



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

의료영상에서 딥러닝을 이용한  
영상분할 연구

A research on medical image segmentation  
with deep learning

울산대학교 대학원  
의학과  
함성원

의료영상에서 딥러닝을 이용한  
영상분할 연구

지도교수 김 호 성 김 남 국

이 논문을 공학박사 학위 논문으로 제출함

2021 년 2 월

울 산 대 학 교 대 학 원

의 학 과

함 성 원

함성원의 공학박사 학위 논문을 인준함

심사위원 김 남 국 (인)

심사위원 김 호 성 (인)

심사위원 서 준 범 (인)

심사위원 윤 지 혜 (인)

심사위원 홍 헬 렌 (인)

울 산 대 학 교 대 학 원

2021 년 2 월

## **Abstract**

Medical image segmentation plays a critical role in computer-aided diagnosis, image quantification, and surgical planning, which identifies the pixels of homogenous regions including in organs and lesions and provides important information about the shapes and volumes of the organs and lesions. However, it could be one of the most difficult and tedious tasks to be performed by humans consistently. Therefore, there has been a good amount of research to propose various semi- or automatic segmentation methods, which depend mainly on conventional image processing and machine learning methods. However, these methods may be vulnerable to variations in image acquisition, anatomy, and disease. Due to the above problems faced in conventional image segmentation methods, many scholars continue to seek more robust medical image segmentation methods.

In recent years, the deep learning model has been widely applied and popularized in computer vision. This success has been rapidly applied in the area of medical imaging. In particular, deep learning has achieved a leap in precision and robustness with regard to variations of anatomy and disease. Several deep convolutional neural network (CNN) models have been proposed such as Residual Net, Visual Geometry Group (VGG), fully convolutional network (FCN) and U-Net. These models provide not only state-of-the-art performance for image classification, segmentation, object detection and tracking tasks but also a new perspective on image processing. Therefore, at present, deep learning can assist radiologists and surgeons to segment various anatomic structures as a reference and multiple abnormalities in computed tomography (CT) or magnetic resonance imaging (MRI) images.

In this research, we conducted various experiments to find and evaluated adequate deep learning based semantic segmentation models in medical images from the viewpoint

of their accuracy in the clinical context. The aim of this study was two-fold as follows: 1) identifying and/or developing a deep learning based semantic segmentation model and the properties of an imaging modality that are adequate for the clinical context. 2) Solving specific tasks including smart labeling with humans in the loop, fine-tuning the models with different label levels in imbalanced datasets, and comparing deep learning and human segmentation where these models are developed and applied. For achieving these tasks or meeting these objectives, we proposed a fully automatic segmentation network with various kinds of CNN models considering organ-, image modality-, and image reconstruction-specific variations. Toward this, segmentation of a glioblastoma and acute stroke infarct in brain MRI, mandible and maxillary sinus in cone-beam computed tomography (CBCT), breast and other tissues in MRI, and pancreas cancer in contrast-enhanced CT were all performed in actual clinical settings. Basically, in case of slices with more thickness, 2D semantic segmentation shows better performances. Additionally, pre-processing is sensitive to developing robust segmentation that needs image normalization and various augmentations. However, because the modern graphics processing unit (GPU) lacks memory for 3D semantic segmentation, cascaded semantic segmentation or patch-based semantic segmentation gives better results. An anatomic variation could be easily trained by semantic segmentation, but disease variation of cancer is hard to be trained. Further, size-invariant semantic segmentation could be one of the important issues in medical image segmentation. Variation of contrast agent uptake may be vulnerable to the overall performance of semantic segmentation. For multi-center evaluation, subtle variation including variations in vendors' image protocols and high noise levels at different centers may cause problems to train robust semantic segmentation. Furthermore, as labeling of semantic segmentation is very tedious and time-consuming, deep learning based smart labeling is needed.

Based on these issues, we have developed and evaluated various applications with

semantic segmentation in medical images including smart labeling, robust radiomics analysis and disease pattern segmentation, and automated segmentation.

We concluded that adequate semantic segmentation with deep learning in medical images can improve the segmentation quality, which can be helpful in computer-aided diagnosis (CAD), image quantification, and surgical planning in actual clinical settings. Medical image segmentation and its application may be sufficient to provide practical utility to many physicians and patients who do not need to learn sectional anatomy.

**Key words:** Smart labeling, Cascaded CNN, CNN, Deep learning, MRI, CT, Semantic segmentation.

# Contents

Abstract.....	i
Contents .....	iv
List of Tables.....	vi
List of Figures .....	vii
1     Introduction.....	1
1.1     Motivations .....	3
1.2     Contributions .....	5
2     Background .....	6
2.1     CNN.....	6
2.2     General segmentation architectures .....	7
2.2.1     FCN.....	7
2.2.2     U-Net.....	8
2.2.3     Cascaded U-Net .....	9
2.2.4     No-new-U-Net .....	10
2.3     Evaluation metrics .....	12
3     Organ specific segmentation .....	13
3.1     Glioblastoma in brain MRI .....	13
3.2     Infarct in Brain MRI.....	20
3.3     Supine-prone tissues in Breast MRI.....	26
3.4     Kidney substructures with Renal Cell Carcinoma (RCC) in kidney CT .....	33
3.5     Pancreatic cancer in Pancreas CT .....	37
3.6     Multi-structures in dental CBCT .....	40
4     Advanced issues in medical image segmentation.....	43
4.1     Strategy for medical image segmentation .....	43



4.2	Smart labeling with human in the loop .....	47
4.3	Fine-tuning with different level labels in imbalanced datasets.....	48
4.4	Comparison between deep learning based and human segmentations in radiomics	
	56	
5	Discussion .....	60
6	Conclusion .....	69
	References.....	70
	Abstract (In Korean).....	75
	Acknowledgements .....	78

## List of Tables

Table 3.1 Evaluation results of fine segmentation. (up: enhance tumor, down: tumor core)	
.....	19
Table 3.2 The DSC results of the application of various networks and losses in the second stage.	26
Table 4.1 Summary of organ specific segmentation.	46
Table 4.2 Ratios of six image patterns	55
Table 4.3 Agreement among radiologists, SVM, and deep learning with and/or without fine-tuning	55
Table 4.4 Ratios of each image patterns by the fine-tuning model in three typical DILDs	55
.....	55

## List of Figures

Figure 2.1 An illustration of the architecture of our CNN.....	7
Figure 2.2 FCN architectures.....	8
Figure 2.3 U-Net architecture. ....	9
Figure 2.4 Cascaded U-Net methods: (a) Step 1 of cascaded U-Net: The first U-Net learns to segment location pancreas from whole CT images, (b) Step 2 of cascaded U-Net: The second U-Net learns to fine-segment pancreas lesions from step 1 of the cascade. ....	10
Figure 2.5 nnU-Net architecture .....	12
Figure 3.1 Sub-regions of GBM [47]. (a) Tumor core visible in T2, (b) Enhancing tumor structures visible in t1c (blue) surrounding the cystic/necrotic components of the core (green), (c) Segmentations are combined to generate the final labels of the tumor structures, (d) Edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), and enhancing core (blue). ....	14
Figure 3.2 Overall flow of GBM segmentation study. ....	16
Figure 3.3 Example of brain tumor and mask image. (a) Original brain tumor image, (b) Overlay GBM mask, (c) Enhance tumor, and (d) Tumor core. ....	18
Figure 3.4 Examples of the results of infarct core segmentation on DWI: (a) ADC maps, B) b1000 image, (c) Initial segmentation result, (d) Final segmentation result, and (e) Gold standard [53]. ....	22
Figure 3.5 Examples of the results of infarct segmentation. (a) Original image, (b) Ground truth, (c) Results of the application of DSC+Focal loss network in the second stage.....	25
Figure 3.6 Overall procedure of semantic segmentation in supine and prone breast MRI.	

.....	27
Figure 3.7 Example multi-region labels. (a) Prone MRI, (b) Supine MRI.....	27
Figure 3.8 Example multi-planner reformatted views. (a) Prone MRI, (b) Supine MRI. .....	28
Figure 3.9 Kidney and RCC label image example.....	34
Figure 3.10 Pancreas and focal legions image example.....	38
Figure 3.11 Examples of parenchyma and focal lesions segmentation. (a) Original image, (b) Ground truth (c) Segmentation results .....	40
Figure 3.12 Multi-structures segmentation examples .....	43
Figure 4.1 Flowchart for strategy in medical image segmentation. ....	45
Figure 4.2 Typical ROIs in six image patterns of diffuse interstitial lung disease. ((a) Normal parenchyma, (b) Ground-glass opacity, (c) Reticular opacity, (d) Honeycombing, (e) Emphysema, and (f) Consolidation.).....	49
Figure 4.3 Examples of HRCT images (upper row) and their corresponding classification results on typical six image patterns of DILD by the gold standard (lower row). (Normal, green; ground-glass opacity, yellow; reticular opacity, red; honeycombing, light blue; emphysema, dark blue; consolidation, pink) 51	
Figure 4.4 Examples of HRCT images with manual drawings of reader 1 and reader 2 and the results of semantic segmentation before fine-tuning and after fine- tuning in the same. Gold standard masks are drawn by (a) Reader 1 and (b) Reader 2 and semantic segmentation results (c) without fine-tuning and (d) with fine-tuning. (Normal, green; ground-glass opacity, yellow; reticular opacity, red; honeycombing, light blue; emphysema, dark blue; consolidation, pink).....	53
Figure 4.5 Further examples of HRCT images (middle row) and the results of semantic	

segmentation before fine-tuning and after fine-tuning in the same HRCT: (a)  
Original images, (b) without fine-tuning, and (c) with fine-tuning. ....56

# 1 Introduction

Generally speaking, image segmentation is one of the most interesting and challenging problems in computer vision. Image segmentation an image area or volume into nonoverlapping, connected regions, being homogeneous characteristics with respect to signal and semantics [1-3]. In medical imaging, these regions often correspond to different tissue classes, organs, pathologies, or other biologically relevant structures [4]. Consequently, medical image segmentation is of substantial importance in providing non-invasive information about human body structures that helps radiologists and physicians to visualize and study the anatomy of the structures, simulate biological processes, localize pathologies, track the progress of diseases, and evaluate the need for radiotherapy or surgeries [4, 5]. For these reasons, segmentation is an essential part of any CAD system, and the system's functionality depends heavily on segmentation accuracy [6-8].

However, low contrast, noise and other imaging ambiguities make medical image segmentation difficult [9]. There are many computer vision techniques for image segmentation, but some have been specifically applied to medical image computing [10]. Historically there have been many kinds of conventional segmentation studies in medical image depending on the expertise of the clinician. Basically, manual segmentation, which explicitly defines the tissue grade of each pixel with a tool such as a paint brush, remains the standard for many imaging applications [11]. In recent years, the principle of feedback control theory has been incorporated into the segmentation, providing users with much greater flexibility and allowing errors to be corrected (semi-)automatically. This interactive method is useful when the clinician can provide some information, such as a rough outline of the area to be segmented or the seed area. The algorithm can then iteratively improve these subdivisions with or without the guidance of the clinician concerned.

The image-based segmentation method starts a template and adjusts its shape according to

the image data while minimizing the measurement of integration errors such as the active contour model and its deformation [12, 13]. Often, many methods parameterize the shape of the template for a given structure and often rely on control points along the boundary. Then the whole shape is transformed to match the new image. These are the subdivisions based on shape. Two of the most common shape-based techniques are active shape models and active appearance models. The subjective surface segmentation method is based on the evolutionary idea of segmentation functions controlled by advection-diffusion models [14, 15]. To segment an object, you need a segmentation seed (i.e. a starting point that determines the approximate position of the object in the image). As a result, the initial segmentation function is configured. The idea of the subjective surface method is that the position of the seed is the main factor that determines the shape of this segmentation function.

In many applications, clinical experts can manually label multiple images. Segmenting the invisible image is a matter of extrapolating from the manually labeled training image. This style of method is commonly referred to as the atlas-based segmentation method [16-18]. Parametric atlas methods typically combine these training images into a single atlas image, and nonparametric atlas methods generally use all training images individually. Atlas-based methods generally require the use of image registration to align an atlas image or image to a new, invisible image. However, accurate segmentation of medical images faces many challenges in this method [19]. First of all, many anatomical structures are inhomogeneous with respect to spatial repetitiveness of individual pixel/voxel intensities or their grouped co-occurrences. Many medical images are difficult to separate because of similar visual appearances of vessels and tissues as well as overlapping organs and background appearances, respectively [20]. Low-contrast medical images present additional challenges as do 3D and 4D (spatial-temporal) images. These challenges include inter- and intrasubject variability of shapes and textures of an object-of-interest in images collected over some time for the same subject or a group of different subjects. Consequently, a particular segmentation technique

using conventional methods may work well for one subject and not for another, or work only for certain images of a particular subject. In particular, contour-, region-, and pixel-/voxel-based techniques aim at getting boundaries of goal objects, forming connected regions occupied by these objects, or performing just pixel-/voxel-wise classification, respectively.

