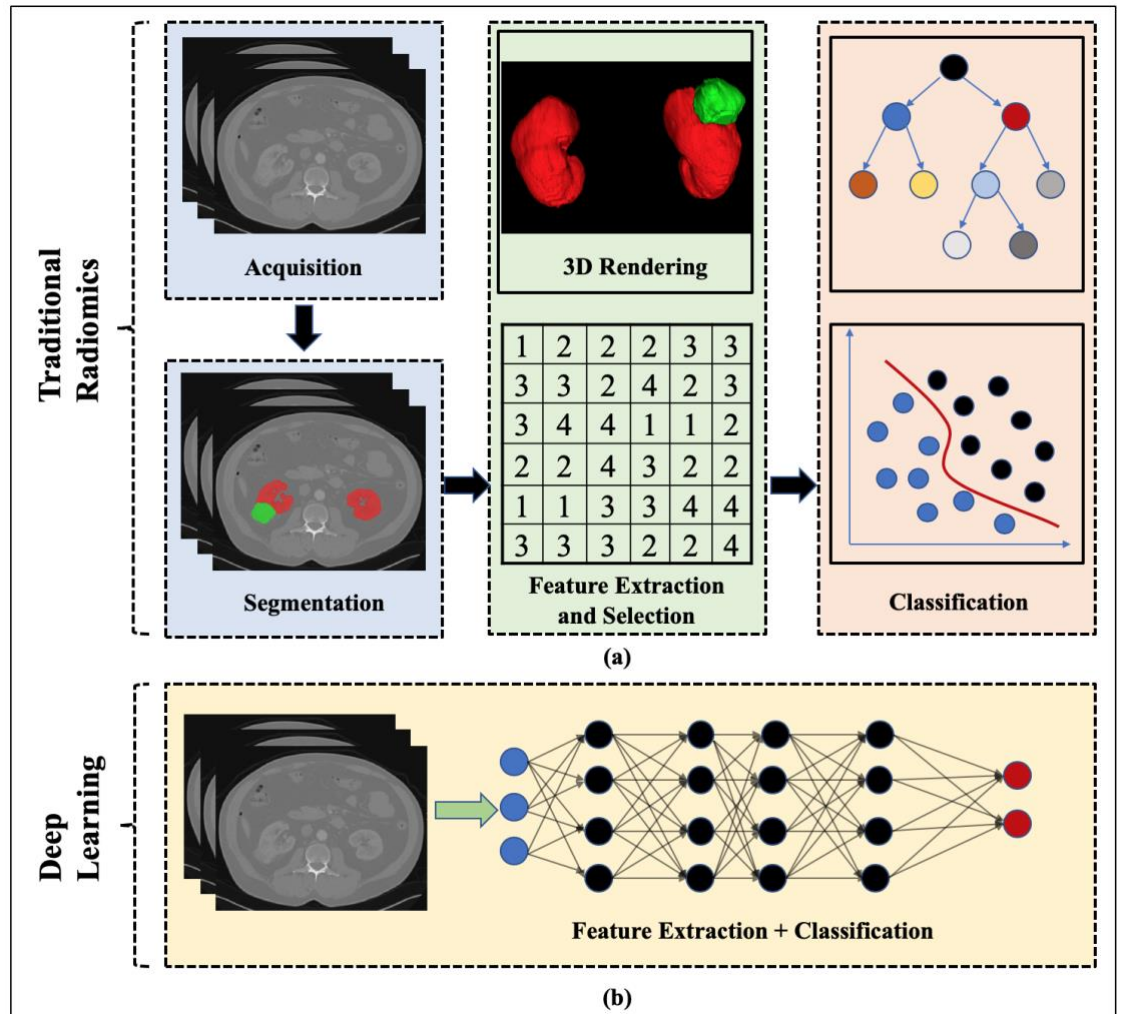
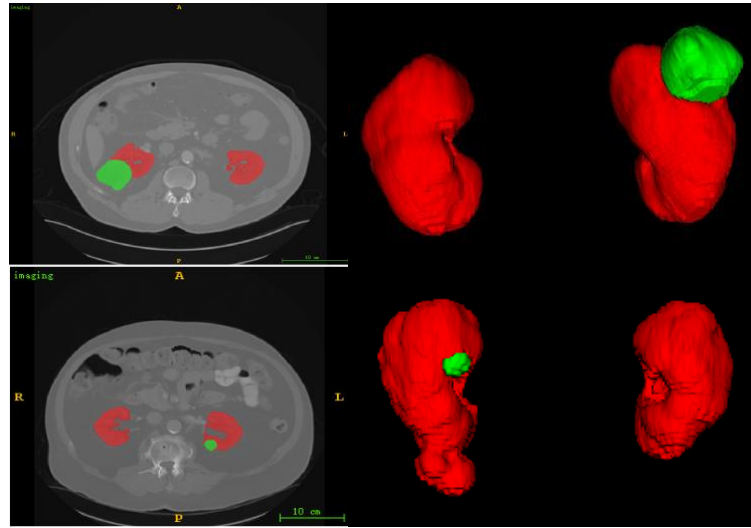


Fig. 1.2: Over-simplified representation of radiomics. (a) The process of traditional radiomics includes image segmentation, feature extraction and selection, and

classification. (b) Deep learning based radiomics integrates autonomous feature extraction and classification into the network with multi layers.

According to the needed stages, there are three technical challenges to settle. First and foremost, the accurate segmentation of ccRCC area is an important precondition for the subsequent feature extraction, selection, and prediction, because it produces a rich quantitative representation of tumors and affected organs. Even for deep learning, accurate segmentation could reduce most input area, and providing smaller ROIs to the network will promote the convergency and its accuracy.



(a)

(b)

Fig. 1.3: Kidney and tumor in abdominal CT images, (a) transverse plane of abdomen where the red area is healthy renal tissue while the green area means the tumor, (b) 3D view of the segmentation of kidney and tumor.

Two segmentation samples of the kidney and tumor is shown as Fig. 1.3. As what can be seen, the kidney and tumor with the irregular shape and varied relative size and

location would consume much time and energy to manually segment. Simultaneously, the above characteristics also play an important role to challenge current computer assisted segmentation methods and diminish their performance. Therefore, as the basis of analysis, the proposal of effective automatic segmentation methods is one significant goal of this dissertation.

Second, although medical images contain abundant clinical and diagnostic information, deeply hided traits are hard to be seen directly by human. Thus it is a huge challenge how to extract these useful features and quantify them to provide diagnostic foundation or information for prediction and prognosis. This is also the main purpose of this dissertation.

At last, deep-learning based methods are always data thirsty, which is especially difficult for medical images. Moreover, the principle of deep learning is still not clear, and appearing as black-box while clinical applications require good reliability and interpretability. Therefore, aiming for how to utilize the powerful feature representation capacity of deep learning under such limitations, this dissertation also makes some attempts.

1.4 Related Researches

This section will describe the related researches. At first, the basic concept and common methods used in radiomics and radiogenomics will be introduced, since these are the inspiration of this dissertation. Because segmentation is the cornerstone for the subsequent analysis and thereby constitutes the essential part in this dissertation, the second part will present the common segmentation algorithms currently, especially deep learning due to its tremendous performance in various computer vision tasks.

1.4.1 Current Situation of Radiomics and Radiogenomics

As the computing capacity increases dramatically, it is now feasible to rapidly extract countless quantitative features from tomographic images, such as CT, magnetic resonance (MR) images [25]. Radiomics is inspired by the idea that biomedical images can be mined more underlying pathophysiology information because the intensity variation might imply more than what can be seen. Under this concept, radiological images are not just intended solely for visual interpretation, but also could be employed to extract sub-visual, yet quantitative, image features. Likewise, radiogenomics specifically denotes mining more information from radiomic data in order to detect the correlations between genomic patterns, and determine genetic causes for variations in radiosensitivity [26], [27].

The process of radiomics or radiogenomics conventionally involves four steps [28]: image acquisition, organs or lesions segmentation, quantitative features extraction and selection, and building predictive and prognostic models. Radiomics presents notable advantage for aided diagnosis since it does not significantly disrupt clinical workflow. Once the radiography is acquired, the next steps including segmentation, feature extraction and analysis are all conducted based on images.

Currently, features used to analyze or build models are typically categorized as two kinds: semantic features and radiomic features, or called quantitative features. Semantic features are observed and described directly by radiologists, with the resistance to inaccurate segmentation, and image acquisition parameters variation. They can also be used to build classifiers. Examples of these features for the prediction of gene mutation in ccRCC include cystic tumor architecture, ill-defined tumor margin, calcification, and nodular enhancement etc. [12], as shown as Fig. 1.4.

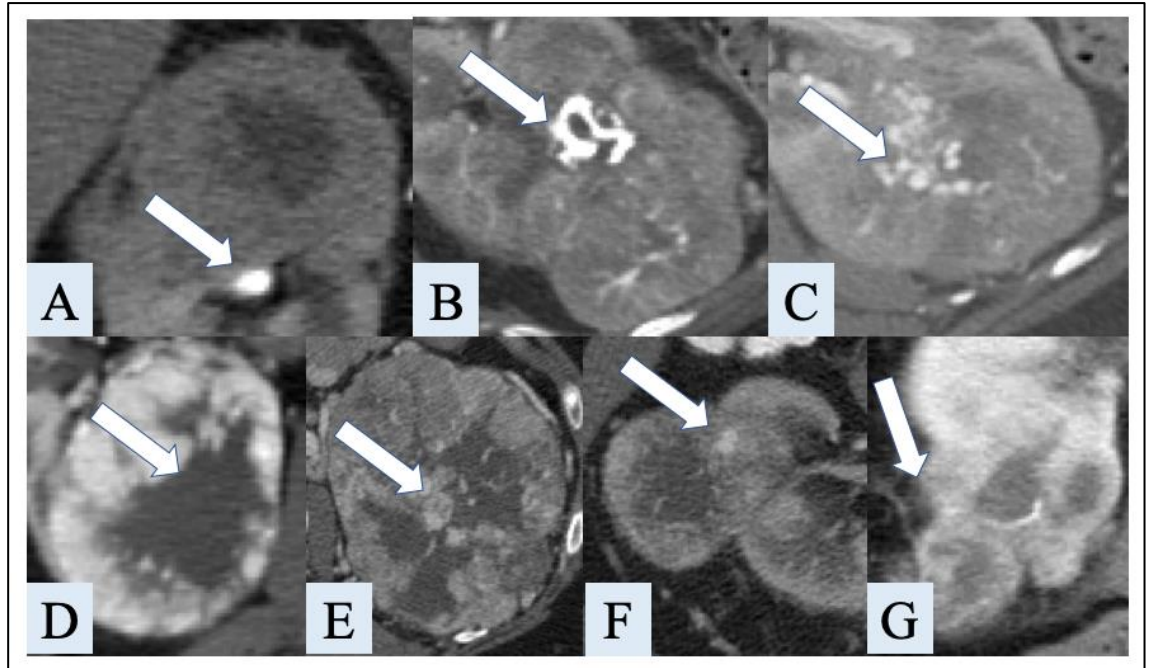


Fig. 1.4: Semantic features identified by experts related to with underlying gene mutation status of ccRCC. A, Calcification. B, Collecting system invasion. C, Gross appearance. D, Necrosis. E, Nodular enhancement. F, Renal vein invasion. G, Ill-defined margin. Arrow indicates feature of interest in each image.

In order to utilize the computation and mitigate the energy and time consumed by radiologists, quantitative features were proposed in radiomic analysis. They are expected to be able to mine not only the visible features but also the invisible information hiding in medical images. Common radiomic features are currently concluded as the following classes: intensity-based [29], structural [30], texture/gradient-based [31], and wavelet [32]. These features are sub-visual image information and extracted by computations. Since these features can be achieved without the expertise and more easily generalized, this dissertation just follows this way.

Even though some semantic features manually extracted from ccRCC mentioned above have been proved correlated with the VHL gene mutation in several studies, there

are still not any specifically designed radiomic features validated in such task. Therefore, in this dissertation we attempted to use generic radiomic features and select distinct features to predict the underlying gene mutation status.

1.4.2 Current Situation of Medical Image Segmentation

Accurate segmentation is fundamental for selectively gathering radiomic features. It provides the exact area of specific tissues or lesions from the huge surrounding tissues, which can be noises for extracted features. Manual segmentation is usually conducted by professional radiologists, and it possesses good accuracy and resistance to the image parameter variation. However, manual segmentation is cumbersome, and easily affected by inter-observer variability. Therefore, semi-automatic and automatic methods are proposed to relieve the burden of radiologists.

Semi-automatic methods combine the efficiency and accuracy. They aim to reduce the processing time or provide an initial segmentation result to radiologists. They always require experts to mark a box to define a region of interest (ROI), some common algorithms are employed to process the segmentation task in the ROI. This can reduce large bulk of pixels surrounding the target tissues to be recognized and hence get a better performance. Another kind of semi-automatic method is interactive segmentation. Typically, in this way, users should provide seed pixels to the active contour model and the algorithm would automatically conduct the next segmentation.

Obviously, a fully automatic segmentation is the final desire that scientists pursue. There are various segmentation methods proposed in the past decades [33]. (1) Threshold-based algorithms simply set a threshold to divide the pixels into background or foreground. (2) Edge detection-based algorithms employ edge detection operators to distinguish the images into different semantic objects. (3) Region-growing methods

compare pixels with their neighbors and conduct cluster. (4) Partial differential equation-based methods evolve an initial curve to obtain the lowest defined cost function. (5) Graph partitioning methods assume different semantic objects are homogeneous in images, and employ the inspirit of undirected graph to describe the pixels.

However, traditional segmentation algorithms rarely achieve sufficient accuracy and robustness to be clinically useful. Various reasons lead to the unsatisfying performance, such as poor image quality, dramatic variations among patients, heterogenous appearances brought by pathology, etc.

As deep learning with convolutional neural networks (CNNs) has achieved tremendous performance in computer vision tasks, current state-of-the-art methods for segmentation are usually deep learning based. Fully convolutional net (FCN) [34] proved an end-to-end easily constructed deep learning model with comparable segmentation accuracy, and drew great attention employing deep learning to handle the segmentation tasks. Afterwards U-Net [35] has been validated as one of the most robust architectures for medical images segmentation, and miscellaneous U-Net based architectures also proposed for different datasets.

Compared with natural images, medical images present huge differences and always appear in a volumetric form. According to the different treatments, researchers proposed two kinds of methods to segment medical images. Traditionally, limited by the algorithm complication and huge computation requirement, original 3D volume images are always sliced into 2D slices, which treats medical images just as natural 2D images and neglects their 3D characteristics. Recently, as the computation capacity increases, it is feasible to directly train 3D networks to segment volume images and take advantage of their specific features [36][37][38]. Note that they all have their own advantages in specific situations. Such as, 3D networks could take use of inter-slice features and achieve more smooth

segmentation appearance, while 2D networks consume much less GPU memory and can be more easily converged.

1.5 Major Works and Highlights of the Research

The goal of this dissertation is to build an accurate prediction, or determination, model for the detection of VHL gene mutation in ccRCC. The main work can be divided into two principal parts: (1) segmentation of kidney tumor based on abdominal CT images, (2) building the prediction model to detect if the ccRCC in patients present VHL gene mutation. Regarding these two tasks, there are three highlights attributed as follows:

(1) Proposed one deep learning based semi-automatic segmentation method. In this workflow, ROIs, as a form of box, were first proposed manually to anchor the specific position and area of tumor. To fully take advantage of 3D characteristics of CT images, a simple but powerful 3D U-Net architecture based network was designed and achieved pretty good performance in the test.

(2) Proposed a multi-scale supervised 3D U-Net to segment kidneys and tumor simultaneously in a fully-automatic way. Such supervision strategy could help the network make correct predictions from the deep layers and hence lay better foundation for the final output in the top layer. The exponential logarithmic loss was also employed in such method, encouraging the model to pay more attention to tumors which was more difficult to recognize. In addition, a specific connected component based post-processing method was also designed to remove obviously mistaken pixels. Experiments and broad evaluations showed its state-of-the-art performance and superiority compared to other algorithms.

(3) Proposed a CT images based prediction method for the VHL gene mutation in ccRCC. Under the premise of accurate segmentation of the kidney and tumor in CT images using the proposed segmentation network above, totally 460 dimensions high-throughput features were extracted including the shape, texture, and wavelet. A sparse representation based method was then adopted to select most discriminating features [49]. At last, it employed a typical classifier to predict if the VHL gene mutation presented in this patient.

1.6 Data and Experiment Conditions

There are two datasets used in this dissertation. Meanwhile, it is not necessary to apply any ethical approval for our experimental retrospective study because all data are public and freely accessed for scientific study.

To train and validate our segmentation models, the cohort of 210 patients' abdominal CT images with their ground truths from KiTS19 [39] was utilized. Unfortunately, this dataset does not contain gene mutation labels. Therefore, another 267 patients' CT scans were accessed from The Cancer Genome Atlas-Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) database [40], and we fetched their corresponding ground truth mutation labels from the cBioPortal for Cancer Genomics [41]. Note that many patients' scans from such cohort do not have the same acquisition protocol with the first KiTS19 dataset, so it is difficult to achieve satisfying segmentation results from the above trained model. In addition, several patients acquired CT scans after nephrectomy. Therefore, we designed inclusion and exclusion criteria to select patients suitable for our experiments. More descriptions about these data would be introduced in the following corresponding sections.

The experiments in this dissertation were basically conducted using python3.7, with the support of deep learning framework Pytorch (<https://pytorch.org/>). The extraction and selection of features was deployed on Matlab2018b (<https://www.mathworks.com/>). The computing server used in this dissertation is with two Nvidia Telsa (32Gb) GPUs.

Chapter 2

Semi-automatic Segmentation of Kidney Tumor in CT Images.

2.1 Introduction

Accurate segmentation is not only the foundation to assess the tumor from a morphology view, but also an important step for the subsequent radiomic analysis. Quantitative features are usually extracted from such segmented regions, and significantly affected by the accuracy of segmentation.

Traditional segmentation algorithms including threshold based methods, edge detection based methods, area based methods or graph cut based methods are now outperformed by deep learning based models. FCN [34] is the first notable deep learning based network for image segmentation. It wisely replaces the last full connection layers in typical classification networks by a convolution network to up sample the extracted feature maps, and hence realizes end-to-end segmentation.

Thereafter, huge amount of segmentation networks based on deep learning has been proposed for various tasks including both natural images and medical images. Due to the complex pattern and rich characteristics, segmentation networks for natural images are always much deeper and more complex than that for medical images, which usually vary

little and have much similar appearance between different images. Therefore, many effective models with tremendous performance in natural image tasks may become invalid when transplanted for medical applications.

However, even medical images are usually simpler and duller than natural images, it is still often challenging to segment tiny organs or lesions from a CT scan, as the objects are always too small to easily detect. Deep neural networks with successful applications on natural images, such as notable FCN [34], Deeplab [42], can be easily confused by complicated background and relative smaller foreground in medical images. Even though MaskRCNN [43] is a two-stage segmentation network and has achieves remarkable success in natural images segmentation, its performance was still very poor when applied in medical images because the kidney tumor with unclear margin and variable appearance is hard to anchored at the first stage, hence the performance of the next segmentation step executed on the previous inaccurate region would not be guaranteed.

Therefore, semi-automatic segmentation is an intelligent compromise considering the actual accuracy and efficiency. It needs experts or radiologists first label out a smaller region containing the targeted lesions from whole scans, which would not take too much time or energy, because the next refining segmentation is processed by the proposed segmentation model automatically. This workflow is shown as Fig. 2.1.

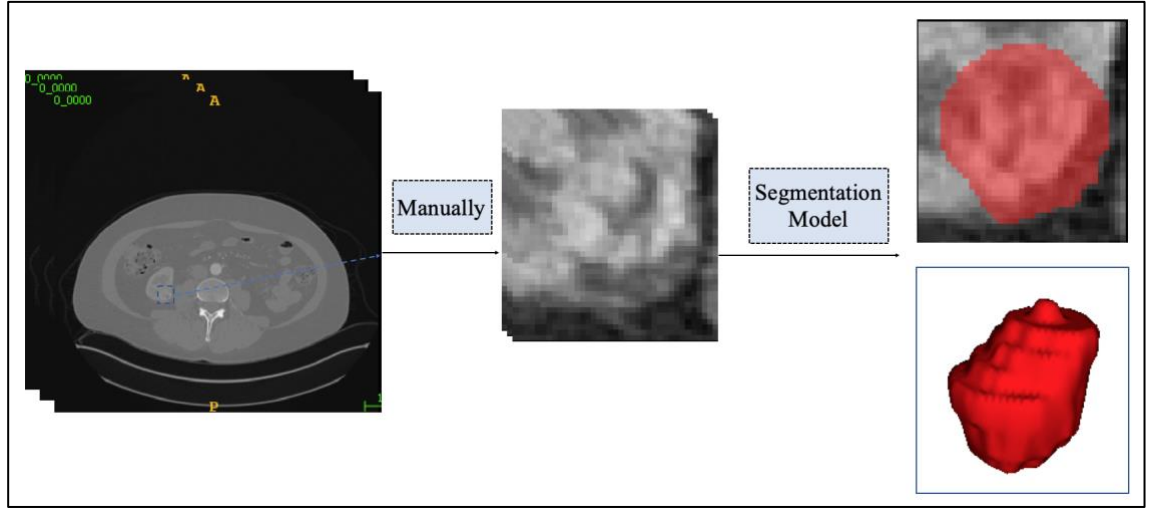


Fig. 2.1: The scheme of semi-automatic segmentation. The dotted blue box is manually labelled by experts and the right column are the segmented slice and its 3D reconstruction view. The red area shows the tumor.

U-Net [35] has generally applied in various medical image segmentation tasks and denotes a kind of neural network architecture proposed based on FCN [34], showing good performance as well as robustness especially in medical image applications. It outperforms FCN through constructing short connection between encoder and decoder and hence providing more accurate localization information to the up-sampling process.

Therefore, in this chapter, based on the ROIs of tumor, we proposed to employ 3D U-Net as the segmentation model to fully utilize the 3D characteristics in CT images and designed a simple but powerful network to segment tumors from the manually extracted ROIs. It takes advantage of 3D convolution kernels to extract 3D features and perform prediction for voxels. Of note, 3D convolution usually has better performance than generic 2D convolution because it has broader receptive field along 3D axes and hence has smoother prediction results.

2.2 Segmentation based on 3D U-Net

2.2.1 3D U-Net Architecture

Because the input of semi-automatic segmentation network are ROIs extracted from original images, their volume sizes are much decreased. In order to avoid severe distortion of images after multi convolution operations, we designed a rather shallow network and it looks very simple but with satisfying performance. The basic architecture is shown as Fig. 2.2.

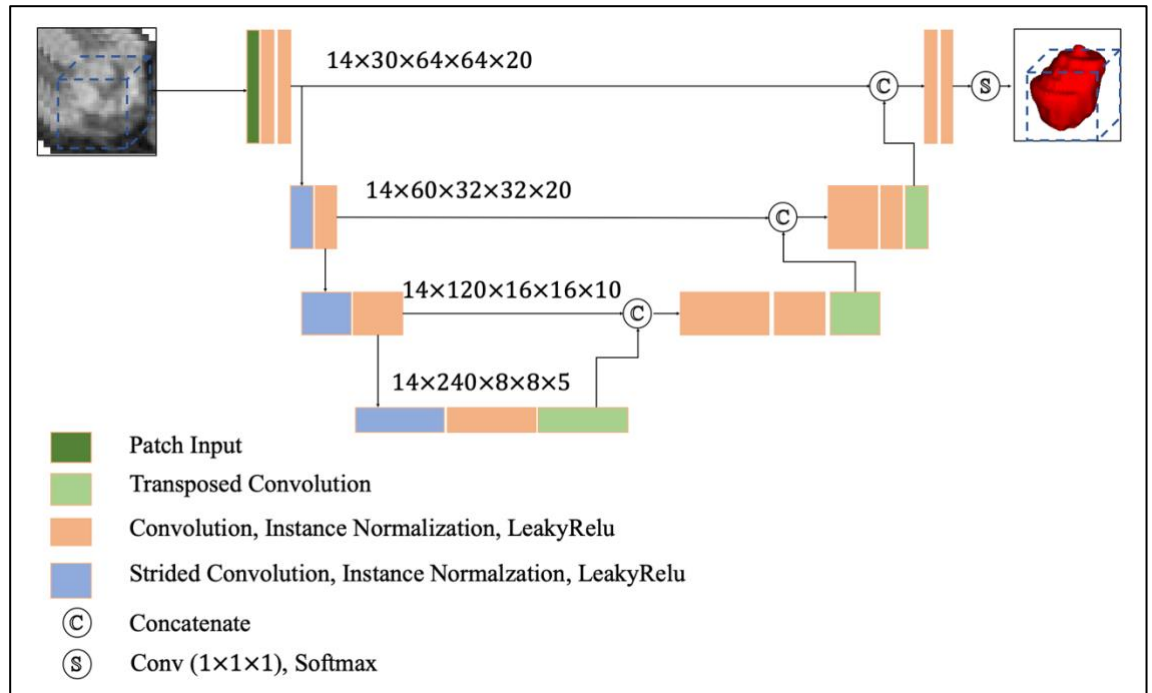


Fig. 2.2: The architecture of our proposed 3D U-Net.

Since deep learning networks only process the size-fixed input, we adopted a patch-based strategy to train and test images with various sizes. The input size of our network is set as $64 \times 64 \times 20$ slightly smaller than the average image size, where 64 means the

pixels in transverse plane and 20 means the pixels in the axis plane. Out of our experience, we replaced common pooling operation by stride convolution to down sample images or extract features.

With the process of convolution operation, the data size would be decreased while more high-level distinguishable semantic information was extracted to determine the class of each pixel. However, if the data size is down sampled too much, it might lead to severe distortion of pixel positions and behave as obstacles for up sampling. Therefore, the maximum depth was designed as small as four including the bottleneck.

2.2.2 Training Settings

Of note, even though 3D networks usually have stronger feature representation capability than 2D networks, the anisotropy existed in training data would diminish their advantages severely. Based on this prior knowledge, we first resampled all data into same voxel spacing and then input them into the network.

As medical data are always scarce and valuable and at the same time deep learning networks are data thirsty to avoid overfitting, we employed a large variety of on-the-fly data augmentation techniques to handle such problem, including mirroring, elastic deforming, scaling, gamma correction and rotation. The final input patch was randomly selected from these augmented and original data.

Cross entropy (CE) and Dice coefficient (Dice) are commonly used loss functions in segmentation networks. Their definitions are shown as equation (1) and (2).

$$CE = -\sum_{classes} y_{true} \log(y_{pred}) \quad (1)$$

where y_{true} is the ground truth, while y_{pred} is the probability of prediction.

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

where $|A \cap B|$ represents the common elements between sets A and B, and $|A|$ represents the number of elements in the set A (and likewise for the set B).

CE is a pixel-wise loss as it computes the entropy of prediction probability and ground truth based on each pixel. However, such characteristic might lead to the network learning dominated by the class with more pixels and hence neglecting the class with less pixels. This situation is more notable in medical images since they are always class imbalanced.

In contrast, Dice is especially useful in tiny target segmentation tasks, because it could reflect the performance of segmentation regardless of the targets' relative size. In order to formulate a loss function which can be minimized, the Dice loss is usually used as $-Dice$. Moreover, we could directly use the predicted probabilities instead of thresholding and being converted into a binary mask, i.e. Soft Dice, which is more sensitive to the prediction confidence and able to instruct the network to adjust its parameters opportunely.

Therefore, we finally adopted Soft Dice + CE as our loss function to train the 3D U-Net, shown as equation (3).

$$Loss = -Soft\ Dice + CE \quad (3)$$

In addition, we used Adam as our optimizer. The learning rate was initialized to be 3×10^{-4} , and would be dropped by the factor of 0.2 if the training loss was no more improved in 30 epochs.

2.2.3 Inference

For testing images, we first resampled them into the same voxel spacing as training data. Since we adopted a patch-based training strategy, testing images should be also

conducted in a patch-based way and set the overlap between patches as half of the patch size. We also weighed more on the central predictions than the borders through allocating a Gaussian distribution weights on the different predictions. This was inspired by the prior knowledge that convolution layers perform better in the central of reception field. Moreover, we employed test-time data augmentation in an 8-fold way by mirroring all axes, which would improve the precision and robustness of the predictions.

2.3 Experiment and Results

2.3.1 Data Preparation

We obtained 210 patient CT scans from KiTS19 challenge dataset and no ethical consent and informal are needed for this retrospective study since they are publicly accessed and free available for the scientific study. We first resampled them into the same voxel spacing, and the dataset properties are shown in Table 2.1.

Table 2.1: Dataset Properties

Property	Value
Number of Patients	210
Modality	CT (late-arterial phase)
Number for Training	168
Number for Test	42
Affine	Array ([[0. , 0. , -0.7816, 0.], [0. , -0.7816, 0. , 0.], [-3.0 , 0. , 0. , 0.], [0. , 0. , 0. , 1.]])

2.3.2 Results and Analysis

Running on one Nvidia Tesla GPU (32GB) and employing Pytorch framework, the training of our network took five hours and more than 600 epochs. The loss variation process is shown as Fig. 2.3.