In recent years, the deep learning model has been widely applied and popularized in computer vision. The amazing success of this model when compared with conventional segmentation and traditional machine learning models is the advancements in neural networks. This model learns high-level features from data in an incremental manner, which eliminates the need for domain expertise and handcrafted feature extraction [21-24]. This success has been quickly applied to the medical image area. In particular, deep learning has taken a leap forward with regard to the accuracy and robustness of medical image segmentation for various anatomies and diseases [4, 25, 26]. Additionally, it solves the problem in an end-to-end manner. Therefore, deep learning could improve medical image segmentation.

## **1.1 Motivations**

There has been a rapid growth in the use of deep learning based medical image segmentation with various kinds of deep CNN models such as Residual Neural Net (ResNet), VGG, FCN, and U-Net [27]. Not only do these models provide a state-of-the-art presentation for image classification, segmentation, object detection, and tracking tasks, but they also afford a new perspective on image processing [20, 28, 29]. Right now, a deep learning based method could be used to help radiologists and surgeons to segment numerous anatomic structures of a reference and multiple abnormalities in CT or MRI images. In this paper, we have evaluated the feasibility and applicability of deep learning based medical image segmentation methods in various experiments and have attempted to find an appropriate deep learning based semantic segmentation model in medical images in terms of accuracy and its use in clinical contexts. In



addition, we have raised several issues including efficiency, robustness, and clinical meaning of deep learning based semantic segmentation models from the viewpoint of their application to actual clinical settings. The purpose of this study is two-fold as follows: 1) Finding and developing various kinds of semantic segmentation models based on deep learning that is appropriate for the properties of medical images and to the medical context. 2) Solving specific issues concerning efficiency, clinical meaning, and robustness of these segmentation methods including smart labeling with a human in the loop, fine-tuning them with different levels of data in imbalanced datasets, and their application to image segmentation in radiomics analysis.

## **1.2 Contributions**

The several contributions to medical image segmentation in this paper include the following:

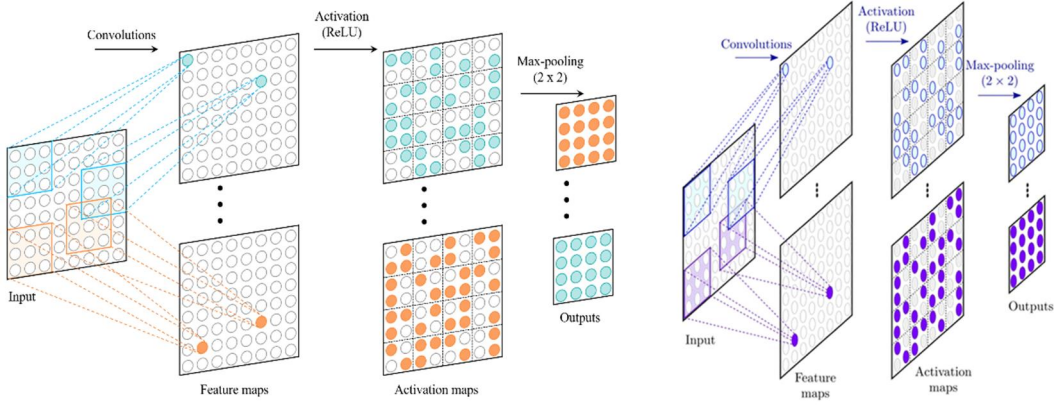
- 1) Finding and developing various kinds of semantic segmentation models based on deep learning that is appropriate for the properties of medical images and medical contexts.
- 2) Solving specific issues on efficiency, robustness, and clinical meaning of these segmentation methods.
  - Developing efficient and robust semantic segmentation on renal cell carcinoma in kidney CT by using smart labeling.
  - Evaluating and fine-tuning with different level data in imbalanced diffuse infiltrative lung disease progression dataset using semantic segmentation.
  - Developing robust semantic segmentation for glioblastoma radiomics analysis in brain MRI.

## **2 Background**

### **2.1 CNN**

A CNN is a type of artificial neural network that is used for image recognition and processing and is specifically designed to process pixel data designed for image analysis as shown in Figure 2.1 [27]. CNN uses deep learning with powerful image processing and artificial intelligence (AI) to both create and describe tasks often with the help of recommendation systems and natural language processing to perform image and video recognition, image classification, medical image analysis, and more [28, 29].

A CNN is a branch of neural networks and consists of a stack of layers each performing a specific operation, e.g., convolution, pooling, loss calculation, etc. Each intermediate layer receives the output of the previous layer as its input. The beginning layer is an input layer, which is directly connected to an input image with the number of pixels in the input image being equal to the number of neurons. The next set of layers are convolutional layers that present the results of convolving a certain number of filters with the input data and perform as a feature extractor [30-33]. The filters, commonly known as kernels, are of arbitrary sizes and defined by designers depending on the kernel size. Each neuron responds only to a specific area of the previous layer called the receptive field. The output of each convolution layer is considered as an activation map, which highlights the effect of applying a specific filter on the input. Convolutional layers are usually followed by activation layers to apply non-linearity to the activation maps. The next layer can be a pooling layer depending on the design, and it helps to reduce the dimensionality of the convolution's output. To perform the pooling, a few strategies exist such as max pooling and average pooling. Finally, high-level abstractions are extracted by fully-connected layers. The weights of neural connections and the kernels are continuously optimized during the procedure of backpropagation in the training phase.



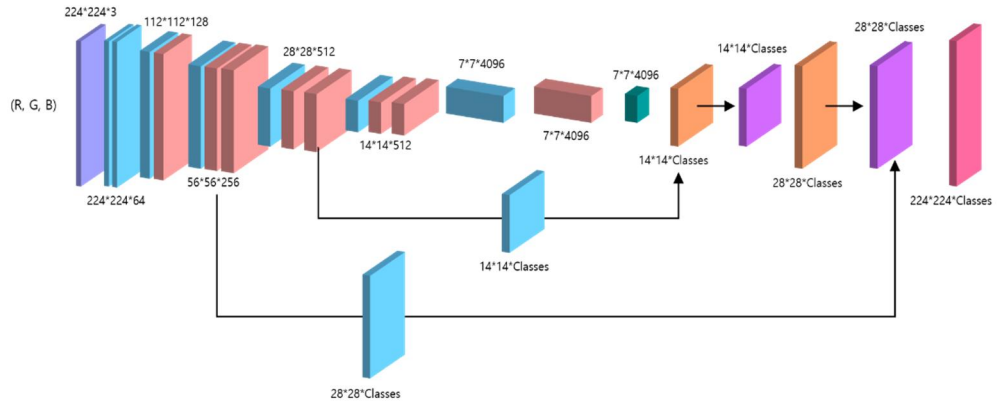
**Figure 2.1** An illustration of the architecture of our CNN.

## 2.2 General segmentation architectures

### 2.2.1 FCN

FCN is a modified CNN-based model that has previously shown excellent performance in image classification for the semantic segmentation model. FCN transforms image pixels to pixel categories using a CNN. FCN transforms the height and width of the middle layer feature map back to the size of the input image via a transposed convolution layer. Consequently, the prediction has a one-to-one correspondence with the input image in a spatial dimension with height and width. In the FCN so developed, the last fully connected layer was replaced with a fully convolutional layer [34]. As shown in Figure 2.2, to get the output of the segmentation map, FCN usually consists of two parts. The downsampling path captures semantic or context information, and the upsampling path recovers spatial information. Down-sampling paths are used to extract and interpret context, while upsampling paths are used to enable precise localization. Upsampling also completely recovers spatial information that was lost in pooling or downsampling.

This major improvement allows the network to have a dense pixel-wise prediction. To achieve better localization performance, high-resolution activation maps are combined with upsampled outputs and passed to the convolution layers to assemble more accurate outputs.

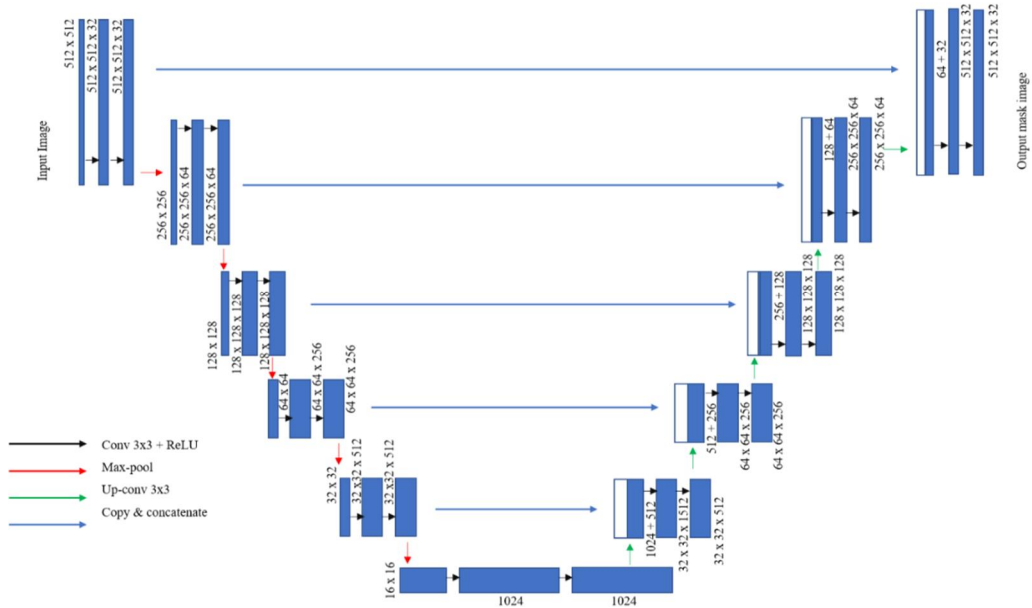


**Figure 2.2** FCN architectures.

### 2.2.2 U-Net

The typical use of convolutional networks is on classification tasks, where the output to an image is a single class label. However, in many visual tasks, especially in biomedical image processing, the desired output should include localization, i.e., a class label is supposed to be assigned to each pixel. Moreover, thousands of training images are usually beyond reach in biomedical tasks. The U-Net [35-37], illustrated in Figure 2.3 which was designed for medical images, concatenates downsample layers and its counterpart of upsampling layers. These networks, mentioned above, emphasize interconnections between downsample layers and upsample layers. It consists of a reducing path (left side) and an extending path (right side). The reducing path follows the typical architecture of a convolutional network. To explain in a 2D network, it consists of the repeated application of two  $3 \times 3$  convolutions (unpadded convolutions) each followed by a rectified linear unit (ReLU) and a  $2 \times 2$  max-pooling operation with stride 2 for downsampling. At each downsampling step, we doubled the number of feature channels. Every step in the extending path consists of an upsampling of the feature map followed by a  $2 \times 2$  convolution (“upconvolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path and two  $3 \times 3$  convolutions, each followed by a ReLU. The cropping is necessary due to

the loss of border pixels in every convolution. At the final layer, a  $1 \times 1$  convolution is used to map each 64-component feature vector to the desired number of classes. In total, the network has 23 convolutional layers. To allow a seamless tiling of the output segmentation map. It is important to select the input tile size such that all  $2 \times 2$  max-pooling operations are applied to a layer with an even x- and y-size.

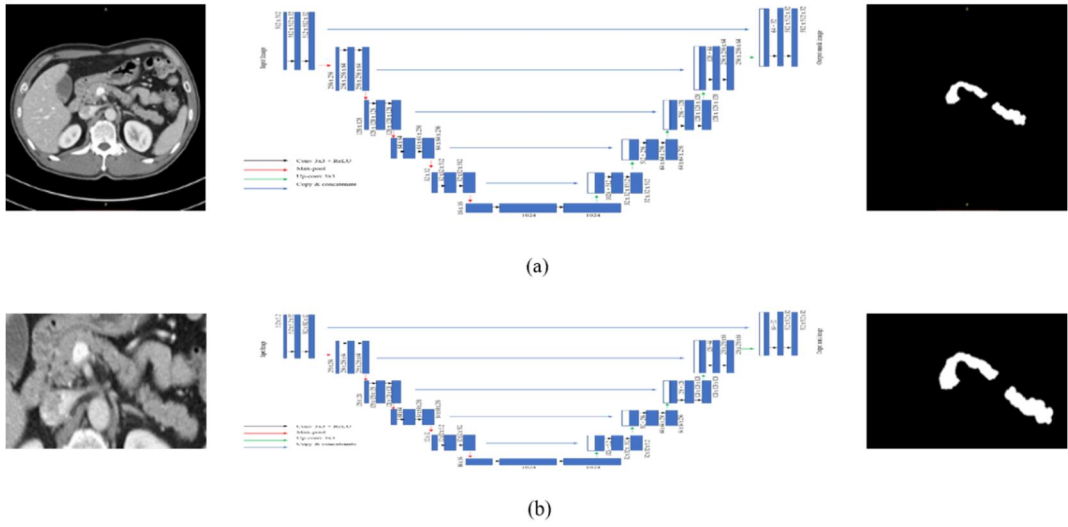


**Figure 2.3** U-Net architecture.

### 2.2.3 Cascaded U-Net

Cascaded U-Net is a method of segmentation an image by devising two stages. The first step was used to train the model with an initial prediction of segmentation labels, and the second step was implemented to further fine-tune labels for the results of the first step. In the first step, the model roughly predicted large and coarse displacements using 3D U-Net. At this time, since the whole size of the image was reduced, the image resolution is quite poor but the computation cost is low. After all, the first step is to determine the approximate location and shape of the whole image. In order to get a more accurate boundary of the region(s) of interest

(ROI), when the approximate location was obtained as a result of the segmentation in the first step, the original image was cropped by a specific size based on the ROI result from the first learning model [38-40]. In this stage, it can provide more accurate segmentation. In fact, this step narrows and simplifies the search space for the network to determine which voxels belong to the background or foreground class. This strategy has been successful in many computer vision problems [41]. This architecture is illustrated in a training example in Figure 2.4.



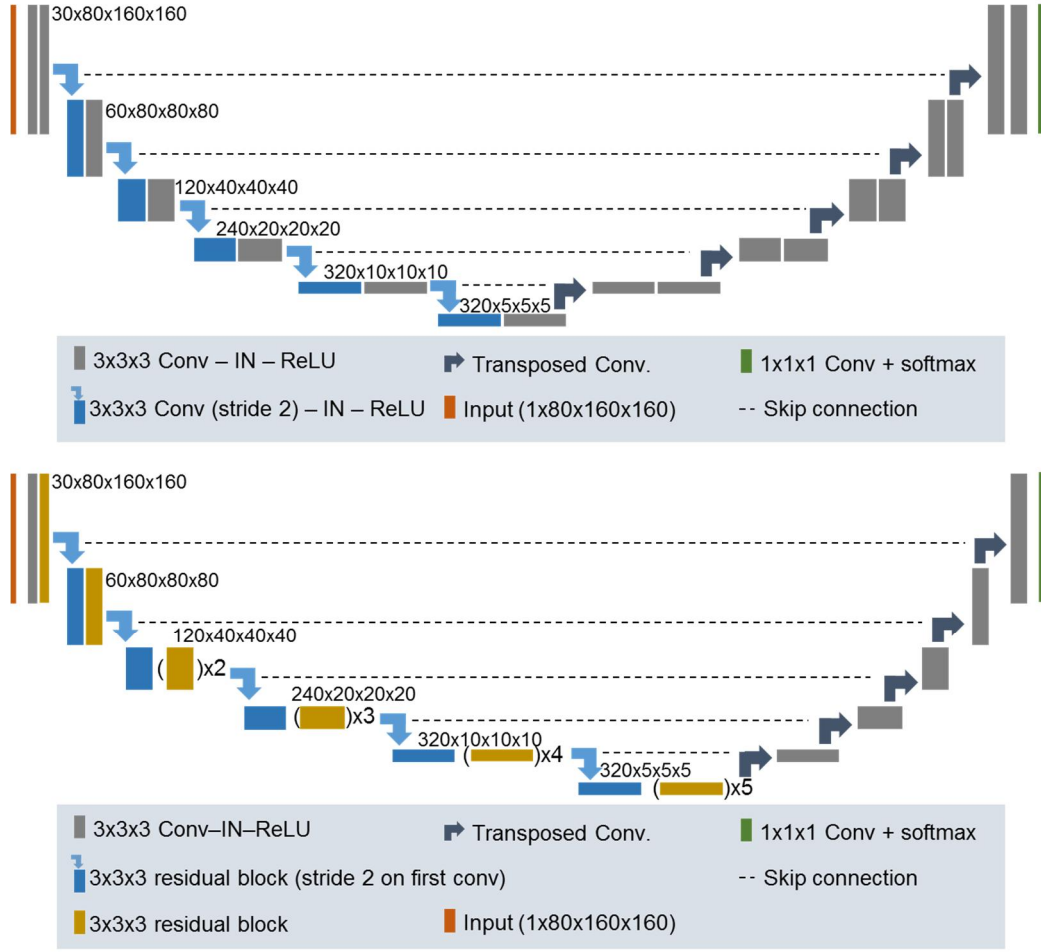
**Figure 2.4** Cascaded U-Net methods: (a) Step 1 of cascaded U-Net: The first U-Net learns to segment location pancreas from whole CT images, (b) Step 2 of cascaded U-Net: The second U-Net learns to fine-segment pancreas lesions from step 1 of the cascade.

#### 2.2.4 No-new-U-Net

Medical images commonly encompass a third dimension, which is why we consider a pool of basic U-Net architectures consisting of a 2D U-Net, a 3D U-Net, and a U-Net cascade. While the 2D and 3D U-Nets generate segmentations at full resolution, the cascade first generates low-resolution segmentations and subsequently refines them. Our architectural modifications, as compared to the U-Net’s original formulation, are close to negligible. and

instead, we focus our efforts on designing an automatic training pipeline for these models. The U-Net is a successful encoder-decoder network that has received a lot of attention in recent years. Its encoder part works similarly to a traditional classification CNN in that it successively aggregates semantic information at the expense of reduced spatial information. Since in segmentation, both semantic as well as spatial information are crucial for the success of a network, the missing spatial information must somehow be recovered. The U-Net does this through the decoder that receives semantic information from the bottom of the 'U' and recombines it with higher-resolution feature maps obtained directly from the encoder through skip connections. Unlike other segmentation networks, such as FCN and previous iterations of DeepLab this allows the U-Net to segment fine structures particularly well. Just like the original U-Net, we use two plain convolutional layers between poolings in the encoder and transposed convolution operations in the decoder. We deviate from the original architecture in that we replace the ReLU activation functions with leaky ReLUs and use instance normalization instead of the more popular batch normalization.





**Figure 2.5** nnU-Net architecture

## 2.3 Evaluation metrics

Evaluation metrics play an important role in achieving optimal classifier during classification training. Therefore, the selection of the appropriate evaluation metric is the key to identifying and obtaining the optimal classifier.

To accurately evaluate our various segmentation method's performance, we compared the conventional and deep learning based methods with the gold standard results using four metrics: Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), Harsdorf distance (HSD) [42], and mean absolute surface distance (MSD). Manual segmentation masks

were considered as the ground truth (GT), and the resultant deep learning masks were considered test mask (TM). The DSC is defined as a measure of the spatial overlap between GT and TM.

$$DSC(GT, TM) = \frac{2 \times |GT \cap TM|}{|GT| + |TM|} \quad (1)$$

JSC: similarity index between two masks, which is related to the DSC:

$$JSC(GT, TM) = \frac{|GT \cap TM|}{|GT \cup TM|} \quad (2)$$

MSD: the mean of the sum of the Euclidean distance (for each voxel) between mask contours, their average gives the MSD as follows:

$$MSD(GT, PM) = \frac{1}{N_{GT} + N_{PM}} \{ \sum_{i \in N_{GT}} d_{\min}(GT, TM) + \sum_{i \in N_{PM}} d_{\min}(TM, GT) \} \quad (3)$$

where  $N_{GT}$  and  $N_{TM}$  are the total number of voxels in the contour for GT and TM respectively. The distance values are obtained through the use of a 3D Euclidean distance transform. HD measures the maximal contour distance between the two segmentations:

$$d(X, Y) = \max[d_i(X, Y), i = 1 \dots N_X]$$

$$HD(GT, TM) = \max[d_{\min}(GT, TM), d_{\min}(TM, GT)] \quad (4)$$

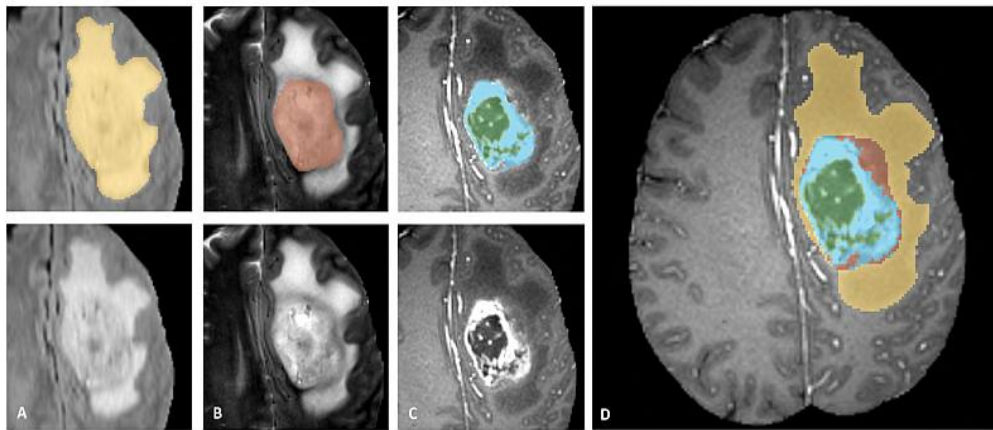
where  $d$  is the Euclidean distance between voxels  $x$  and  $y$ . All statistical tests were carried out using SciPy in Python.

### 3 Organ specific segmentation

#### 3.1 Glioblastoma in brain MRI

Though less common but really fatal, the brain tumor is one of the most infamous medical threats to human beings [43]. Patients with the most aggressive tumors have a life expectancy of fewer than two years [44]. Brain tumors can be divided into primary tumors or metastatic tumors based on their origin. Glioblastoma malformation (GBM), as one of the most frequent

chief tumors, is one of the main objects studied in the field of brain tumor segmentation. GBM has been graded into 4 levels by the World Health Organization (WHO). In particular, the Gliomas of grade III and IV are named as High-Grade Gliomas (HGG). They comprise malignant gliomas and mostly cause death [44-46]. GBM is also divided into several classes such as the entire tumor, enhancing tumor, tumor core (see Figure 3.1). The tumor core is the necrotic part, and the tumor enhancement is an area where the tumor progresses vigorously. Registration tracks the spatial mapping that aligns the moving images with the fixed images



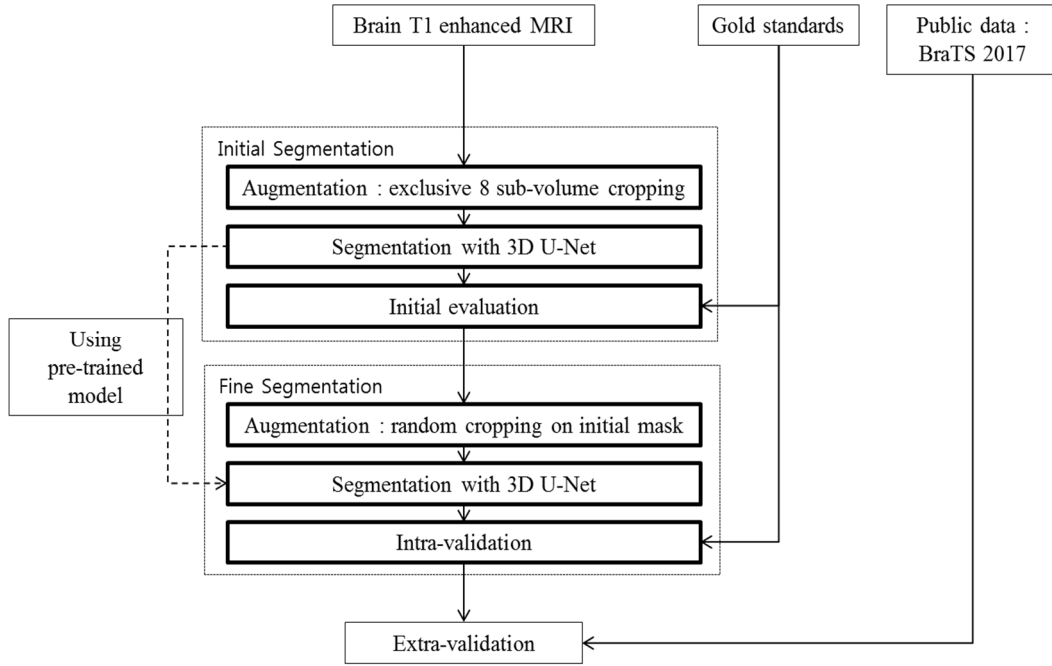
**Figure 3.1** Sub-regions of GBM [47]. (a) Tumor core visible in T2, (b) Enhancing tumor structures visible in t1c (blue) surrounding the cystic/necrotic components of the core (green), (c) Segmentations are combined to generate the final labels of the tumor structures, (d) Edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), and enhancing core (blue).