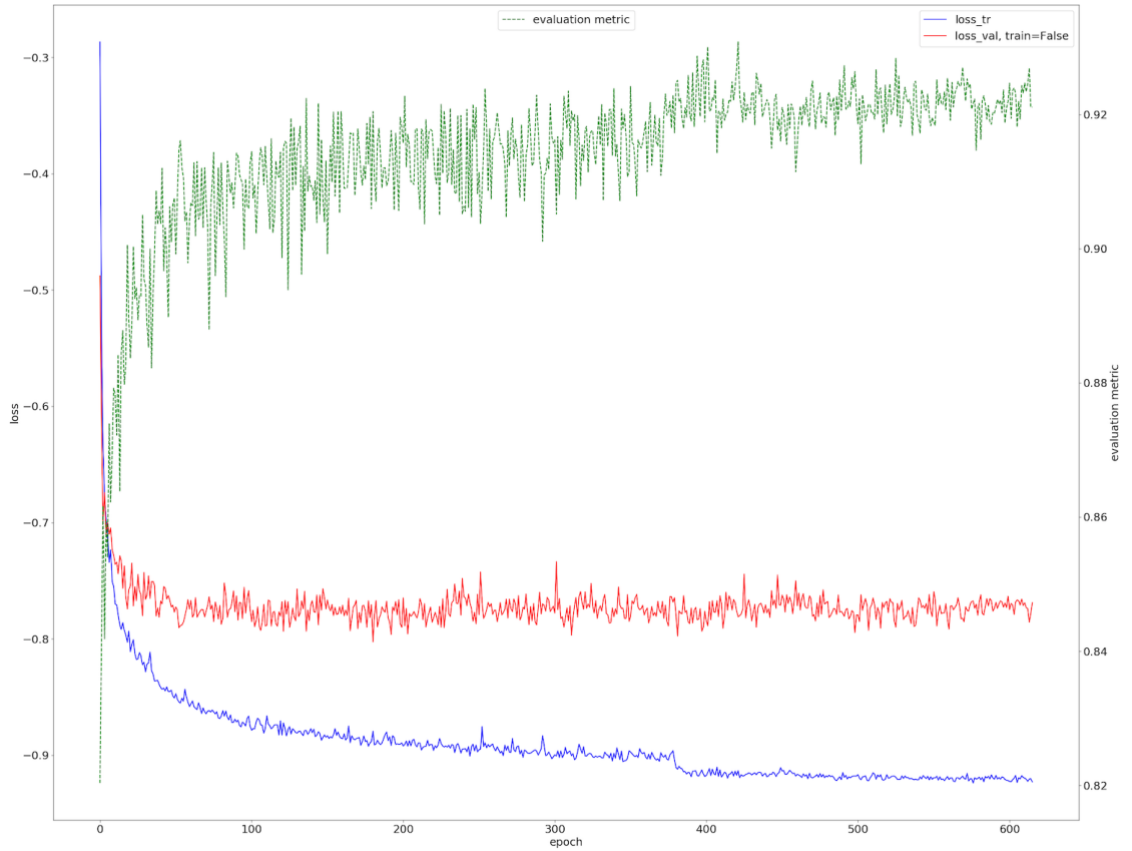


Fig. 2.3: The loss variation process. The blue curve is the training loss; the red curve is the validation set loss; the green curve is the evaluation metric used to find the best checkpoint to save.

Some segmented examples with their ground truth are visualized in Fig. 2.4. It is obvious that the predictions are rather close to the labels, but some details still differ.

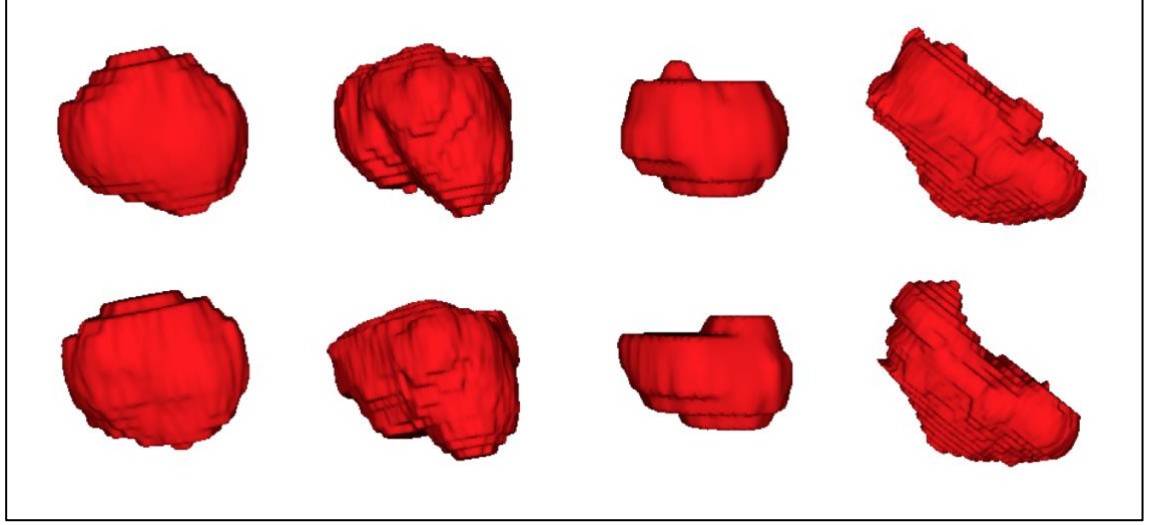


Fig. 2.4: Visualization of segmentation results. The top row is the prediction, while the bottom row is the ground truth.

In order to quantify the performance of the segmentation network, we used Dice Coefficient as our main evaluation metric, and also employed other four metrics as supplement to fully evaluate the segmentation performance, i.e. Jaccard, pixel accuracy, precision, and recall.

For clarity, we rewrote the definition of Dice Coefficient as Equation (4):

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (4)$$

where TP, FP, FN and TN (used later) denotes the true positive, false positive, false negative and true negative.

Jaccard index between two sets is defined as the intersection between them divided by their union, as Equation (5):

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (5)$$

Pixel accuracy is an alternative metric and is simply report the percent of pixels in the images which were correctly classified, as Equation (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision effectively describes the purity of the positive classifications related to the ground truth, defined as Equation (7).

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall describes the completeness of our positive predictions relative to the ground truth. It means computing how many positive pixels in the ground truth have been captured by our prediction, defined as Equation (8).

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

After the model was well trained, we tested it using the independent 42 patient scans, and the average of corresponding metrics are listed in Table 2.2.

Table 2.2: The segmentation performance (average values of all test images)

Metric	Value
Dice	0.8731
Jaccard	0.7878
Accuracy	0.9522
Precision	0.8432
Recall	0.9180

In order to describe the detailed performance and show the distribution of each patient, we plotted the boxplot in Fig. 2.5. The performances on these five metrics are all competitive and pretty close to the expert labels. It should be noted, even being re-annotated by experts, the results still cannot be completely the same. The agreement for repeating annotation Dice is 0.923 in terms of the tumor segmentation [39], while our semi-automatic segmentation model achieved 0.873.

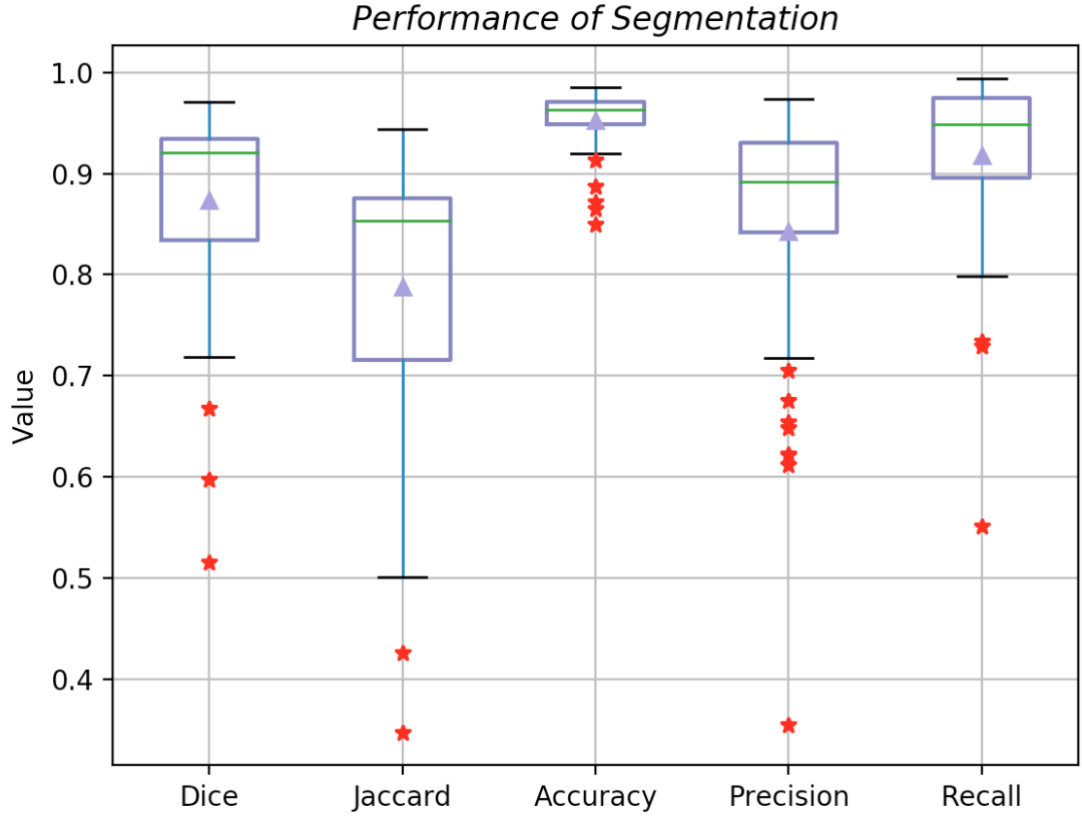


Fig. 2.5: Boxplot of the tumor segmentation. The purple triangle denotes the average value, while the green line denotes the median of all 42 testing patients. The red stars mean some outliers.

2.4 Summary

In this chapter, we proposed a semi-automatic segmentation method. It was constructed based on the basic but effective 3D U-Net architecture. According to the dataset size, the network was designed rather shallow, only four layers including the bottleneck. Also, we adopted a variety of data augmentation techniques in the training

process to avoid overfitting and employed test-time augmentation to improve the accuracy and robustness.

In the test, this simple but awesome network showed good performance, achieving 0.87 on Dice coefficient. It can be concluded that if applied in clinical practice, this model would reduce the tedious segmentation workload, because radiologists only need to label out the tumor area with this semi-automatic model's help.

Chapter 3

Automatic Segmentation of Kidney and Kidney Tumor in CT Images

3.1 Introduction

Even manual or semi-automatic segmentation could achieve accurate segmentation, they definitely consume radiologists time and energy in clinical practice since there are usually hundreds of slices for one patient. Therefore, the final goal of academia is to develop full-automatic segmentation methods with comparable performance with manual segmentation.

Actually, there have been various attempts to approximate the above object. Several approaches aiming to segment kidneys in CT or MRI images were proposed. Cuingnet et al. employed random forest and segmented kidneys in a two-step way [44]. Yang et al. [45] proposed a multi-atlas images-based method and also segment kidneys in a coarse-to-fine way. However, this traditional method using the prior shape as starting point can easily fail in the presence of large exophytic tumors. Also, a few research works targeted to kidney tumor segmentation. A level-set based method was proposed by Linguraru et

al. [46] to extract the renal tumors. Qian et al. [47] designed a CNN based architecture named Crossbar-Net to segment the kidney tumor.

There is also similar work as this chapter to segment the kidney and kidney tumor at the same time. For example, Yang et al. [48] proposed a two-stage method which first cropped ROIs of kidneys from the original CT images by multi-atlas-based approach, and then the ROIs were used to train a new 3D_FCN_PPM network.

However, these previous works are usually either multi stages based or just targeting for one class. In this chapter, aiming to build one end-to-end multi-task segmentation network trained from scratch and improve the performance further, we proposed multi-scale 3D U-Net. This method was then validated in the International Conference on Medical Image Computing and Computed-assisted Intervention (MICCAI2019) challenge KiTS19 public training dataset [39] and outperformed the classic 3D U-Net.

Furthermore, in order to compare our model with the state-of-the-art ones, we tested the proposed algorithm in the KiTS19 test dataset, and finally our method achieved the 7th place among 106 international teams with the comparable performance both in terms of kidney and tumor.

In section 3.2 we describe details of this architecture-simple but rather powerful model; the experiments and analysis are introduced in section 3.3; section 3.4 gives a brief summary for this chapter.

3.2 Multi-Scale Supervised 3D U-Net

As described above, there has existed various segmentation methods with spells and whistles covering kinds of organs and modalities. After comparison and test of typical approaches, their generalization was not guaranteed even though they might have

achieved good performance in specified datasets and organs. Moreover, most experiments shown in previous papers regarding the kidney and kidney tumor were always conducted in private datasets, therefore, their results can be hard to reproduce.

Inspired by Fabian [50], who used 3D U-Net with minor modification and achieved champions or top ranks in various medical images segmentation challenges, a multi-scale supervised 3D U-Net with other useful enhancement was proposed in this chapter. We also discarded some common tricks on architecture, such as residual block [51], dense block [52], attention mechanism [53], feature pyramid network [54] and feature recalibration [55]. From the point of our view, medical images are far less diverse than nature images, so they don't need too deep convolution layers or too much connections. The basic U-Net with only six layers is enough to represent or learn features to be used to classify.

Following such suggestion, we focused our work on training 3D U-Net better and utilizing the limited training dataset more effectively. Therefore, we proposed three strategies to enhance the basic 3D U-Net.

(1) As the final full resolution prediction is up-sampled from deeper low-resolution layers, it is very important to guarantee the accuracy of the prediction in deep layers. Hence, we designed a multi-scale supervised 3D U-Net to encourage the network to predict correctly not only in the last layer, but also in every resolution level, which would improve the performance in the final layer consequently.

(2) Noteworthily, the voxels of background in CT volumes are much more than that of kidneys and especially tumors. To mitigate the negative effect brought by imbalanced class data, we used the enhanced focal loss [56], exponential logarithmic loss [57], which would pay more attention on the difficult elements such as tumors.

(3) At last, with the prior knowledge of maximum two kidneys in one patient, our post processing would keep up to two maximum connected components of the kidney and remove the scattered kidney voxels or tumor not attached with kidneys.

In the next sections, several related details of the workflow were introduced respectively. In section 3.2.1, the preprocessing and augmentation of training data was described. Section 3.2.2 delineated the main architecture of our network. In section 3.2.3, the training procedure was given and in the last section 3.2.4 the methods of inference and post-processing were explained.

3.2.1 Preprocessing and Data Augmentation

Data quality is vital for training a well performing network. Usually, machine learning works under the assumption that the input data including both training data and testing data are independent and identically distributed (*i.i.d.*). Therefore, some preprocessing strategies should be conducted to guarantee this requirement.

CT images acquired by different protocols or different phases will produce huge difference of intensity distribution. Hence, the data we used below are all from the same late arterial (AP) phase. Even in such circumstance, there still are some impacts on the acquired data due to the patients or the devices, such as the prohibited materials on patients or different settings of devices.