GBM segmentation plays an important role during treatment planning and follow-up evaluation. But it is time-consuming and prone to inter- and intra-rater variability. Therefore, automatic and reliable methods are desirable. Traditional image process algorithms and machine learning algorithms have been applied to brain tumor segmentation since a long time. Havaei et al. [46] proposed a semi-automatic classification method using the support vector machine (SVM). D Kwon et al. [44] proposed a generative model which generate tumors and

edemas from several seed points by giving priority to the tumor shape. They achieved top ranking in BRATS 2013 challenge. D Zikic et al. [48] differentiate brain tumor and its sub regions with a discriminative approach based on decision forests. Recently, GBM segmentation has been achieved using CNN. However, brain tumor segmentation is a challenging task, due to their irregular shape, appearance, and location [46, 49].

In this work, we proposed a two-phase, fully-automated GBM (tumor enhancement, tumor core) segmentation method using 3D U-Net architecture. We then performed extra-validation for stability using the BRATS data.

On the whole, the overall procedure consisted of three main steps as shown in Figure 3.2. The first step was the pre-processing step, in which the data was intensity normalized using white stipe package in r and removed of air region. The second step involved the U-Net training in a supervised mode. We divided this step further into two phases. First, the initial segmentation of the brain tumor was performed using a conventional image processing method in a scratch manner. To find the tumor, we divided 8 subvolumes. In the second phase, the model was trained by the initial segmentation model by fine-tuning. We call this transfer learning. Moreover, by choosing a well-trained model, the segmentation of the GBM region was extracted and evaluated by DSC, JSC, MSD and HD with metrics.



**Figure 3.2** Overall flow of GBM segmentation study.

### *Dataset*

All the MR imaging data obtained from Asan Medical Center (AMC). The MRI scanning sequence is used in T1-weighted contrast-enhanced (T1ce) imaging. We defined the GBM and generated a gold standard mask based on the following criteria. We searched the electronic database of the Department of Radiology at our institution, retrospectively reviewed records for patients between March 2011 and March 2017, and identified 208 patients pathologically confirmed to have de novo glioblastoma. T1ce image was obtained for the patients ( $n = 200$ ). Patients were excluded if no histopathological specimen was available ( $n = 96$ ), or the image was unreadable (because of an artifact) ( $n = 1$ ). These steps yielded 103 consecutive patients (mean age: 62.4 years; male–female ratio: 61:42). The number of data is 103 patients included only Glioblastoma disease. These data consist of brain and mask. GT data was created by experts through manual drawing. We divided the dataset as follows: The training set contains

75 subjects. Each of the validation and testing set contains 14 subjects. Every patient has 2 classes-tumor regions, *viz.* enhance tumor, tumor core.

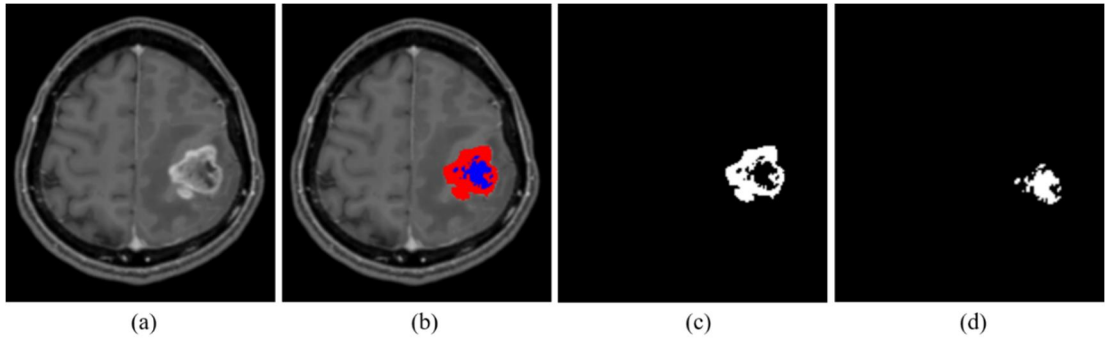
#### *MR protocol*

In the brain image study, MRI is an important diagnostic tool for the precise detection of various clinical symptoms. MRI images can extract visual information using multiple sequences such as T1-weighted (T1), T1ce, T2-weighted (T2) and T2-weighted fluid-attenuated inversion recovery (FLAIR). In particular, the T1ce image contrast permits the visualization of the enhancing part of the tumor as well as the necrotic part. The T1ce image was obtained with a high-resolution three-dimensional (3D) volume, using a gradient-echo T1-weighted sequence with the following parameters: repetition time (TR)/echo time (TE): 9.8/4.6 ms; flip angle: 10°; field of view (FOV): 256 mm; matrix: 512 x 512; and slice thickness: 1 mm with no gap. T1ce were obtained using a gradient-echo T1-weighted sequence – TR/TE: 7.6/3.7 ms; flip angle: 10°; FOV: 24 mm; matrix: 512 x 512; and slice thickness: 1.2 mm with no gap

#### *Initial and Fine Segmentation*

The size of GBM relative to the entire brain was too small. We needed to know where the GBM is located. So, we split the data into 8 sections considering the capacity of the network. Also, if the data were resized, it would have affected the raw image quality. So, we simply cropped the data into 8 sections. First, the images were cropped into exclusive 8 subvolumes to cover the entire brain to detect GBM regions. This step was called the initial segmentation phase. The datasets were divided into 600, 112, and 112 for training, validation, and test, respectively. We used the 3D U-Net, one of the most widely used CNN architectures, for image segmentation. These datasets were cropped (352 x 416 x 192) and resized (176 x 208 x 96) for

inputting into the network. Thereafter, data augmentation by random cropping on the mask was performed. This phase was called fine segmentation. For training, we used the pre-trained model from the initial segmentation. During training, we used the cross-entropy loss, the Adam optimizer with a learning rate of  $5 \times 10^{-5}$ , weight decay of  $1 \times 10^{-6}$ , and spatial dropout probability of 0.05. These methods with transfer learning have been developed. The number of random cropping was expanded to 32 on a log scale. All of the network's input size ( $176 \times 208 \times 96$ ) was resized from the original image ( $512 \times 512 \times 256$ ). As the number of random cropping increased, the time required per epoch also increased. However, the total number of epochs decreased.



**Figure 3.3** Example of brain tumor and mask image. (a) Original brain tumor image, (b) Overlay GBM mask, (c) Enhance tumor, and (d) Tumor core.

### *Results*

The result consisted of three stages including detection and segmentation accuracies of initial segmentation, segmentation accuracy of fine segmentation, and extra-validation. For the detection of the initial tumor region, we cropped the data per subject in 8 cubes. The dataset consisted of 3D image patches. The number of patches for the training stage is 600 and for the validation and test stages is 112 each. Table 1 shows the accuracy of how well the tumor region is found in cube units. In practice, we mainly use 3 metrics. They are precision, recall, and DSC. In the initial segmentation stage, we obtained detection accuracy scores of more than

90% for enhancing tumor and tumor core.

In the fine segmentation step, we performed data augmentation by random cropping. This data training takes the trained model from the initial tumor detection stage. Each random-cropped group increased the number of data on the log scale. As can be seen from the results in Table 1, the bigger number of data, the higher the dice score is. Also, in this step, we mainly used 4 metrics, i.e., DSC, JSC, MSD, and HSD. When the number of random cropped data was 16, the dice score was saturated. The mean and SD of DSC, JSC, MSD, and HSD of enhanced and necrosis regions at 16-random cropping with converged evaluation metrics were  $84.40\pm0.09\%$ ,  $72.50\pm0.12\%$ ,  $1.16\pm0.81$  mm, and  $25.38\pm22.86$  mm, and  $84.80\pm0.14\%$ ,  $75.90\pm0.20\%$ ,  $1.63\pm1.19$  mm, and  $15.87\pm5.10$  mm, respectively.

**Table 3.1** Evaluation results of fine segmentation. (up: enhance tumor, down: tumor core)

Number of random cropping	DSC	JSC	MSD	HD
1	$0.648\pm0.213$	$0.511\pm0.226$	$2.592\pm2.176$	$19.812\pm12.224$
2	$0.707\pm0.158$	$0.567\pm0.181$	$1.942\pm1.325$	$27.700\pm18.736$
4	$0.771\pm0.126$	$0.642\pm0.158$	$1.577\pm1.227$	$15.817\pm6.135$
8	$0.823\pm0.104$	$0.711\pm0.140$	$1.105\pm0.787$	$15.802\pm10.010$
16	$0.844\pm0.092$	$0.725\pm0.127$	$1.167\pm0.810$	$25.383\pm22.862$
32	$0.804\pm0.095$	$0.675\pm0.194$	$1.181\pm0.912$	$22.473\pm17.766$
Number of random cropping	DSC	JSC	MSD	HD
1	$0.445\pm0.282$	$0.324\pm0.231$	$4.446\pm4.172$	$22.538\pm8.971$
2	$0.641\pm0.262$	$0.517\pm0.258$	$2.517\pm2.236$	$17.402\pm4.541$
4	$0.717\pm0.202$	$0.594\pm0.242$	$1.608\pm1.246$	$16.019\pm4.130$
8	$0.778\pm0.150$	$0.658\pm0.198$	$1.363\pm0.927$	$15.933\pm4.155$
16	$0.848\pm0.147$	$0.759\pm0.202$	$1.639\pm1.199$	$15.870\pm5.105$
32	$0.811\pm0.147$	$0.707\pm0.231$	$1.884\pm1.207$	$15.991\pm4.753$



### **3.2 Infarct in Brain MRI**

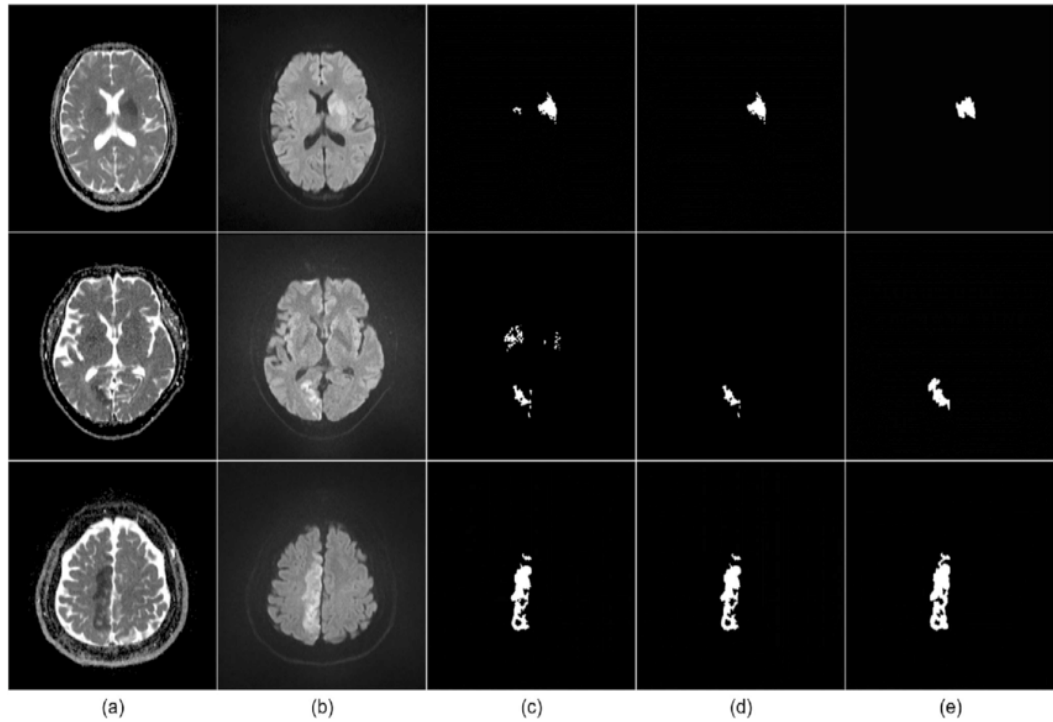
Stroke is a leading cause of death and incapacity worldwide. Acute ischemic strokes due to arterial obstruction account for 80% of all strokes [50, 51]. However, obstruction of an artery by blood clotting does not lead to instant necrosis of the brain area that is supplied blood by the blocked artery as the collateral vascular connection can partially reimburse for the blocked blood flow in the main artery. In general, it is assumed that if the blood clot does not dissolve, the infarct core will gradually expand into this hypoperfusion brain region over time. This area is commonly referred to as semi-shaded or at-risk tissue and represents the target of ischemic stroke treatment. Recent prospective randomized trials have shown the success of mechanical thrombectomy to be overwhelming [52]. Of late, more and more devices for mechanical thrombectomy have been developed as more and more research focuses on the development and evaluation of novel treatment approaches such as the use of neuroprotective drugs. In any case, clinical studies are required to show the efficacy of these new devices or treatment options. Although the clinical outcome (e.g., modified Rankin scale at 90 days post-stroke) is typically used as the primary endpoint for such studies, of late the follow-up stroke lesion volume is becoming more important as the alternative primary or secondary study endpoint. This is so not only because the continuous lesion volume has a higher statistical power compared to categorical outcome measures but also because it can be measured at an earlier time point. MRI is one of the most used techniques for subsequent brain lesion evaluation. However, quantitative measurement of lesion volume in subsequent imaging data requires precise segmentation, which is a tedious and complex task if done manually. This is so because the shape, size, and location of the brain differ considerably. Diffusion-weighted imaging (DWI) of patients with ischemic stroke during the acute or super acute stage presents a replacement for the infarct lesion to the infarct core. Infarct lesions in DWI are associated with infarct volume. Ischemic infarction can occur anywhere in the brain in various forms.

Therefore, segmentation of infarct lesions is important for quantifying infarct volume and determining treatment options for patients who have suffered an ischemic stroke.

Because ischemic infarctions of various shapes and sizes can occur anywhere in the brain, infarct segmentation must capture local information along with a global context dictionary. The lesions in the image constitute a high signal intensity at b1000, and low signal intensity at the apparent diffusion coefficient map (ADC) is a well-known image representing acute infarction. Acute ischemic infarct segmentation can help monitor infarct lesions, determine treatment options, and predict response to treatment in stroke patients.

Due to the importance of tracking lesion segmentation, several methods have been presented for semi-automatic or automatic stroke lesion segmentation in the past. For example, unsupervised k-means clustering (9), an active learning approach (10), etc. Advanced additional tree forests (11), Markov random field (MRF) models (12) and convolutional neural networks (13). Almost all recently proposed lesion segmentation techniques utilize local features and contexts that can be affected by signal noise, geometric distortion, magnetization non-uniformities, and anatomical changes. In addition, the previously described methods often suffer from normalization problems inherent in non-quantitative imaging methods, which is particularly important in the case of multicenter datasets that are usually collected using various imaging parameters.

Despite the widespread use of U-Net, nnU-Net, and Dense U-Net to enhance functionality by introducing skip-connected bridges between the encoding layer and the decoding layer, these methods were insufficient to capture the small and uncertain boundary regions' segmentation. These small and uncertain boundary regions were difficult to fine-segment. To solve this problem, we proposed an inexpensive but effective method of integrating into a recent semantic segmentation network using a two-mode image b1000 and ADC, including various lossy features such as focal and boundary loss.



**Figure 3.4** Examples of the results of infarct core segmentation on DWI: (a) ADC maps, (b) b1000 image, (c) Initial segmentation result, (d) Final segmentation result, and (e) Gold standard [53].

#### *Dataset*

In this study, a dataset of 429 patients with acute ischemic stroke was used for acute infarct segmentation. Patients visited the AMC emergency room from September 2005 to August 2015 with symptoms of acute ischemic stroke and had DWI (including both b0 and b1000 images) to confirm acute infarction. The demographic and clinical characteristics of all 429 patients are as follows: 269 male, 160 females; mean age: 72 years; age range: 24–98 years. This study was approved by the institutional review board (IRB), which gave up the requirement for informed consent.

### *MRI protocol*

The MRI scans were obtained using various machines including 1.5T (Magnetom Avanto, Siemens Healthcare, Erlangen, Germany; Genesis Signa, GE Healthcare, Milwaukee, WI, USA; Intera, Philips Healthcare, Best, Netherlands) and 3.0T (Ingenia CX, Philips Healthcare; Achieva, Philips Healthcare) with the following parameters: TR: 400–8000 ms; echo time (TE): 16.1 to 103 ms; flip angle: 18°–90°; matrix:  $256 \times 256$  to  $512 \times 512$ ; FOV:  $220 \times 220$  to  $250 \times 250$  mm; number of excitations: 1–6; number of slices: 20–45; and slice thickness: 3–5 mm. From DWI scan, ADC map was automatically created built-in software. The GT for acute infarct lesions in DWI was described by two expert neuroradiologists with 6 and 9 years of experience, in consensus. In detail, the radiologists described the maximum visual range of high signal intensity in the b1000 image with infarct lesion. They were also instructed to exclude areas of T2 shine-through artifact with the corresponding ADC images. These ROIs were labeled using in-house software (modified ImageJ).

### *Preprocessing*

In this study, the MRI images were obtained from different scanners according to different protocols. So, several pre-processing steps were performed before the experiments. Since MRI intensities do not have tissue-specific values (i.e. they have a wide range of intensities even within the same tissue, the same protocol, target, and scanner), intensity normalization is an important pre-processing step for MRI image analysis. For each image, we applied a linear transformation to a certain percentile (i.e., 0.005–0.995) of the intensity histogram within the brain regions. The brain regions were automatically determined based on tissue probability maps calculated using the Segment module of Statistical Parametric Mapping (SPM12). Subsequently, homogeneous linear resampling was performed in 3D volume. All images were resampled to a uniform matrix size of  $384 \times 384$  in the axial slice.

### *Cascaded method*

Our proposed infarct segmentation consisted of the following two phases: initial localization and fine segmentation.

#### *Phase 1: Initial localization - Instance segmentation using Mask R-CNN(Regions with CNN features)*

Instance segmentation identifies each object instance for every known object in the image and assigns a label to each pixel in the image. The representative network of instance segmentation is Mask R-CNN. This network extends Fast R-CNN by adding branches to predict object masks in parallel with existing branches for bounding box recognition. In the ROI pooling extracts fixed-size windows from the feature map and uses these features to get the final class label and bounding box.

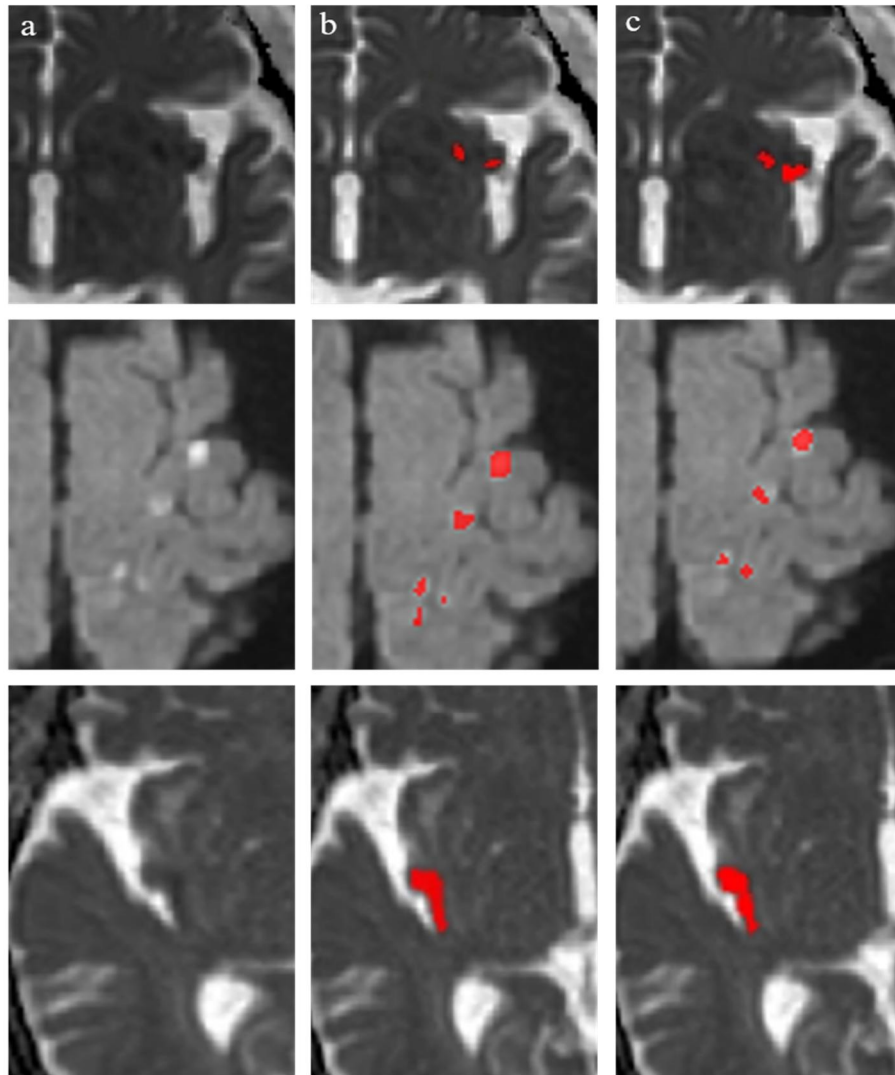
#### *Phase 2: Fine segmentation - Semantic segmentation using U-Net with SE block*

The SE block is provided with inputs such as the convolution block and the number of channels currently held. Using average pooling, each channel was compressed into a single numeric value. A fully connected layer followed by the ReLU feature added the necessary nonlinearity. Output channel complexity is reduced by a certain ratio. A fully connected second layer and S-shaped activation provide smooth gating on each channel. Finally, we weighted each functional map of the convolutional block-based on the results of the lateral network.

### *Results*

In the second stage, we have applied the variant network and losses. Table 3 shows the DSC results of applying various networks and losses in the second step. Overall, the DSC results of infarct segmetnation showed more than 80%, and the combined result of DSC and focal loss

was the highest at 86.6%. The U-Net with SE block achieved better segmentation than the corresponding network without SE block. In addition, the infarct segmentation result was higher when adc and b1000 were combined and trained than when adc and b1000 were trained respectively. Figure 3.5 shows an example of an acute ischemic stroke lesion segmentation learned using DSC and focal loss in the second stage.



**Figure 3.5** Examples of the results of infarct segmentation. (a) Original image, (b) Ground truth, (c) Results of the application of DSC+Focal loss network in the second stage.

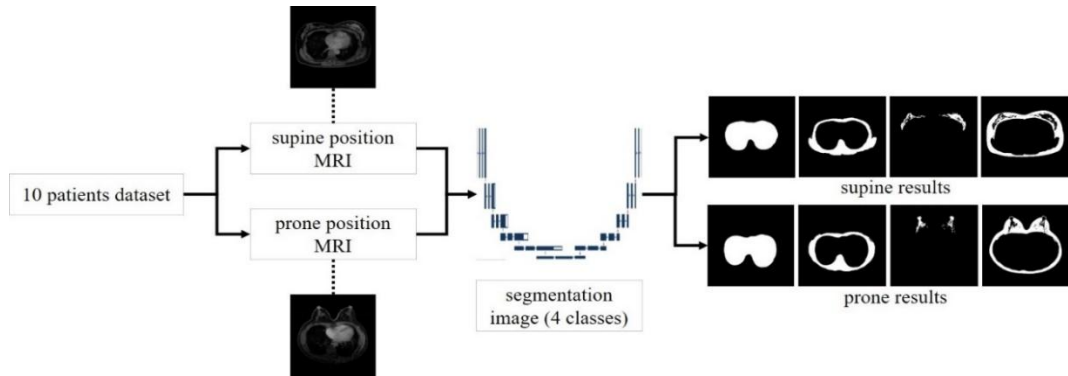
**Table 3.2** The DSC results of the application of various networks and losses in the second stage.