To remove the abnormal intensity values, which might be from some metal things, we clipped the intensity values of CT images into their 0.5 and 99.5th percentile. Then by convention, we normalized the data with the mean and standard deviation of foreground pixels following the typical weight initialization method. It should be emphasized that the anisotropy of 3D data would destroy the advantage of 3D convolution

since it cannot learn a unified representation for different voxel space data with the same receptive field. Thus, we resampled the data into a same voxel space if they are not.

It is always tedious to manually annotate medical images, so the labeled datasets volume is usually limited. But deep learning networks are data thirsty, hence we employed strong data augmentation to avoid the model being overfitted, including random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring. The effect of adopted augmentation techniques is shown in Fig. 3.1.

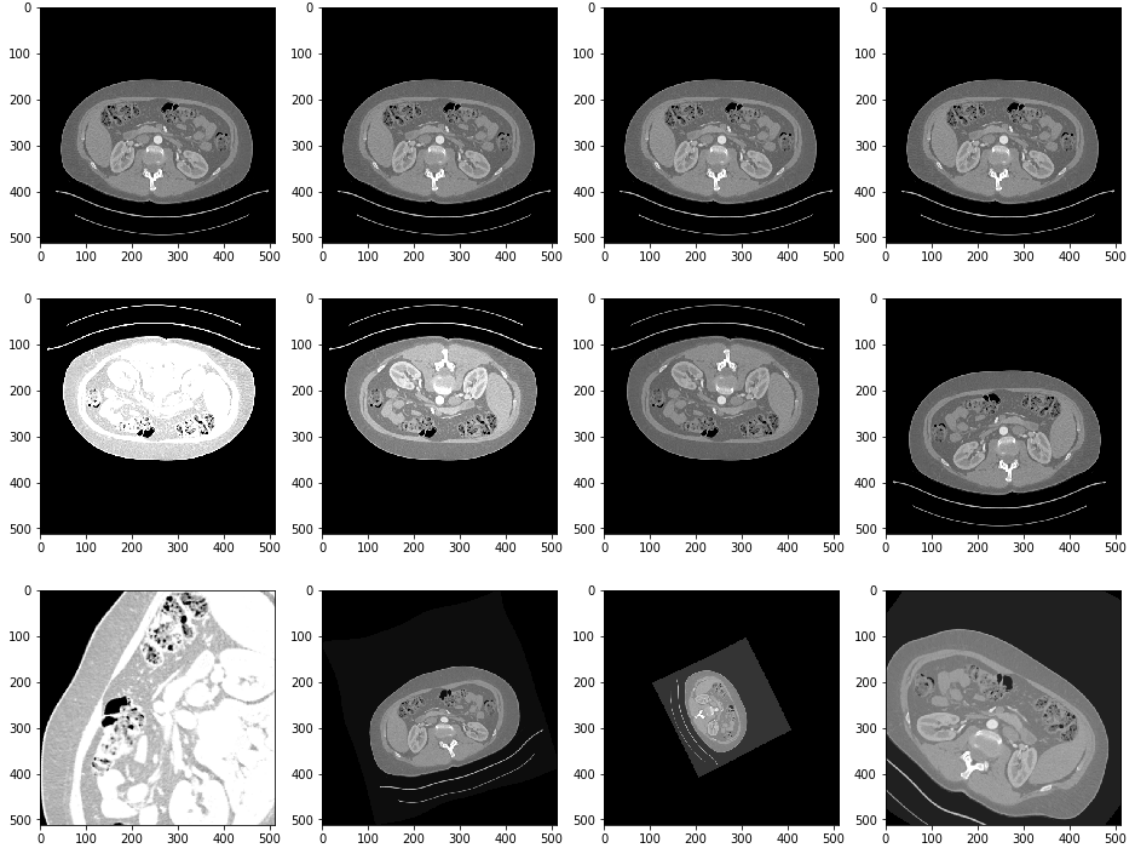


Fig. 3.1: The effect of augmentation techniques used in our method. Although we employed the same augmentation in 3D data, here to be best viewed we just show 2D images. The first row is the original images; the second row shows the effect of contrast

augmentation and mirror; the third row shows elastic deforming, scaling, gamma correction and rotation.

3.2.2 Architecture

U-Net is a classical encoder-decoder segmentation network and has drawn a lot of attention in recent years. The encoder pathway is similar with the typical classification network to extract more high-level semantic feature layer by layer. Then the decoder pathway recovers the localization for every voxel and utilizes the feature information to classify it. To employ the position information embedded in encoder, direct connection is constructed between the layers in the same stage.

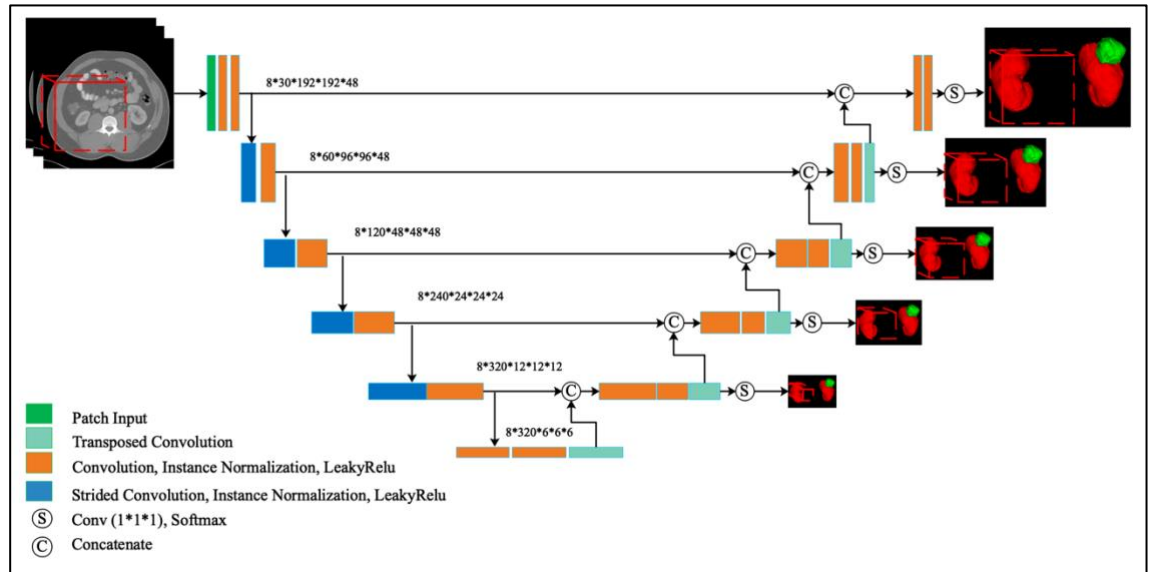


Fig. 3.2: The scheme of our multi-scale supervised 3D U-Net (best viewed in color). The actual architecture is 3D, but for simplicity we use 2D here.

We designed our network based on 3D U-Net, shown as Fig. 3.2. It is widely known that the pooling operation including max pooling and average pooling would fuse the

position information and thus impact the up sampling to predict correct labels. Therefore, we adopted the strided convolution instead of pooling operation to implement down sampling but keep as more original position information as possible. Also, we replaced the trilinear interpolation by the transposed convolution to enable adaptive up sampling. Normalization is usually between any two layers to obtain fixed input distribution, and has been proved beneficial for network training and performance. However, since the batch size we used was limited due to the GPU memory capacity, we hence employed instance normalization instead of batch normalization which is suitable for bigger batch size.

The spirit we followed when designing our network is to avoid the feature maps over sampled. Therefore, we set the architecture of our network only 6 layers including the bottleneck to meet the condition that the deepest feature map is not smaller than $8 \times 8 \times 8$. At the same time, to reduce the model volume, we set the basic feature number as 30.

In addition, we creatively proposed multi-scale supervision to enhance the performance. This multi-scale supervised network makes the prediction from different layers of decoder pathway, unlike classical 3D U-Net which only predicts from the last layer. These segmented outputs would be compared with corresponding resolution labels and then used to calculate the final loss function.

Such supervision format was inspired by the well-known deep supervision, but showed different implementation thought. Common deep supervision works just like feature pyramid network which up samples all the feature maps in different layers and add them together or perform concatenation to get the final output. From the point of our view, this just utilizes the deep features indirectly, because all the features are fused together and cannot take full advantage of deep layers.

However, the multi-scale supervision we proposed could employ the different layers directly by collect all the outputs of them. Thereby, the following comparison with

different resolution labels will aggregate the information of every layer into the final loss. It works in a similar way with human. When we human are labeling something in the computer, we usually first zoom out the target and give a coarse contour. Base on this result with low resolution, we then perform fine tuning in every part to obtain the final contour with full resolution. Overall, this mechanism encourages the network to predict correctly from the low-resolution feature maps which will be up-sampled to be full resolution feature maps and reasonable improvement of the network performance could be expected.

3.2.3 Training Procedure

Limited by the GPU memory, we chose the patch size as $192 \times 192 \times 48$ and set the batch size as eight using Data Parallel in two GPUs (Tesla, 32GB). The patch was random sampled and we refer to an epoch as 250 iterations. We used Adam as our optimizer. The learning rate was initialized to be 3×10^{-4} , and would be dropped by the factor of 0.2 if the training loss was no more improved in 30 epochs.

For the kidney and tumor segmentation in CT images, the sample of the background is far more than kidney and tumor voxels, which would reduce the learning opportunities for tumor. Also, the tumor is more difficult to classify due to its morphological heterogeneity. To handle such generic problem, we employed an exponential logarithmic loss. This loss emphasizes the effect of difficult samples and gives them more weight by making the loss nonlinear. The theory can be shown in Fig. 3.3.

We could find that the linear loss $f(x) = 1 - x$, will reduce the loss value rapidly when the prediction becomes better and hence obstacle the performance being improved further. However, when employing a suitable nonlinear loss, the loss will increase dramatically if the prediction accuracy is poor. Note that inappropriate exponential index

would repress the continuous learning capacity for good prediction. As the orange line shows, if we adopt exponential index as 2.0 the loss value will decline drastically when the prediction probability is higher than 0.5. Therefore, we finally chose $f(x) = -\ln(x)^{0.3}$ as our loss function, which not only encourages the network to pay more attention on obscure pixels, but also presents higher loss even though the prediction is rather accurate. Only when the prediction accuracy is very high, will the loss decrease dramatically to zero.

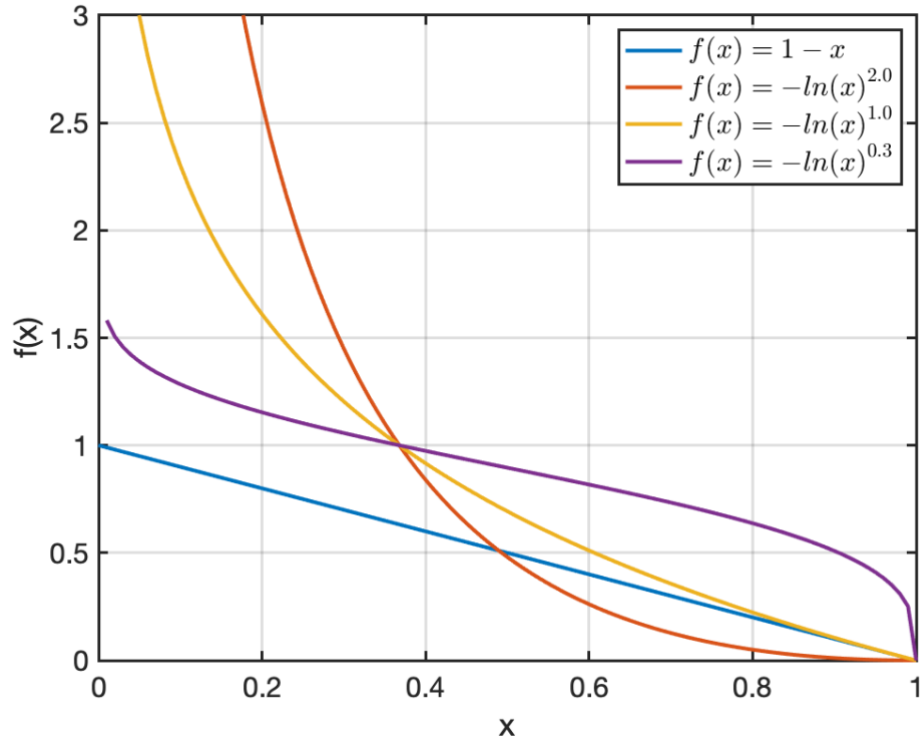


Fig. 3.3. Loss functions with different nonlinearities.

We combined the Soft Dice transformed by our nonlinearities function and Cross Entropy to train our model. At the same time, we attributed different weights for the background, kidney and tumor manually. The final format of our loss can be summarized as follows:

$$Loss = DICE + CE \quad (9)$$

$$DICE = (-\log^{Dice_{kidney}})^{0.3} \times 0.4 + (-\log^{Dice_{tumor}})^{0.3} \times 0.6 \quad (10)$$

$$CE = 0.28 \times CE_{bg} + 0.28 \times CE_{kidney} + 0.44 \times CE_{tumor} \quad (11)$$

Specifically, we set the weight of kidney and tumor as 0.4 and 0.6 respectively in terms of DICE. As the part of CE take consideration of both background and foreground, we adopted the same weight of background and kidney as 0.28, and gave 0.44 to tumor.

3.2.4 Inference and Post Processing

Due to the fact that our training is conducted by patch, all inference should also be done patch-based. Since the accuracy of the border of patches is decreased compared with the patch center, we adopted the overlap prediction and weighed more on the center values when aggregating predictions across patches. Patches were chosen to overlap by $\frac{1}{2}$ patch size and the weights of voxels in patches were given following Gaussian distribution. Furthermore, we employed test-time data augmentation by mirroring all patches along all valid axes to aggregate more predictions for voxels. Finally, for every voxel there were multi predictions aggregated to decide its optimal value. In the center of patient data, up to 64 predictions from several overlaps and mirroring were aggregated up.

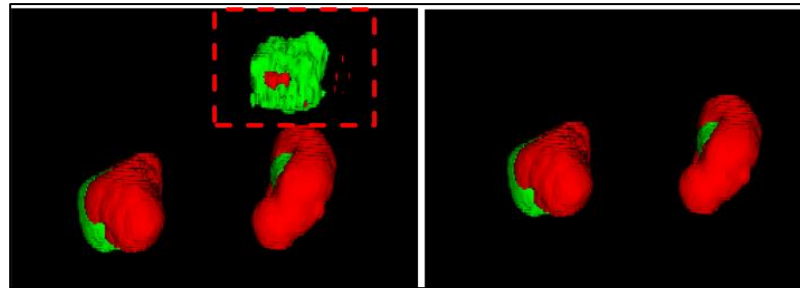


Fig. 3.4: The effect of our post processing method. The left image is before post processed, and the volume in the dotted box are some voxels obviously wrong. The right image is after post processed, where the extra voxels have been removed.

Some human common knowledge can be used to further improve the performance of original prediction. There are up to two kidneys in one person, and the tumor should be attached on or embedded in kidneys. Although it seems very simple, this information can be employed to act as post processing. Hence, we designed a connected-component based post processing method to remove the obvious wrong voxels.