	<b>ADC + B1000</b>	<b>ADC</b>	<b>B1000</b>
U-Net (2D)	0.821 $\pm$ 0.045	0.801 $\pm$ 0.081	0.811 $\pm$ 0.082
U-Net with SE (2D)	0.833 $\pm$ 0.027	0.823 $\pm$ 0.063	0.821 $\pm$ 0.071
DSC + Boundary loss	0.858 $\pm$ 0.070	0.844 $\pm$ 0.071	0.837 $\pm$ 0.069
DSC + Focal loss	0.860 $\pm$ 0.077	0.845 $\pm$ 0.068	0.848 $\pm$ 0.062

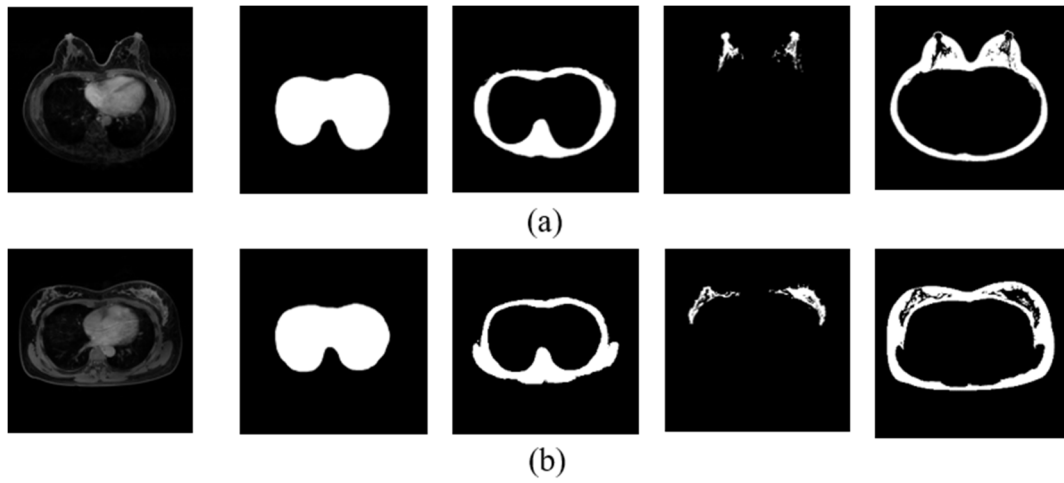
### 3.3 Supine-prone tissues in Breast MRI

Breast cancer is one of the most common cancers among women worldwide [54-56]. Early diagnosis and treatment have been proven to reduce the mortality rate [57]. In general, compared to mammography and ultrasonography (USG), MRI has been shown to have high sensitivity and high resolution for detecting primary lesions in the body [58, 59]. MRI is a non-invasive method and is preferred because it allows evaluation of preoperative staging and high-risk screening [60, 61].

Automatic segmentation of breast and surrounding tissue in MRI is a key step in the automated analysis for clinically relevant applications including CAD and registration considering material properties between prone and supine MRI for surgery. Manual segmentation of the MRI, however, is time-consuming and error-prone. MRI provides a high contrast among breast and surrounding tissues including lungs and heart, muscles and bones, parenchyma with cancer, and skin and fat all of which could have different material properties and deformations in supine and prone poses. Besides, the intensity non-uniformity of MRIs leads to major difficulties in this segmentation. Therefore, several specific systems have been developed to help radiologists or surgeons detect and segment breast lesions in supine MRIs from prone MRIs, greatly improving the clinician efficiency.

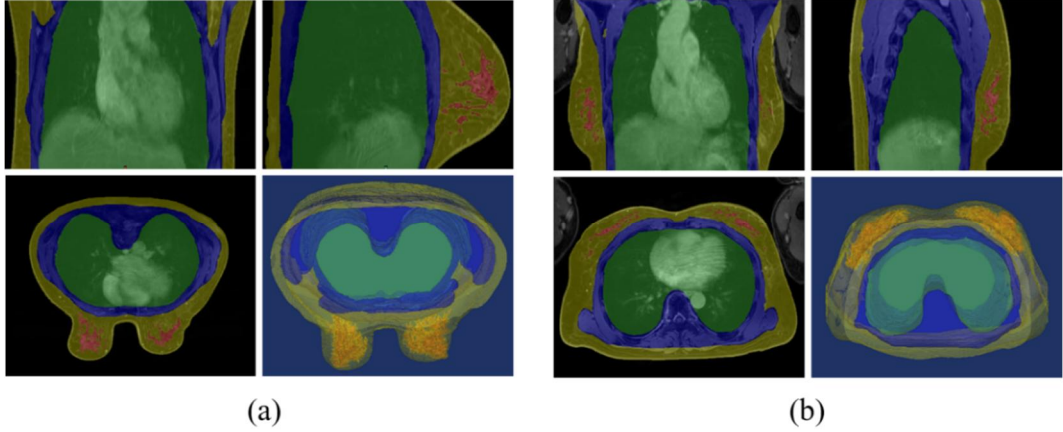


**Figure 3.6** Overall procedure of semantic segmentation in supine and prone breast MRI.



**Figure 3.7** Example multi-region labels. (a) Prone MRI, (b) Supine MRI.





**Figure 3.8** Example multi-planner reformatted views. (a) Prone MRI, (b) Supine MRI.

There have been several studies investigating breast segmentation in MRI. Niukkanen et al. [62] used k-means clustering for breast and fibro glandular tissue (FGT) segmentation based on MRI images. Nguyen et al. and Nie et al. [63, 64] developed an algorithm for semi-automatic segmentation using fuzzy c-means (FCM) clustering to identify breast FGT with breast cancer risk. To correct for the field inhomogeneity, they applied the bias field correction algorithm. Lin et al. [65] suggested a fully automatic segmentation using patient-specific chest template model mapping, which showed segmentation accuracy similar to that achieved by experts. Milenkovich et al. [66] reported a fully automatic method using edge maps obtained by applying a tunable Gabor filter, and they obtained 0.96 for the average DSC. These methods, however, are usually limited by the characteristics of the MR images used in the study datasets. Breast MRI varies for different contrast injection methods, MRI scan protocols, and MR conditions [67]. Even in a single hospital, this variability would be expected in MRI data across years, as protocols are changed from time to time due to the improvements in acquisition or MRI units. In addition to the variability of the MRI protocol, there are additional variabilities in terms of breast shapes, sizes, density, and chest muscle shape all of which could cause various deformations between prone and supine poses and MRI artifacts such as

inhomogeneous intensity, or alias effect and so on.

In recent years, it is to overcome these variabilities that deep CNNs have enabled a significant improvement in computer vision tasks such as image classification, object detection, and segmentation. Dalmis et al. [67] used deep learning segmentation methods of 2D and 3D U-Net architectures in three regions including non-breast, fat inside the breast, and intra-breast FGT. Their average DSC values were 0.933, 0.944, 0.863, and 0.848 for 3D U-Net, 2D U-Nets, atlas-based, and sheetness-based methods, respectively. X Zheng et al. [68] developed a coordinate-guided U-Net to identify breast boundaries on MRI by obtaining breast location information for segmentation. However, this study only focused on prone MRI because MRI scans in the supine position are quite difficult to be performed due to the low accuracy of tumor diagnosis and low contrast among surrounding tissues. Therefore, in an actual clinical setting, scans are usually performed in a prone position using a specialized breast coil and contrast injection protocol to increase the diagnostic accuracy of MRI. However, a surgical pose is supine, which may differ significantly from the prone position. Wang et al. [69] have shown how to move breast tumors three-dimensionally from the prone to a supine position in the operating room (OR). Further, given that prone MRI has a strong signal to noise ratio (SNR) of cancer tissues with adequate contrast agent protocol, and supine MRI has weak SNR, deformable registration between prone and supine MRIs has been required. However, because cancer and surrounding tissues of the breast are significantly changed between supine and prone position, the ordinary registration algorithm does not work. Therefore, fine registration based on the material properties of various breast tissues is needed to overcome the difficulty of registration between supine and prone MRIs. Therefore, in this study, we proposed a semantic segmentation to differentiate four regions with different material properties including lungs and heart, muscle and bone, parenchyma with cancer, and skin and fat in prone and supine breast MRIs with deep learning. This study aims to employ and compare various kinds

of networks and strategies with supine, prone, transferred from prone to supine and pooled supine and prone MRIs.

### *Dataset*

This study was approved by the IRB of the AMC (IRB No. 2017–1341) and was performed following the principles of the Declaration of Helsinki. The requirement for patients' informed consent was waived by the IRB. The imaging data were anonymized in accordance with the Health Insurance Portability and Accountability Act's privacy rules. The study was conducted using a total of the following two types of MRI scans for 29 patients: (1) Pre-prone MRI, prone position before neoadjuvant systemic therapy (NST) and (2) Pre-supine MRI, supine position before NST. The patient age range was 36–65 years, and the mean age was 47.2 years

### *MR Protocol*

Breast imaging was performed with a 3.0 T MRI system (Ingrain; Philips Healthcare, Netherlands) with a bilateral dedicated four-element breast coil. Patients underwent a routine standard MRI protocol performed in the prone position and then were repositioned in the supine position. A dynamic perfusion study with an intravenous injection of 0.1 mmol/kg of gadopentetate dimeglumine (MultiHance, Gd-BOPTA; Bracco Imaging SpA, Milan, Italy) was followed by a flush of 20 ml of saline solution at 2 ml/s. The dynamic study involved one pre-contrast acquisition followed by five post-contrast acquisitions of T1-weighted high-resolution volume examinations (THRIVE); TR = 4.1 ms; TE = 1.8 ms; slice thickness = 0.9 mm; pixel size = 0.9 x 0.9 mm). Immediately after the acquisition of the last dynamic series, the patient was extracted from the magnet, the breast coil was removed, and the patient was invited to assume the supine position. Thereafter, a thoracic four-channel surface coil was positioned over the breast surface. A multi-point Dixon (mDixon) sequence was used for

acquiring MRI in the supine position with the following technical parameters: TR/TE = 4.9/0.0 ms; fractional anisotropy (FA) = 10°; voxel size = 0.742 x 0.742 x 1.0 mm; and matrix = 512 x 512.

### *Gold standards*

All breasts (N = 116) in prone and supine MRIs were flipped into the right breast. The 4-class tissues including lungs and heart, muscles and bones, parenchyma with cancer, and skin and fat of prone and supine MRIs were manually delineated by a breast MRI specialist and confirmed by a surgeon. All labels were drawn using Mimics Medical 17 (Materialise Inc, Belgium), which is software for manual imaging segmentation. At first, we used thresholding and region growing with a cropped mask. The expert then modified it manually

### *Preprocessing*

The objective of preprocessing is to improve the quality of the image and make it ready for further processing by removing the irrelevant noise and unwanted portions from the background of the MRI. Even though breast images were acquired with an MRI scanner, image normalization was required to correct the intensity of the image. Then, we did the normalization by subtracting the average image intensity from every pixel in the image and dividing the pixel by the standard deviation (SD) of the intensities. Further, we used to vertically flip in the prone direction images to align them with the supine direction. Since the patient's two breasts are often symmetrical, we made it a single breast by cutting in half around the x-axis to create two images per patient to make the dividing problem easier and to have the effect of increasing the size of the training dataset. Additionally, after this division, the left breast was mirrored so that it was in the same direction as the right breast. Therefore, the width and height sizes of the input images were converted from 512 x 512 to 256 x 512. As another

way to increase the size of the training dataset, we used the data generator functions in Keras to randomly crop, flip, rotate, scale, zoom, add gaussian noise, and more on the training data.

#### *Semantic Segmentation Network with CNN - U-Net and nnU-Net*

U-Net architecture is a defined expression segmentation learning method with multiple levels of expression obtained by constructing simple but nonlinear modules that convert expressions from one to another. The name "U-Net" indicates the "U" shape of the network as shown in Figure 2.5. This is a fully convolutional network consisting of convolution and maximum pooling layers in descending order or, in other words, in the initial part of U. This part can be seen as a downsampling step because the input image size is divided into the maximum pull kernel size in each maximum pull layer. Upsampling is performed at the latter part of the network or at the rising part of U, which is implemented by a convolution where the kernel weights are learned during training. An arrow between the two parts of the "U" shows that the information available in the downsampling step is integrated into upsampling operations performed in the rising part of the network. Accordingly, the finely detailed information that is captured in the lowering part of the network is used in ascending and descending parts. The nnU-Net framework, which automatically adjusts the architecture to the given image shape, has been modified based on the U-Net architecture. The nnU-Net framework automatically defines several steps such as preprocessing, resampling, and normalization by setting loss optimization during training and post-processing operations. In this study, the U-Net and nnU-Net architectures were applied to both 2D and 3D data for segmentation involving breast and other tissues, and 4 labels including the background were learned simultaneously.