As the segmentation of kidneys is always better and more credible, the method first keeps the top two biggest kidney components. If there is only one kidney, we should account that the patient has ever received nephrectomy. It treats other components as wrong voxels and removes them. At last all tumor components attached with the kidney should be kept while others should be removed. The effect of our post processing is shown in Fig. 3.4.

3.3 Experiments and Results

In order to validate our algorithm, KiTS19 data was employed to train and test our model. The following sections introduce details of experiment and analysis of the results respectively in section 3.3.1 and section 3.3.2.

3.3.1 Experiment Data and Set Up

There are 210 patients' volumetric scans published. These scans are all preoperative abdominal CT imaging in the late-arterial phase, with unambiguous definition of kidney tumor voxels in the ground truth images.

Scans of different patients have different properties like, voxel size along three plans, and affine. Due to different voxel spacings, i.e. affine, affect the learning of network a lot,

we used the interpolated dataset, which interpolate the original dataset and keep same affine. The statistic of properties is shown in Table 3.1.

Table 3.1: Dataset Properties

Property	Value
Number of Patients	210
Modality	CT (late-arterial phase)
Number for Training	168
Number for Test	42
Min Patient Size in Voxels	[434, 434, 69]
Max Patient Size in Voxels	[639, 639, 182]
Median Patient Size in Voxels	[523, 523, 116]
Affine	Array ([[0. , 0. , -0.7816, 0.], [0. , -0.7816, 0. , 0.], [-3.0. , 0. , 0. , 0.], [0. , 0. , 0. , 1.]])

We randomly selected 20% patients as the test dataset, among the left there were 20% scans set as validation dataset and the others were used to train out network.

The experiment was conducted based on Pytorch framework, and ran on our severer (two× Nvidia Tesla 32GB, one× Intel Xeon Gold 6130 CPU). Similar with chapter 2, we still use Dice as the main metric with other four accompanying metrics (Jaccard, precision, accuracy, and recall) to quantify our segmentation performance.

3.3.2 Results and Analysis

Training the network from scratch nearly consumed five days in the above conditions. The loss variation is shown in Fig. 3.5, from which it shows that the loss decreases stably with more epochs, meaning our network has good convergence.

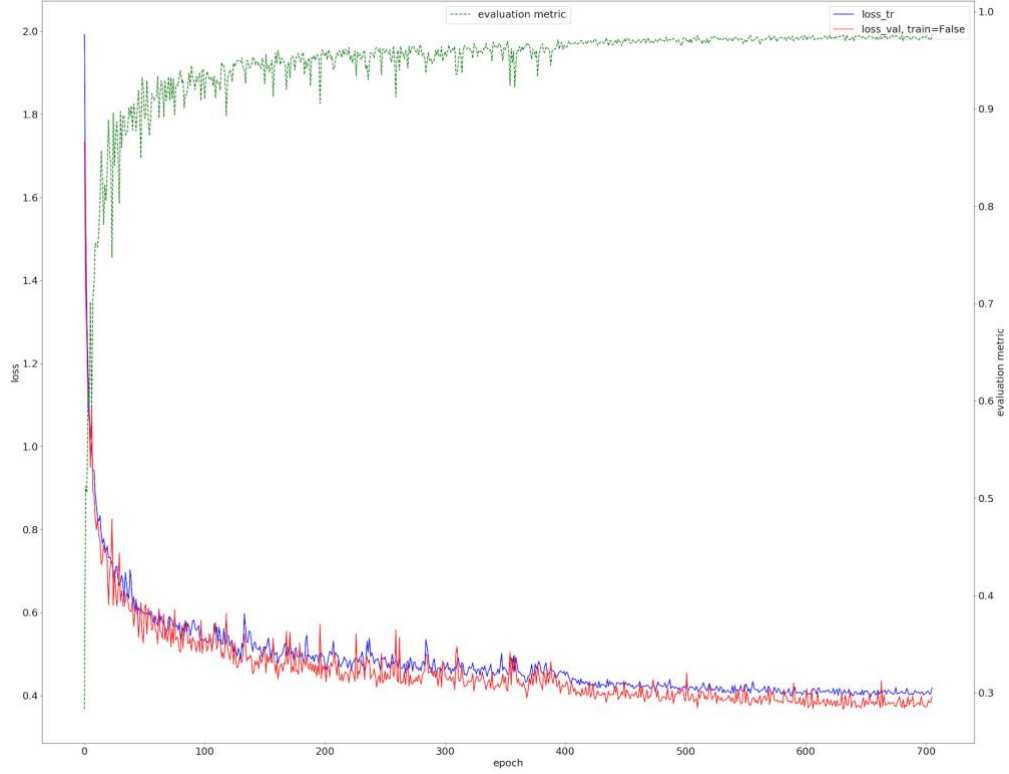


Fig. 3.5: The loss variation during our training. Red and blue lines mean validation and training loss respectively. The green line is a sliding validation loss metric to choose the best check point.

Some examples of visualized segmentation are shown in Fig. 3.6. We can see that kidneys are very close to the ground truth even when severe deformity exists in some cases. In addition, most tumors are also segmented very well except some too small ones which only occupy little percentage of the whole voxels.

We also evaluate the results using five quantified metrics as in chapter 2. Fig 3.7 and Fig 3.8 describe the boxplot of the kidney and tumor segmentation. We can find that the

Dice coefficient of kidney segmentation is up to 0.969, and it shows very good robustness with rather small variance. The other metrics for the kidney, including Jaccard, Accuracy, Precision and Recall, are also pretty good, all higher than 0.825. In terms of the tumor segmentation, the average Dice coefficient is 0.805, and median value a little higher than 0.90. However, it should be notable, several cases have very low values in terms of Dice and other metrics, and pulls down the average severely. The main reason is that the tumor in such cases are too small to detect. Therefore, even slight difference or deviation between prediction and ground truth would lead to very poor quantified values.

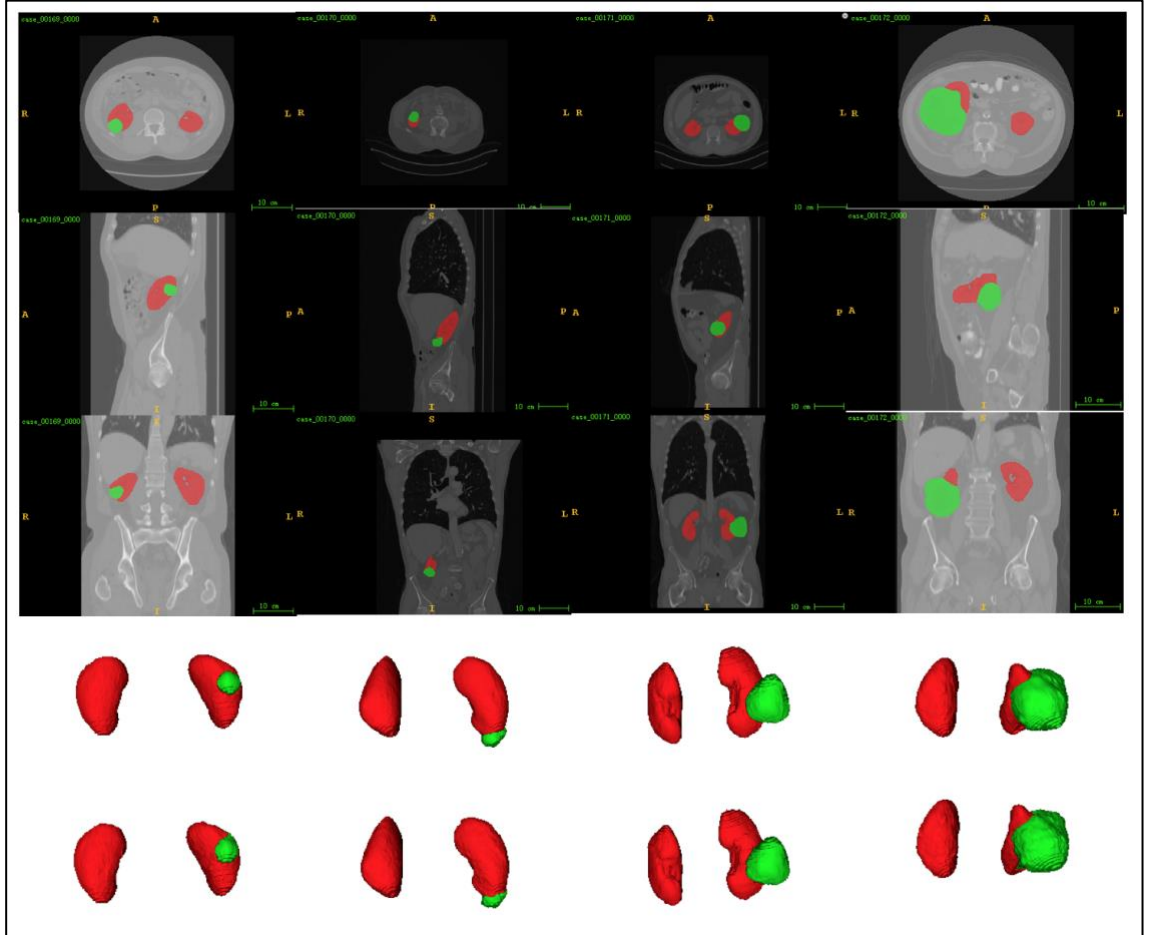


Fig. 3.6: Examples of segmentation results. Each column denotes one patient, and from top down, they are images in transverse plane, sagittal plane, coronal plane, 3D view of

predictions and the ground truth. The red mask means the kidney area, while the green mask means the tumor area.

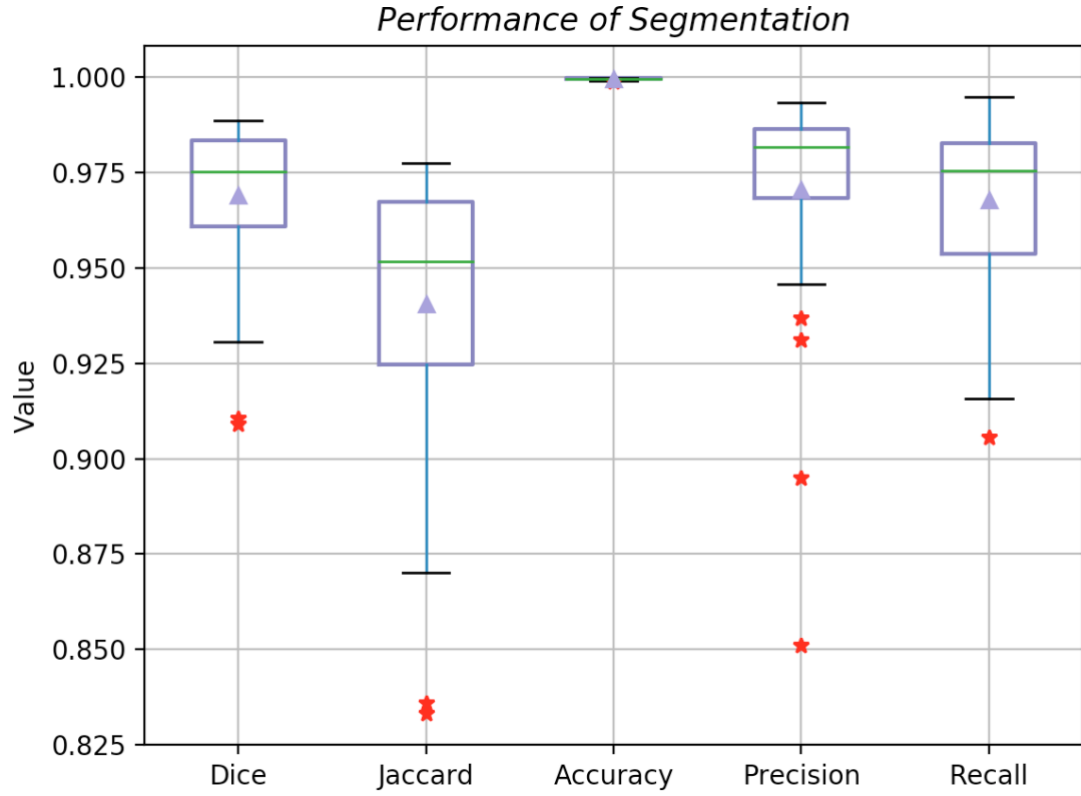


Fig. 3.7: Quantified metrics of the kidney segmentation.

Table 3.2: Comparison between the proposed method and classic 3D U-Net (kidney), the bold is higher.

Metric	Proposed	3D U-Net
Dice	0.9690	0.9621
Jaccard	0.9405	0.9297
Accuracy	0.9996	0.9995
Precision	0.9705	0.9610
Recall	0.9679	0.9652

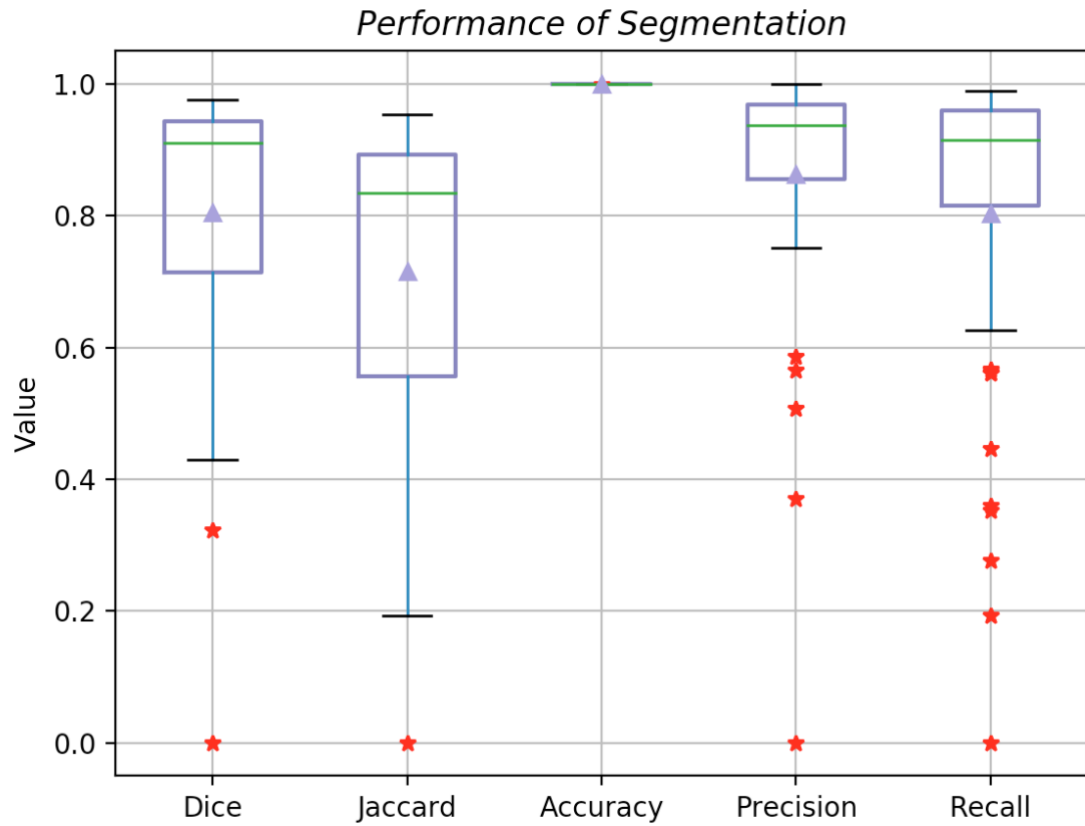


Fig. 3.8: Quantified metrics of the tumor segmentation.

Table 3.3: The comparison between the proposed method and classic 3D U-Net (tumor).

Metric	Proposed	3D U-Net
Dice	0.8053	0.7813
Jaccard	0.7160	0.6989
Accuracy	0.9998	0.9996
Precision	0.8627	0.8410
Recall	0.8029	0.8098

As our method is based on the 3D U-Net architecture, in order to quantify the performance of our enhancing strategies, we compared its results with the original 3D U-Net and listed the values in Table 3.2 and Table 3.3. It can be found that most metrics except Recall are improved after using our strategies. Of note, we achieved 0.02 growth on tumor Dice and this proved that the exponential logarithmic loss had worked as expected.

Additionally, in order to evaluate our method broadly, we compared its performance with other algorithm proposed in recent literatures even though they might use private dataset. From the comparison listed in Table 3.4, we can find that this multi-scale supervised 3D U-Net outperformed other two recently proposed algorithms.

Table 3.4: Dice of the proposed method and other algorithms used in recent literatures.

Method	Kidney	Tumor
2D_PSPNET [58]	0.902	0.638
3D_FCN_PPM [48]	0.927	0.802
Proposed	0.969	0.805

3.4 Summary

In this chapter, aiming to develop fully automatic method for segmentation of the kidney and tumor in CT images, we proposed a simple but powerful multi-scale supervised 3D U-Net. There are three useful strategies employed to enhance the classic 3D U-Net. First, we proposed multi-scale supervision, which encourages the decoder to predict correct segmentation results in every resolution level. Second, we absorbed the spirit of the recently proposed exponential logarithmic loss, and combined it with the Soft

Dice. At last, utilizing the prior knowledge, we designed the connected-component based post processing method to remove the obvious wrong voxels in the prediction results.

We tested our method in the public KiTS19 dataset, and it presented the pretty good segmentation results both in visualization and quantified metrics. Furthermore, our algorithm, with simple architecture, showed superiority and powerful capability when compared with the state-of-the-art ones.

Chapter 4

VHL Gene Mutation Prediction of ccRCC in CT Images

4.1 Introduction

There have existed a few researches to explore the relationship between CT image features of ccRCC and underlying gene mutation status **Error! Reference source not found.**, [60]. However, they usually employed manual features, such as tumor margin and calcification, calculated or quantified by experts. Expertise can well find the metaphysical features related to genomics, but this method relies on experienced radiologists which brings interobserver differences and consumes too much energy and time. In addition, it might loss some valuable information embedded in images but unseen by naked eyes.

Recently, some new presented work used high-throughput radiomic features to uncover genomic characteristics. Typical radiomic features used are shape, intensity, texture and wavelet, regardless of the targeted diseases or genomic status. Since these algorithm-based features produce high-dimension data, researchers always reduce the

dimensions or select some most distinct features from them to perform the next prediction. As a consequence, the final features used for different targets still differentiate even though they are from the same feature cluster.

Feature reproducibility is important for building a reliable prediction model. Burak et al. [61] presented that radiomic features extracted from manually segmented images generally have poor reproducibility despite only marginal difference between two segmentations. In order to avoid the possible uncertainties in the whole radiomics workflow, which may bring the impact of the reproducibility, here we developed fully algorithm-based automatic prediction models of VHL gene mutation in ccRCC.

In this chapter, we first employed the typical radiomic features extracted from CT images segmented by our multi-scale supervised 3D U-Net, then we adopted a sparse representation based method to select the most discriminating features and developed a machine learning based model to predict VHL gene mutation status.

4.2 Radiomic Features Based Prediction Model

Following the typical radiomic workflow, we first extracted quantified high-throughput features from segmented images. Since there are a lot of redundant features and this will diminish the performance of classifiers, the feature selection was performed to reduce the dimension of features, creating a reduced feature sets balanced between the predictive power and non-redundancy. At last, a classifier was employed to make the prediction. The basic scheme is shown in Fig. 4.1 [62].

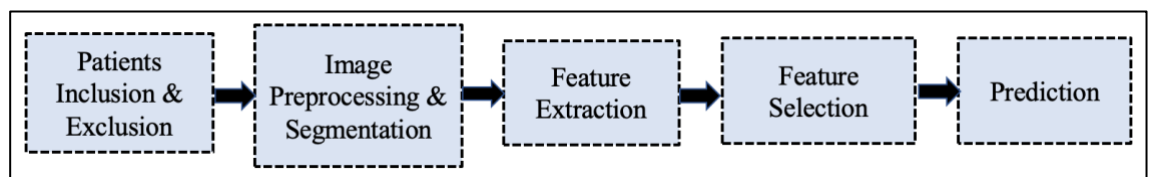


Fig. 4.1: The over-simplified workflow of the prediction model.

4.2.1 Introduction of Radiomic feature Sets

In this section, we introduced the feature sets used in our method. In total, we extracted 460 features from each slice, which can be categorized into four different classes, including 16 intensity features, 15 shape or morphology features, 73 texture features and 356 wavelet features [63]. They would be introduced one by one.

(1) Morphological Features or Shape Features (total number, 15): usually these are used to represent the geometric characteristics of the tumor, shown in Table 4.1.

Table 4.1: Morphological features

Cohort 1	Cohort 2
Compactness	Length of width ratio
Roundness	Standard deviation (Std) of normalized radius
Solidity	Elliptic-normalized angle
Extent	Diameter of equivalent circle
Spiculation	Moment difference
Convexity	Extreme point number
Edge roughness	Elliptic-normalized eccentricity
Entropy of normalized radius histogram	

(2) Intensity Features (16): these features indicate the intensity distribution and histogram distribution of the tumor area in the original image, details shown in Table 4.2.

Table 4.2: Intensity features

Cohort 1	Cohort 2
T-energy	H-mean
T-entropy	H-variance
T-kurtosis	H-skewness
T-mean absolute deviation	H-kurtosis
T-mean	T-skewness
T-median	T-Std
T-range	T-uniformity
T-root mean square	T-variance

(3) Texture Features (73): they imply the underlying change of intensity which are usually invisible. We attributed six kinds of texture features to represent tumor's corresponding properties.

a. Texture features based on contour (13): these features reveal the differences between the intensity inside and outside of the tumor, listed in Table 4.3.

Table 4.3: Contour based texture features

Cohort 1	Cohort 2
Local windows mean	Variance of inside region
Std of annular region	Std of correlation coefficient contrast
Std of inside region	Acutance 1
Inside region signal noise ratio (SNR)	Acutance 2
Deviation ratio	Annular region SNR
Variance of annular region	Mean of correlation coefficient contrast
Relative brightness	

b. Texture features based on ROI (6): these features present the intensity distribution in the whole ROI, shown in Table 4.4.

Table 4.4: ROI based texture features

Cohort 1	Cohort 2
ROI contrast mean	ROI covariance Std
ROI contrast Std	ROI dissimilarity mean
ROI covariance mean	ROI dissimilarity Std

c. Gray-level co-occurrence matrix (GLCM) texture features (23): these features depict the spatial linear relationship between the frequency of two intensities, listed in Table 4.5.

Table 4.5: GLCM texture features

Cohort 1	Cohort 2
Correlation 1	Contrast
Correlation 2	Sum average
Run percentage	Sum variance
Autocorrelation	Sum entropy
Low gray-level run emphasis	Difference variance
Cluster Prominence	Difference entropy
Cluster Shade	IMC 1
Dissimilarity	IMC 2
Energy	Inverse difference
Entropy	Inverse difference normalized
Homogeneity 1	Homogeneity 2
Maximum probability	

d. Gray-level run-length matrix (GLRLM) texture features (13): these features are proposed in order to characterize the roughness of the texture, by computing a set of consecutive collinear points with the same gray level in a given direction of the image, shown in Table 4.6.

Table 4.6: GLRLM texture features.

Cohort 1	Cohort 2
Gray-level variance	Long run high gray-level emphasis
Short run emphasis	Run-length variance
Long run emphasis	High gray-level run emphasis
Run-length nonuniformity	Low gray-level run emphasis
Long run low gray-level emphasis	Short run low gray-level emphasis
Gray-level nonuniformity	Short run high gray-level emphasis
Run percentage	

e. Gray-level size zone matrix (GLSZM) texture features (13): these features represent the uniformity of pixel clusters, shown in Table 4.7.

Table 4.7: GLSZM texture features.

Cohort 1	Cohort 2
Zone percentage	Large zone high gray-level emphasis
Zone-size variance	Gray-level nonuniformity
Large zone emphasis	Gray-level variance
Low gray-level zone emphasis	High gray-level zone emphasis
Large zone low gray-level emphasis	Small zone low gray-level emphasis
Zone-size nonuniformity	Small zone high gray-level emphasis
Small zone emphasis	

f. Neighborhood gray-tone difference matrix (NGTDM) texture features (5): they are utilized to describe the intensity difference of pixels and their neighborhoods, details are shown in Table 4.8.

Table 4.8: NGTDM texture features.

Cohort 1	Cohort 2
Busyness	Contrast
Complexity	Coarseness
Strength	

(4) Wavelet Features (356): these features are extracted from images applied wavelet transform. As a common image-processing technique, wavelet transformation with high or low frequency filters is always employed to decompose images in order to obtain possible hidden information from the raw images. After such process, more information which cannot be directly represented by intensity or texture features, could be extracted. Here we employed four kinds of wavelet transformations to decompose the original images, including Low/ Low pass (LL), Low/ High pass, (LH), High/ Low pass (HL), and High/ High pass (HH). After wavelet transformation, we got four components and computed their corresponding intensity and texture features. There were 89 features obtained for each component, and thereby totally 356 wavelet features were acquired.

Among above 460 features, shape features are computed based on the contour of the tumor, hence they are easily affected by the accuracy of segmentation. While the ROI based texture features are calculated based on the ROI, other intensity and texture features are acquired through the whole tumor area, and so are wavelet features.

4.2.2 Features Selection

Radiomic approaches generally lead to high-dimension features, producing a large number of features to be dealt with. However, they are usually redundant and correlated. Directly employing these original features would consume too much computing and time. For example, there are a lot of similar features remaining in both the VHL images and images from patients without VHL gene mutation. In addition, it is difficult to train classifiers for high-dimension data, since most of the correlated or similar features bring noises to the model and make it easily overfitting.

Therefore, it is a common practice to bring the high-dimensionality to low-dimensionality in order to optimize the classifier performance. There existed different approaches to achieve dimension reduction, such as feature reproducibility analysis [61], collinearity analysis [64], and algorithm-based feature selection. In the next paragraphs we would simply introduce these general feature selection methods and propose to employ a sparse representation-based feature selection approach.

(1) Feature reproducibility analysis. Some radiomic features are sensitive to variabilities such as segmentation accuracy [61], image acquisition protocol [65], and the parameter settings in the process of feature extraction. The goal of this reproducibility analysis is to reduce the dimension by excluding features with the relatively poor reproducibility. One common statistical tool for this analysis is the intra-class correlation coefficient (ICC) [66].

(2) Collinearity analysis. A large number of radiomic features contain the similar information and the item to represent it is called the strength of collinearity [67]. Thereby, the collinearity analysis is proposed as a plausible way of dimension reduction. Pearson's correlation coefficient and correlation-based algorithms can be used to determine the redundant features [68].

(3) Algorithm-based feature selection. As the most widely used dimension reduction technique, various algorithms have been proposed with different functions such as least

absolute shrinkage and selection operator (LASSO) [69], ReliefF [70], and Gini index [71].

Some existed approaches always analyze the correlation between feature and the label one by one, which might be difficult to achieve the global optimum. Thereby we proposed to employ a sparse representation-based method to filter the high-dimension features. It considers all the selected features and then determines which combination is the best for prediction, this method is more reasonable to obtain better performance.

Generally, the model of sparse representation model is [49]:

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_p \quad (12)$$

where \mathbf{y} indicates the target signal; \mathbf{D} is the dictionary shown as:

$$\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_i, \dots, \mathbf{d}_k] \quad (13)$$

where \mathbf{d}_i is called an atom in the sparse representation dictionary. α is the sparse representation coefficient (SRC) to approximate \mathbf{y} based on the fixed dictionary, and $\hat{\alpha}$ indicates the estimated value of α but with few nonzero elements, which is also the spirit of sparse representation. $\|\cdot\|_p$ is called the l_p norm and used to regularize the structure of α . λ is a weight employed to keep balanced due to the consideration for both the sparsity of matrix and the representation capacity. Usually we adopt orthogonal matching pursuit (OMP) method to solve (12).

In terms of our feature selection, if we suppose to extract Q features from each image, i.e. each image feature set is written as $\mathbf{f} \in \mathbf{R}^Q$, the sparse representation model can be formulated as:

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{l} - \mathbf{F}\alpha\|_2^2 + \lambda \|\alpha\|_0 \quad (14)$$

in which $\mathbf{l} \in \mathbf{R}^K$ represents the sample label set; K shows the amount of samples; $\mathbf{F} = [\mathbf{f}_1^T, \mathbf{f}_2^T, \dots, \mathbf{f}_i^T, \dots, \mathbf{f}_K^T] \in \mathbf{R}^{K \times Q}$ denotes the sample feature set; \mathbf{f}^T is the transpose of \mathbf{f} ; and λ is the regularization parameter.

Once the estimated SRC $\hat{\alpha}$ in equation (14) has been calculated, the value of each element in this vector means the importance of its corresponding feature. Then according to the absolute value, we could rank them in a descending order and thus remove the low-ranked features.

Of note, the number of samples K plays an important role in the process of feature selection. We should consider two kinds of situations due to its considerable influence on the selection results.

When $K \ll Q$, the matrix of dictionary can be regarded as underdetermined, hence it is easily to find one solution for the representation coefficient $\hat{\alpha}$. But it might be not the optimum and cannot effectively reflect the important of features.