### *Statistical evaluation*

We performed a Willcoxon rank-sum test to compare the model results for prone, supine, and combined positions. To compare the performances of different segmentation models, we applied a paired t-test to the DSC, JSC, and HD values obtained for each MRI with a p-value less than 0.05 considered as significant.

### *Results*

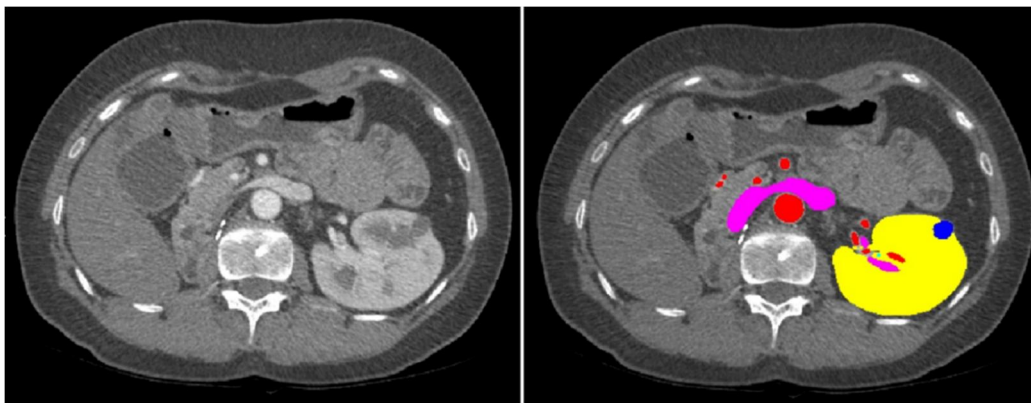
The DSC, JSC, and HD of the segmentation results of breast and surrounding tissues for each method segmentations are given in Table 1 and its supplementary respectively. Comparing the overall results, it was found that 2D U-Net with prone MRI showed the best performance in lungs and heart, muscles and bones, parenchyma with cancer, and skin and fat (mean  $\pm$  SD of DSC:  $0.987 \pm 0.003$ ,  $0.966 \pm 0.011$ ,  $0.870 \pm 0.085$ , and  $0.956 \pm 0.016$ , respectively). Comparing the overall results of different U-Net methods, 2D segmentation showed 1.1% results better compared to 3D segmentation. The parenchyma with the cancer label showed significant differences for each model. Additionally, the base U-Net architectures showed better performance than the nnU-Net architectures. The results of training with the data of the prone and supine postures combined showed lower results compared to the training results for each posture. Regarding the 4 class labels, the least segmentation performance was observed in parenchyma with cancer with an average DSC value of around 0.80.

## **3.4 Kidney substructures with Renal Cell Carcinoma (RCC) in kidney CT**

It is a prerequisite for CAD and provides quantitative information for treatment, surgical planning, and 3D printing in medicine. Recent advances in deep learning such as the emergence of FCNs have enabled the training of models for semantic segmentation tasks. In

particular, the 3D U-Net, which has a contracting path and a symmetric expanding path, has been proven to be effective for 3D medical image segmentation. Some authors have proposed novel cascaded architectures such as segmentation-by-detection networks and cascaded 3D FCN to improve segmentation performance using region proposal network before segmentation.

RCC is the most common of renal malignancies, and it is the eighth most common cancer overall with over 60,000 new cases and nearly 15,000 deaths in the United States each year. Most RCCs are asymptomatic and are accidentally detected when CT scans are performed for other indications. Kidney tumors vary widely in size, location, and depth. There has been much interest recently to study the outcome of partial or radical nephrectomy concerning the tumor morphology. To this end, accurate segmentation of kidney tumors in the imaging data is paramount for reliable disease classification and treatment planning. However, manual annotation processing is a difficult and time-consuming process and is sensitive to errors and inaccuracies. The most popular and promising method for general medical imaging computer vision in the past few years has been CNN.



**Figure 3.9** Kidney and RCC label image example.

CNNs are often learned manually from previously partitioned data. They have achieved state-of-the-art performance in organ and lesion detection, localization, segmentation and

classification. However, the specific work of renal and renal tumor segmentation, as previously mentioned, has addressed a high degree of variability in tumor location and morphology.

### *Dataset*

A total of 50 kidneys from 36 patients in abdominal computer tomography (CT)-scans (Sensation 16, Siemens Healthcare) were utilized—30 of them had RCC and 20 were normal kidneys, with each having a slice thickness of 1–1.25 mm. There were four phases in the CT scans—the non-contrast, renal cortical, renal parenchymal, and renal excretory phases. We used the renal cortical phase, which enhanced the arteries, and classified kidneys into five subclasses of an artery, vein, ureter, parenchyma with the medulla, and RCC for kidneys with RCC. We excluded those kidneys with cysts or stones. The IRB for human investigations at AMC approved the retrospective study and waived the need to acquire informed consent. The imaging data were de-identified in accordance with the privacy rule of the Health Insurance Portability and Accountability Act.

### *Experiment*

First, we trained the model using a cascaded 3D U-Net with exceedingly small amounts of training data, and corresponding GTs were generated through manual labeling at the initial stage. The cascaded architecture was designed to improve segmentation performance using RPN before segmentation within the available memory of the GPU. Second, the results of the additional data through the trained network were manually corrected instead of creating new GTs from scratch. This step is known as convolutional neural network-corrected segmentation. Third, all of the initially used and newly added data were used again for subsequent training.



### *Cascaded 3D U-Net.*

We used a cascaded 3D U-Net architecture that replaced the first RPN in the study by Tang et al. into 3D U-Net.<sup>7</sup> This shows superior accuracy in the detection of the ROI in the kidney through the abdominal CT scans. The 3D U-Net technique can be divided into two main sections: the left side, which reduces the number of dimensions, and the right side, which extends to the original number of dimensions. The two sides consist of convolution and upsampling or downsampling layers. The downsampling layer was implemented through max-pooling (3 x 3 x 3). The prominent feature of 3D U-Net is its concatenation function to the left and the right. The concatenation results lead to an improved segmentation through the prevention of the loss of information.

Cascaded 3D U-Net was separately trained in an end-to-end manner. The ROI was marked as a cuboidal-bounding box around the kidney after the first U-Net module. Subsequently, the second U-Net module for final segmentation was trained to make masks for five subclasses of the kidney. Gaussian noise was added to each input image. Moreover, the errors were calculated using the DSC, like in equation (1). The loss function, which is denoted by dice loss (DL), was defined as equation (2) in each 3D U-Net.  $V_{GT}$  and  $V_{CNN}$  were defined as the volume of GT and CNN segmentation, respectively.

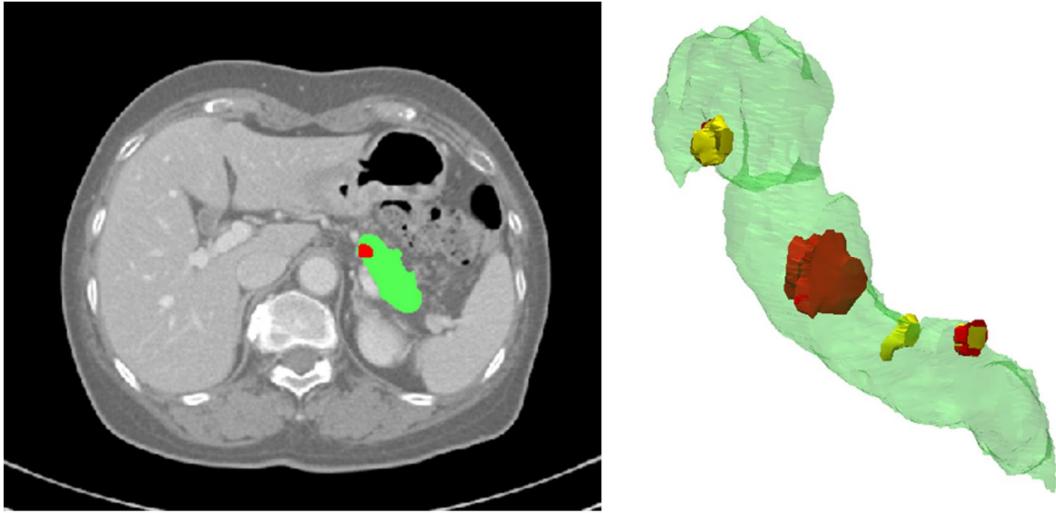
### *Experimental settings.*

The model was executed in Keras 2.2.4 with the backend of Tensorflow 1.14.0 and trained with a GPU of NVIDIA GTX 1080 Ti. Usually, in both the steps of our cascade method, a large number of epochs are required. In the first stage, the training was saturated at about 150 epochs due to the small number of datasets ( $N = 20$ ). The second and third stages required 300 epochs due to the increased numbers of datasets ( $N = 40$  and  $N = 50$ ). In addition, Adam optimizer with a learning rate of  $10^{-5}$ , a weight decay of 0.0005, a momentum of 0.9, the

training loss as the average dice coefficient loss, and a batch size of 1 was used. For testing the overfitting of the model, the difference of the overall DSC accuracies between the validation and test datasets of the final model was 6.17, which demonstrates that this model is not overfitting.

### **3.5 Pancreatic cancer in Pancreas CT**

Pancreatic cancer is one of the common malignancies whose incidence is increasing worldwide every year. According to the most recent statistics, in 2017, the estimated number of pancreatic cancer deaths in the United States was 43,090, with malignant tumor mortality in the fourth position. In the ‘China Cancer Registry Annual Report’, in China, pancreatic cancer was ranked sixth in the mortality rate of malignant tumor patients in the same year. With the current growth in the incidence of pancreatic cancer worldwide, the need to lower the mortality rate and improve the survival rate of pancreatic cancer patients has also increased, along with enhancements of diagnostic capabilities and treatment techniques. The segmentation of the pancreas can be used meaningfully to support doctors’ diagnosis, treatment, and surgery. Automatic and reliable segmentation of the pancreas is an important but challenging task for numerous clinical applications, such as pancreatic cancer radiation therapy and computer-aided diagnostics (CAD). The main challenge with accurate CT pancreas segmentation lies in the following two aspects: (1) Large shape change of the pancreas in various patients and (2) low contrast and blur around the pancreatic border. In this paper, we have proposed using deep learning to automatically segment the pancreas and compare its performance.



**Figure 3.10** Pancreas and focal lesions image example.

#### *Dataset*

For this study, we obtained a dataset of CT images of 862 patients, 360 with solid lesions like cancer and 142 with cystic lesions. The diagnostic decisions made for these data, based on a combination of clinico-radiologico-pathological discussions and consensus opinions, were regarded as gold standards by experts.

#### *CT protocol*

CT imaging was performed using a Somatom 64 scanner (Siemens AG, Healthcare Division, Erlangen, Germany) with a consideration of the following parameters: craniocaudal abdominal scan – 120 kV, pitch – 0.9, collimation – 0.6 mm, inter-slice spacing – 5 mm, and soft recon kernel. CT images were acquired at the portal venous contrast agent phase (with an intravenous application of weight-adopted and warm Imeron® 400 (Bracco Imaging, Konstanz, Germany), followed by a saline flush with a flow rate of 3 ml/s through a 20-gauge catheter inside an antecubital vein.