Whereas, if $K > Q$, i.e. in the overdetermined case, it is not wise to directly solve $\hat{\alpha}$ because the features are insufficient to reflect the differences of so many images. Therefore, we propose a compromised method to obtain the coefficient vector. In this approach, we iteratively selected k random samples among the total K images to implement the sparse representation:

$$\hat{\alpha}^{(j)} = \arg \min_{\alpha} \|\mathbf{l}^{(j)} - \mathbf{F}^{(j)}\alpha\|_2^2 + \lambda \|\alpha\|_0 \quad (15)$$

in which $\mathbf{l}^{(j)}$ and $\mathbf{F}^{(j)}$ respectively denote label and the feature set of selected samples in the j th iteration. $\hat{\alpha}^{(j)}$ is the SRC computed through the OMP algorithm. Thereafter, we simply estimate the average of $\hat{\alpha}^{(j)}$ iteratively:

$$\alpha^{(j)} = \frac{1}{j} \sum_{i=1}^j \hat{\alpha}^{(i)} \quad (16)$$

Conventionally, we set a small positive constant ε as a sign to stop the iteration:

$$d(j) = \|\alpha^{(j)} - \alpha^{(j-1)}\|_2 < \varepsilon \quad (17)$$

under this condition, we suppose the final $\alpha^{(j)}$ has been optimized enough.

This iterative method provides an effective solution for the problems regarding of sample size. For each iteration, there are certain portion of samples randomly selected to

perform the sparse representation. Additionally, the average operation not only takes advantage of all samples but also improves the robustness of the feature selection.

4.3 Experiments and Analysis

4.3.1 Data Preparation

There were 267 patients' CT scans from The Cancer Genome Atlas-Kidney Renal Clear Cell Carcinoma (TCGA-KIRC) database[40], and their corresponding gene mutation ground truth were obtained from the *cBioPortal for Cancer Genomics* [41]. In this dataset, only 138 patients were diagnosed with gene mutation because of ccRCC and of note, some patients had multiple types of mutation. Statistically, 105 patients were marked with VHL gene mutation, 60 patients with PBRM1 and SETD2 respectively, and 20 patients had BAP1 mutation.

However, according to the experiment to be conducted, we had to carry out inclusion and exclusion of this database due to kinds of requirements. For example, nine patients had their CT scans after nephrectomy, and thus their images cannot be used. Some patients only had unenhanced CT scans while our segmentation model was trained using specified late arterial phase enhanced CT scans, so these patients should also be excluded because of their bad automatic segmentation results caused by the different acquisition protocols. Some scans were also excluded because their abdominal images did not cover the integrated kidney and tumors. Therefore, only 90 patients were reserved as the data used in our experiments. This process is shown in Fig. 4.2.

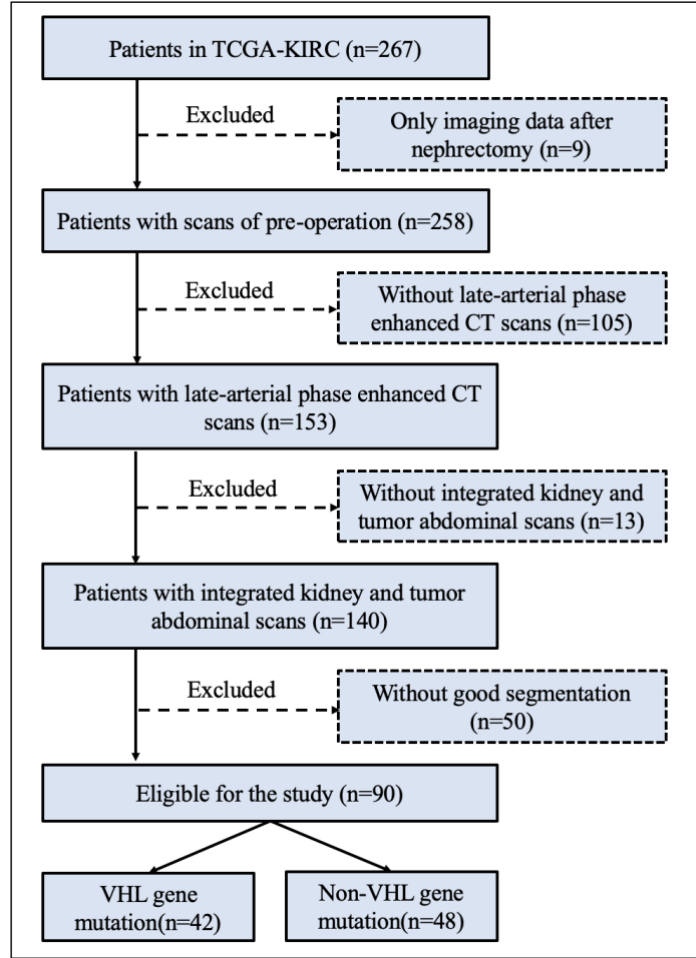


Fig. 4.2: The flowchart of the inclusion and exclusion criteria.

4.3.2 Image Segmentation

Tumor segmentation is a fundamental step in radiomic analysis, and all kidney and tumor segmentation were automatically implemented using the pre-trained multi-scale 3D U-Net proposed in Chapter 3. Since the original image scans after the process of inclusion and exclusion have various spacing, which are not consistent with the training data of segmentation model, we first resampled all images into the same voxel spacing as

$0.7816 \times 0.7816 \times 3.0 \text{ mm}^3$. As we guaranteed the same voxel spacing, the image size would vary and thus we still conducted 3D patch-based segmentation.

Examples of segmentation are shown in Fig. 4.3 and Fig. 4.4. We can find that even though the model was trained using KiTS19 dataset, it performed rather well on this TCGA-KIRC dataset as long as the testing images were obtained by the same protocols with training data. In Fig. 4.3, the first row is one patient scans without VHL gene mutation, while the second row shows one with VHL gene mutation scans. Fig. 4.4 shows their 3D view reconstructed through the ITK Snap.

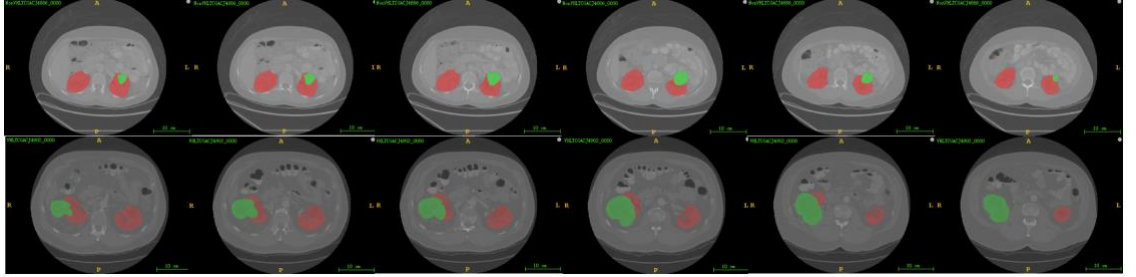


Fig. 4.3: The transverse planes of segmentation results. The top row is one patient without VHL gene mutation while the bottom row is with VHL gene mutation.

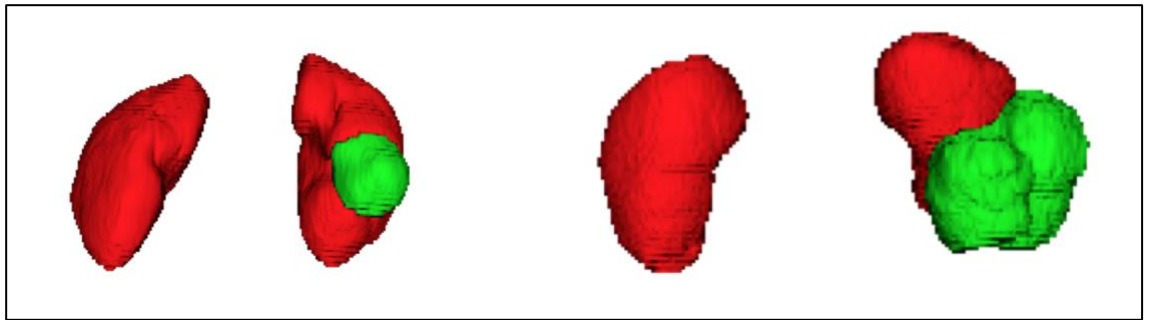


Fig. 4.4: The 3D reconstruction of two scans in Fig. 4.3.

After the kidney and tumor segmentation, along the contour of tumor, we extracted the smallest boundary box containing the whole tumor while excluding the top and bottom two axis slices since these marginal cross-sections are rarely proper for mutation detection

[73]. This small ROI is prepared as the source data of the feature extraction. This process is simplified and shown in Fig. 4.5.

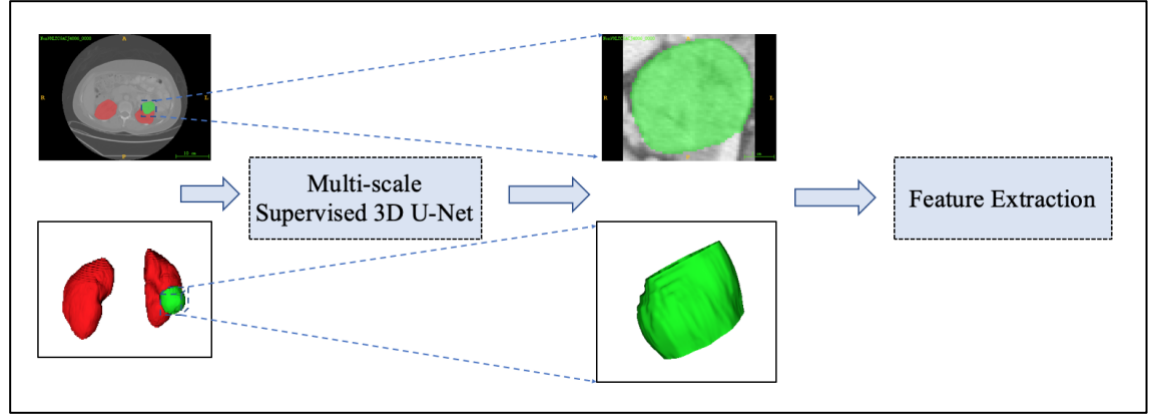


Fig. 4.5: The process before the feature extraction. The original images are segmented by pre-trained multi-scale supervised 3D U-Net, and then the tumor area is extracted in as a small boundary box, at the same time excluding the top and bottom two axis slices. Then the ROI is used to perform the next feature extraction.

4.3.3 Multi-Slice Based Experiment

From 90 patients selected in section 4.3.1, we randomly reserved 10 patients with VHL gene mutation and 10 patients without VHL gene mutation respectively as the independent test dataset. The partition is shown in Table 4.9.

Table 4.9: The number of patients used for the training and testing.

Gene Status	Training Number	Testing Number
With VHL	32	10
Without VHL	38	10

After obtaining tumor ROIs, the ccRCC contained in these ROIs were large enough because the marginal cross-sections had been removed, we employed the feature extraction described in section 4.2.1 and obtained a 460-dimension feature set for each slice. As these features were redundant and harmful for the classifier training, we adopted a sparse representation based feature selection. The iterative OMP solver was employed to implement the selection, and we ranked the discriminating features in Fig. 4.6. In order to balance the feature amount and their representation capability, we finally selected the top ranking 300 features to represent each slice.

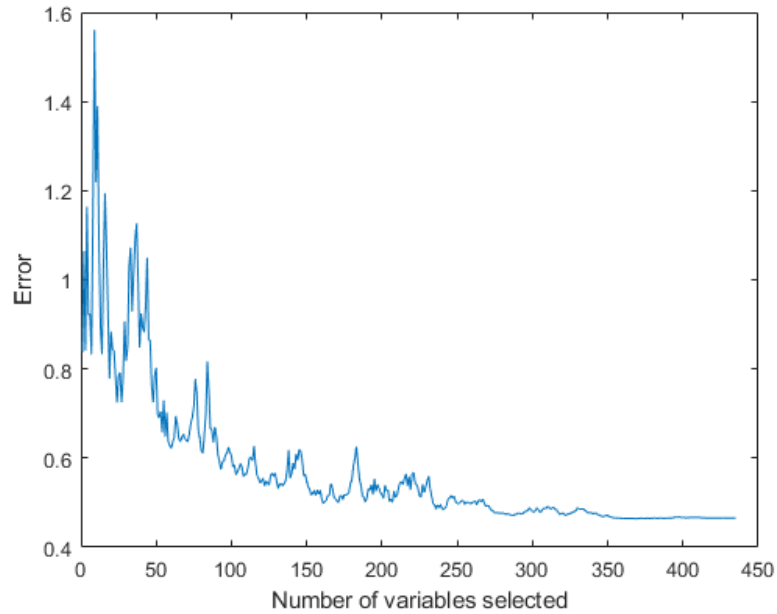


Fig. 4.6: The error variation when employing more features.

Then we employed 5-fold cross-validation for the training dataset in a slice-based way. Using an ensemble subspace KNN as the classifier [72], the receiver operating characteristic (ROC) curve and area under area (AUC) is shown in Fig. 4.7.

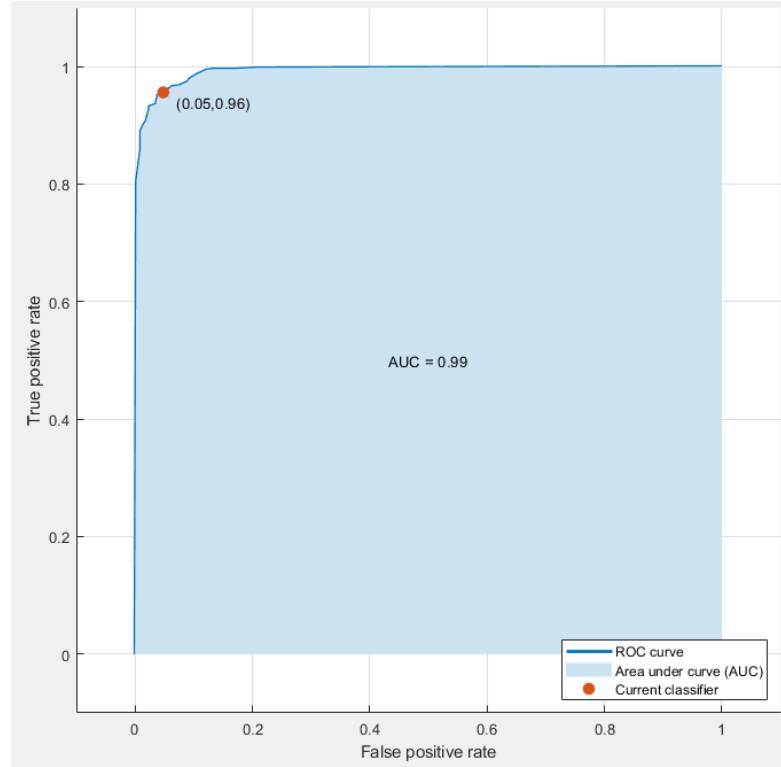


Fig. 4.7: The ROC curve of the validation dataset.

Since each patient contains multiple slices, we designed a multi-slice based prediction method for each patient. In such way, we first aggregated all single-slice based predictions of one patient, and then conducted simple voting for the final binary prediction of the patient. Finally, the accuracy of testing dataset achieved 0.8, and shown in Table 4.10.

Table 4.10: The prediction accuracy for the testing dataset.

Gene Status of Patients	Testing Number	Correct Prediction
With VHL	10	8
Without VHL	10	8

4.4 Summary

In this chapter, we proposed a machine learning based VHL gene mutation prediction model. First, the patient CT images were segmented by our pre-trained multi-scale supervised 3D U-Net, and extracted the tumor area as ROIs. Thereafter, 460-dimension radiomic features were extracted and furthermore selected by using a sparse representation based method. At last, after the slice-based predictions of each patient were obtained by an ensemble subspace KNN, we then employed the voting strategy to aggregate the predictions and determine the final result for the testing patient. The final accuracy achieved 0.8 in the independent test dataset.

Chapter 5

Conclusion and Discussion

5.1 Conclusion

Kidney tumor, especially ccRCC, is one of the most common lethal diseases. Based on recent research, genetic mutations are proved associated with advanced stage and grade, as well as patients' possible survival. Hence there is great prognostic value to get known of the genetic constitution of a patient's ccRCC.

The traditional method for identifying the mutations is through genome sequencing, which requires to sample the ccRCC biopsy. This detection method costs a lot and will bring great physical impact on the patient. Therefore, in this dissertation we proposed a radiomic method to invasively predict the underlying VHL gene mutation based on patients' CT images. With such purpose, there are three main work conducted as follows:

(1) We first proposed a 3D U-Net based semi-automatic segmentation method. This is a wise compromise between the accuracy and workload since kidney tumors with the rather small size and severe morphological heterogeneity are difficult to precisely detect and segment. This network only has four layers including the bottleneck, but shows the powerful performance. We employed a variety of data augmentation to cope with the

limitation of the small dataset. Also, the Soft Dice was used to replace the traditional Dice loss. At last, we adopted the test-time data augmentation to aggregate up to 64 predictions for one voxel, improving the accuracy and robustness.

(2) A multi-scale supervised 3D U-Net was proposed as an attempt to fully automatically segment the kidney and tumor in CT images. We proposed the multi-scale supervision to encourage the network to make correct predictions from deep layers since they play a fundamental role to guarantee the prediction accuracy of the next up-sampled layers. Furthermore, in order to stimulate the network to pay more attention on the tumor and voxels hard to recognize, we combined the exponential logarithmic loss with the Soft Dice and CE as the final loss function, and attributed higher weights on tumors. In addition, inspired by the prior knowledge that there are up to two kidneys in one patient, we designed a connected-component based post processing method to remove the scattered kidney voxels and tumor voxels not attached on kidneys.

(3) We proposed multi-slice aggregated prediction method to detect the underlying VHL gene mutation status of ccRCC in CT images. In this process, the original CT images should be segmented using the pre-trained multi-scale supervised 3D U-Net model, and then the ROIs were drawn out with the exclusion of marginal cross-sections. Thereafter, 460-dimension radiomic features were extracted and selected through a sparse representation based method. At last, we employed an ensemble subspace KNN as the classifier in a single-slice based way and the predictions were then aggregated using the voting strategy to determine the final result for the patient.

5.2 Future Work

Limited by the short time, there is still some work to be carried out.

Firstly, in terms of the segmentation method, even though it shows the pretty good performance in the testing dataset, we should note that some patients with a very small tumor are still very hard to detect and achieve satisfying Dice coefficient. The spirit of deformable convolution provides possibility for networks to focus on the tumor area accordingly by variable receptive field, which might help us to improve the method sensitivity to small tumors.

Furthermore, it is a generic problem to improve the generality of deep learning models. As to our pre-trained multi-scale supervised 3D U-Net, it is maybe meaningful to research how the CT images distribution from different devices and protocols should be processed to suit the unified segmentation model.

Lastly, as we only utilized 2D radiomic features and a voting strategy to aggregate them in the prediction model proposed in chapter 4, it is maybe beneficial to utilize the 3D features directly because it is more reasonable to treat the CT volume as a whole. In addition, some recent studies showed that CNN features extracted from segmentation model might contain minable information for the prediction. They are both practical thoughts and worthy to explore further.

References

- [1] Lindblad P. Epidemiology of renal cell carcinoma[J]. Scandinavian journal of surgery, 2004, 93(2): 88-96.
- [2] Cairns P. Renal cell carcinoma[J]. Cancer biomarkers, 2011, 9(1-6): 461-473.
- [3] Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012[J]. International journal of cancer, 2015, 136(5): 359-386.
- [4] Jemal A, Bray F, Center M M, et al. Global cancer statistics[J]. CA: a cancer journal for clinicians, 2011, 61(2): 69-90.
- [5] Fitzmaurice C, Dicker D, Pain A, et al. The global burden of cancer 2013[J]. JAMA oncology, 2015, 1(4): 505-527.
- [6] Lopez-Beltran A, Scarpelli M, Montironi R, et al. 2004 WHO classification of the renal tumors of the adults[J]. European urology, 2006, 49(5): 798-805.
- [7] Chen F, Zhang Y, Şenbabaoğlu Y, et al. Multilevel genomics-based taxonomy of renal cell carcinoma[J]. Cell reports, 2016, 14(10): 2476-2489.
- [8] Young J R, Margolis D, Sauk S, et al. Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT[J]. Radiology, 2013, 267(2): 444-453.
- [9] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of papillary renal-cell carcinoma[J]. New England Journal of Medicine, 2016, 374(2): 135-145.
- [10] Rini B I, Campbell S C, Escudier B. Renal cell carcinoma[J]. The lancet, 2009, 373(9669): 1119-1132.
- [11] Nabi S, Kessler E R, Bernard B, et al. Renal cell carcinoma: A review of biology and pathophysiology[J]. F1000Research, 2018, 7.
- [12] Karlo C A, Di Paolo P L, Chaim J, et al. Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations[J]. Radiology, 2014, 270(2): 464-471.
- [13] Motzer R J, Jonasch E, Agarwal N, et al. Kidney cancer, version 2.2017, NCCN clinical practice guidelines in oncology[J]. Journal of the national comprehensive cancer network, 2017, 15(6): 804-834.
- [14] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma[J]. Nature, 2013, 499(7456): 43.
- [15] Dalglish G L, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes[J]. Nature, 2010, 463(7279): 360-363.
- [16] Shinagare A B, Vikram R, Jaffe C, et al. Radiogenomics of clear cell renal cell carcinoma: preliminary findings of the cancer genome atlas–renal cell carcinoma (TCGA–RCC) imaging research group[J]. Abdominal imaging, 2015, 40(6): 1684-1692.
- [17] Rutman A M, Kuo M D. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging[J]. European journal of radiology, 2009, 70(2): 232-241.
- [18] Jaffe C C. Imaging and genomics: is there a synergy?[J]. Radiology, 2012, 264 (2): 329-331.
- [19] Bishop C M. Pattern recognition and machine learning[M]. Springer, 2006.
- [20] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436.

- [21] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [22] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]//Proceedings of the 25th international conference on machine learning. ACM, 2008: 160-167.
- [23] Carrim Z I, Murchison J T. The prevalence of simple renal and hepatic cysts detected by spiral computed tomography[J]. Clinical radiology, 2003, 58(8): 626-629.
- [24] Zhang J, Lefkowitz R A, Bach A. Imaging of kidney cancer[J]. Radiologic Clinics, 2007, 45(1): 119-147.
- [25] Gillies R J, Kinahan P E, Hricak H. Radiomics: images are more than pictures, they are data[J]. Radiology, 2015, 278(2): 563-577.
- [26] Guo Z, Shu Y, Zhou H, et al. Radiogenomics helps to achieve personalized therapy by evaluating patient responses to radiation treatment[J]. Carcinogenesis, 2015, 36(3): 307-317.
- [27] West C M, Barnett G C. Genetics and genomics of radiotherapy toxicity: towards prediction[J]. Genome medicine, 2011, 3(8): 52.
- [28] Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician[J]. Lung cancer, 2018, 115: 34-41.
- [29] Parmar C, Leijenaar R T H, Grossmann P, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer[J]. Scientific reports, 2015, 5: 11044.
- [30] Lee G, Lee H Y, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art[J]. European journal of radiology, 2017, 86: 297-307.
- [31] Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology[J]. Annual review of biomedical engineering, 2016, 18: 387-412.
- [32] Fogel I, Sagi D. Gabor filters as texture discriminator[J]. Biological cybernetics, 1989, 61(2): 103-113.
- [33] Suri J S, Singh S, Reden L. Computer vision and pattern recognition techniques for 2-D and 3-D MR cerebral cortical segmentation (Part I): a state-of-the-art review[J]. Pattern Analysis & Applications, 2002, 5(1): 46-76.
- [34] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [35] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
- [36] Merkow J, Marsden A, Kriegman D, et al. Dense volume-to-volume vascular boundary detection[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2016: 371-379.
- [37] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). IEEE, 2016: 565-571.
- [38] Kamnitsas K, Ledig C, Newcombe V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical image analysis, 2017, 36: 61-78.
- [39] Heller N, Sathianathan N, Kalapara A, et al. The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes[J]. arXiv preprint arXiv:1904.00445, 2019.
- [40] Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public

- information repository[J]. Journal of digital imaging, 2013, 26(6): 1045-1057.
- [41] Gao J, Aksoy B A, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal[J]. Science signaling, 2013, 6(269): p11-p11.
- [42] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.
- [43] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [44] Cuingnet R, Prevost R, Lesage D, et al. Automatic detection and segmentation of kidneys in 3D CT images using random forests[C]//International conference on medical image computing and computer-assisted intervention. Springer, Berlin, Heidelberg, 2012: 66-74.
- [45] Yang G, Gu J, Chen Y, et al. Automatic kidney segmentation in CT images based on multi-atlas image registration[C]//2014 36th annual international conference of the IEEE engineering in medicine and biology society. IEEE, 2014: 5538-5541.
- [46] Linguraru M G, Wang S, Shah F, et al. Automated noninvasive classification of renal cancer on multiphase CT[J]. Medical physics, 2011, 38(10): 5738-5746.
- [47] Yu Q, Shi Y, Sun J, et al. Crossbar-Net: a novel convolutional neural network for kidney tumor segmentation in CT images[J]. IEEE transactions on image processing, 2019.
- [48] Yang G, Li G, Pan T, et al. Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module[C]//2018 24th international conference on pattern recognition (ICPR). IEEE, 2018: 3790-3795.
- [49] Wu G, Chen Y, Wang Y, et al. Sparse representation-based radiomics for the diagnosis of brain tumors[J]. IEEE transactions on medical imaging, 2017, 37(4): 893-905.
- [50] Isensee F, Petersen J, Kohl S A A, et al. nnU-Net: breaking the spell on successful medical image segmentation[J]. arXiv preprint arXiv:1904.08128, 2019.
- [51] Milletari F, Navab N, Ahmadi S A. V-net: fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). IEEE, 2016: 565-571.
- [52] Li X, Chen H, Qi X, et al. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes[J]. IEEE transactions on medical imaging, 2018, 37(12): 2663-2674.
- [53] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
- [54] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
- [55] Roy A G, Navab N, Wachinger C. Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 421-429.
- [56] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
- [57] Wong K C L, Moradi M, Tang H, et al. 3d segmentation with exponential logarithmic loss for highly unbalanced

- object sizes[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 612-619.
- [58] Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks[J]. Medical image analysis, 2017, 35: 18-31.
- [59] Karlo C A, Di Paolo P L, Chaim J, et al. Radiogenomics of clear cell renal cell carcinoma: associations between CT imaging features and mutations[J]. Radiology, 2014, 270(2): 464-471.
- [60] Chen X, Zhou Z, Hannan R, et al. Reliable gene mutation prediction in clear cell renal cell carcinoma through multi-classifier multi-objective radiogenomics model[J]. Physics in Medicine & Biology, 2018, 63(21): 215008.
- [61] Kocak B, Ates E, Durmaz E S, et al. Influence of segmentation margin on machine learning-based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas[J]. European radiology, 2019, 29 (9): 4765-4775.
- [62] Koçak B. Radiomics with artificial intelligence: a practical guide for beginners[J]. Diagn interv radiol, 2019.
- [63] Hu Y, Qiao M, Guo Y, et al. Reproducibility of quantitative high-throughput BI-RADS features extracted from ultrasound images of breast cancer[J]. Medical physics, 2017, 44(7): 3676-3685.
- [64] Kocak B, Durmaz E S, Ates E, et al. Unenhanced CT texture analysis of clear cell renal cell carcinomas: a machine learning-based study for predicting histopathologic nuclear grade[J]. American journal of roentgenology, 2019, 212(6): 132-139.
- [65] Ahn S J, Kim J H, Lee S M, et al. CT reconstruction algorithms affect histogram and texture analysis: evidence for liver parenchyma, focal solid liver lesions, and renal cysts[J]. European radiology, 2019, 29(8): 4008-4015.
- [66] Koo T K, Li M Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research[J]. Journal of chiropractic medicine, 2016, 15(2): 155-163.
- [67] Dormann C F, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance[J]. Ecography, 2013, 36(1): 27-46.
- [68] Hall M A. et al. Correlation-based feature selection for machine learning[D]. Hamilton: Univ. of Waikato, 1999.
- [69] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective[J]. Journal of the royal statistical society: series B (statistical methodology), 2011, 73(3): 273-282.
- [70] Kononenko I. Estimating attributes: analysis and extensions of RELIEF[C]//European conference on machine learning. Springer, Berlin, Heidelberg, 1994: 171-182.
- [71] Lings G, Menze B H, Lashkari D, et al. Detecting stable distributed patterns of brain activation using gini contrast[J]. NeuroImage, 2011, 56(2): 497-507.
- [72] Hassanat A B, Abbadi M A, Altarawneh G A, et al. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach[J]. arXiv preprint arXiv:1409.0919, 2014.
- [73] Hussain M A, Hamarneh G, Garbi R. Noninvasive determination of gene Mutations in clear cell renal cell carcinoma using multiple instance decisions aggregated CNN[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2018: 657-665.

Acknowledgement

This dissertation was implemented under the supervision of Prof. Yuanyuan Wang, Prof. Juha Plosila, and Dr. Mohammad Hashem Haghbayan.

Although I have to admit the stressful master life, I am still grateful for all of this because more than two years in our medical signal processing lab has brought a lot to me. Not only did I learn a lot of basic medical image processing knowledge, but also there was huge growth on my personality.

First of all, I am really thankful for Prof. Yuanyuan Wang. He played an important example to show me what an excellent and wise scholar, teacher should be. It is he that promoted me to grow fast and taught me to be down-to-earth. As my supervisor, he gave me a lot of tolerance even when I made serious faults, and continually encourage me to do better.

At the same time, I would like to thank Prof. Juha Plosila and Dr. Mohammad Hashem Haghbayan. They gave me important supervision and advices to review this thesis. Also, with Prof. Juha, I have experienced impressive and valuable study and life in University of Turku. This brought a lot of beautiful memory to me.

Secondly, I want to thank other teachers in our lab. We had the group meeting together each month, and they all gave me help or advices for my study or life. Thank you!

Finally, I also need to thank the guys in Genetics building. It is you that accompanied me for every day in our master life. We shared our joy and even anxiety together. It is really lucky for me to meet you.