### *Manual Segmentation*

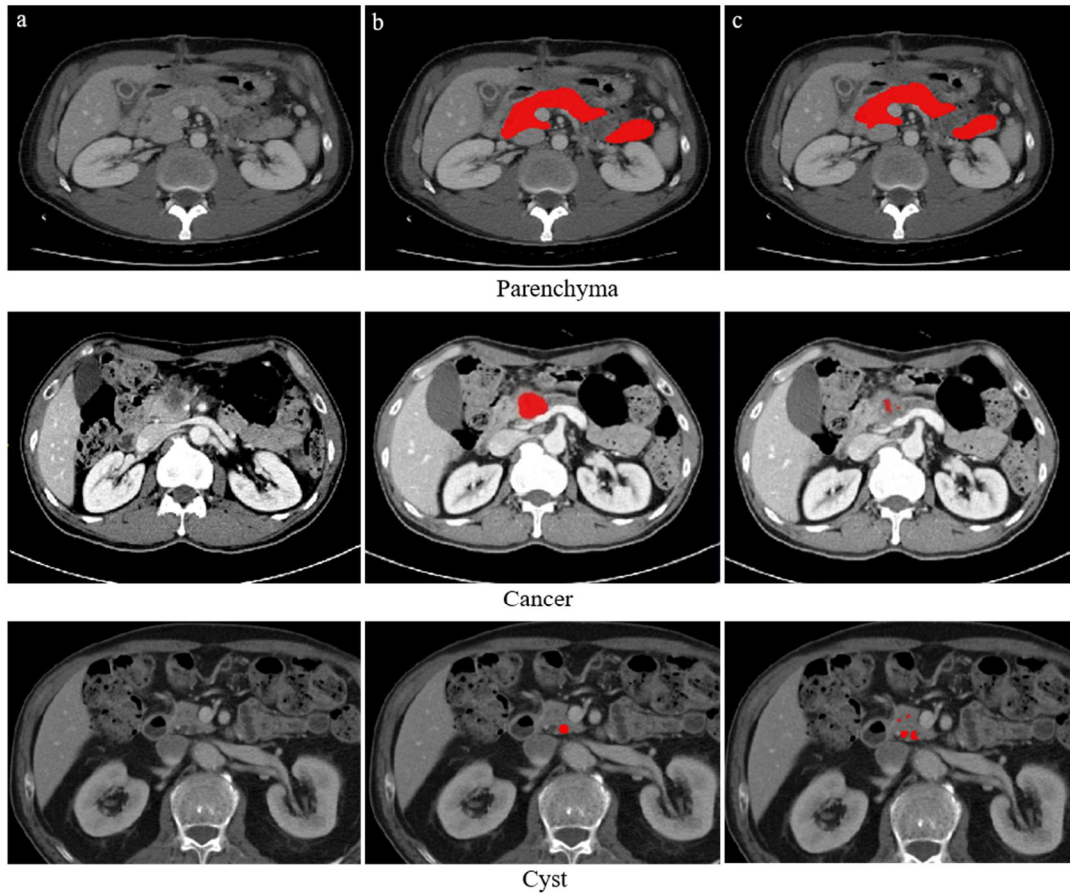
The radiologist (a board-certified radiologist with 10 years of experience (H.J.K)) collected the CT data and manually segmented the pancreas using the CT images in consensus. These delineations were taken as the standard in all tests.

### *Method*

We presented a completely 3D- and 2D-cascaded frameworks for pancreas segmentation in the 3D CT images. We developed a 3D detection network (U-Net with squeeze and excitation blocks) to regress the locations of the pancreas regions; in the second step, we implemented a fine segmentation using a 2D segmentation network to fragment the pancreas in a cascaded manner based on the detection results.

### *Result*

In general, applying 2D-based methods to 3D data can give rise to various problems. If the model cannot be generalized to the implemented for slices that are severely disconnected along the  $z$ -axis, the focal lesions of the pancreas volume boundary in that direction will most likely be inaccurate, resulting in critical errors. To test our model's generalizability, we created an independent set of volumetric CT images. Figure 3.11 shows the segmentation result of parenchyma and focal lesion like cancer and cyst in pancreas. We calculated the accuracies of the result of the pancreas segmentation and focal lesions for abdominal CT. The DSC of pancreas segmentation results was over 90%. However, the focal lesions of the pancreas were low DSC at around 60%.



**Figure 3.11** Examples of parenchyma and focal lesions segmentation. (a) Original image, (b) Ground truth (c) Segmentation results

### 3.6 Multi-structures in dental CBCT

Cone-beam computed tomography (CBCT) has been adopted over the past decades at a rapid pace and is widely used in dent maxillofacial imaging and orthodontic practice. Most importantly, the current effective dose of CBCT for head scans (from tens to hundreds of  $\mu\text{Sv}$ ) is much lower than that of CT for head scans (from hundreds to thousands of  $\mu\text{Sv}$ ). Additionally, the reported spatial resolution (voxel resolution) varies from 0.076 mm to 0.4 mm. In CT, the spatial resolution of the slice can be as small as 0.4 mm, with the thinnest axial thickness being 0.625 mm. On the other hand, CBCT offers special access to most medical

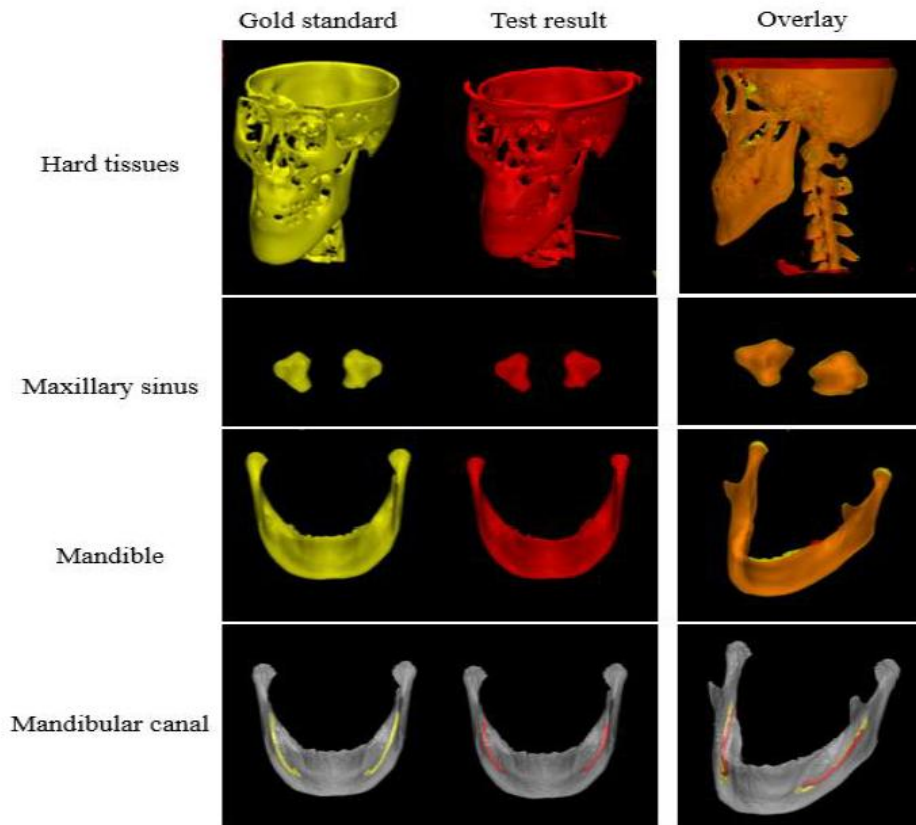
devices since it is less expensive than CT. Commercially available CBCT requires patients to sit or stand vertically during the scan. A natural head position can be acquired directly for 3D head measurement. These advantages make it possible for CBCT to replace CT for 3D imaging and CASS modeling. Segmentation of multi-facial structures is essential for dental implantations and orthognathic surgeries to draw safety margins around the facial nerves and surgical lines in the facial bones. However, the segmentation of multi-facial structures is extremely challenging due to the structural irregularities, complex forms, and heterogeneity of imaging contrast in the voxel-by-voxel approach, especially CBCT.

### *Method*

We obtained 100 datasets of i-Cat CBCT scans (Imaging Science International, Hatfield, PA, USA). In addition, the training and validation datasets of hard tissues, maxillary sinus, and mandible include 7 and 4 cases, 20 and 4 cases, and 20 and 4 cases, respectively. The training and validation datasets of mandibular canals consist of 40 and 15 cases, respectively. Each test dataset was used for all seven patients. Depending on the structures, the segmentation result was made as per the gold standard differently. The hard tissues and mandibular canals were manually drawn by an expert and confirmed by an expert dentist. The initial mandible and maxillary sinus masks were created by in-house software with conventional image-processing techniques, including 3D sculpting and thresholding. These masks were filled per slice and manually corrected by experts. Before commencing with the training, we pre-processed the CT images to change their brightness levels (window level = 450) and contrast settings (window width = 5000). Furthermore, the number of each dataset was augmented by flip and rotation. We used 3D U-Net, one of the most widely used CNN architectures, for image segmentation.

### *Result*

For the evaluation metrics, the DSC, JSC, MSD, and HD were  $82.60 \pm 0.04\%$ ,  $70.47 \pm 0.06\%$  in hard tissues,  $86.60 \pm 0.01\%$  and  $92.80 \pm 0.01\%$ , and  $82.00 \pm 0.07\%$ ,  $0.35 \pm 0.29\text{mm}$ ,  $0.49 \pm 0.09\text{mm}$  in maxillary sinus,  $90.00 \pm 0.04\%$ ,  $0.89 \pm 0.66\%$ ,  $11.29 \pm 10.21\text{mm}$ ,  $3.21 \pm 2.56\text{mm}$  and  $18.42 \pm 19.72\text{mm}$  in mandible, respectively. In the mandibular canals, mean error distances were  $0.89 \pm 0.40\text{mm}$ .



**Figure 3.12** Multi-structures segmentation examples

## 4 Advanced issues in medical image segmentation

### 4.1 Strategy for medical image segmentation

We applied various kinds of deep learning based segmentation, including U-Net, U-Net with SE, nnU-Net, cascaded U-NET to various lesions, and organs, including glioblastoma, infarct, breast, kidney, pancreas, and dental multi-structures with/without contrast agent depending on the clinical context. These are summarized in Table 3.3. These automated segmentation methods were taught taking into account all the characteristics of each lesion, organ, and the clinical context in cross-sectional imaging modalities such as CT and MRI.

Since there are various kinds of configurations for training a deep learning based model,



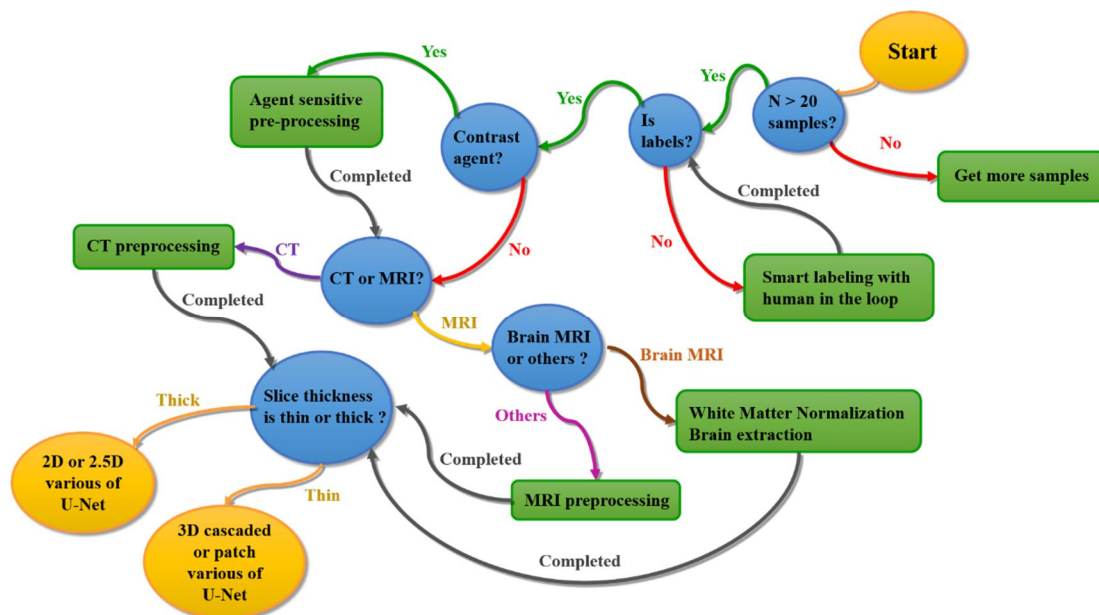
several types of pre-processing techniques were applied depending on the characteristics of each lesion, organ, and the clinical context. Especially, in brain MRIs, specialized pre-processing methods implement a bias-field correction caused by inhomogeneity of MRI B0 field, brain extraction to segment brain regions, and intensity normalization to reduce intra- and inter-image intensity variability. Otherwise, in CT images, only intensity normalization was performed as a pre-processing method since CT density is already calibrated in general. In addition, deep learning based models suffer from small amounts of datasets. Therefore, various augmentation methods, including random rotation, flipping, resizing, and others were performed to make variations in the dataset in order to train more robust deep learning based models. The details of augmentation methods were chosen and controlled depending on the characteristics of each lesion, organ, and the clinical context.

The number of subjects depends on the difficulty of labeling and the possibility of disease incidence. Therefore, the number of scans with thin-sectional thickness with multi-labels needed in kidney and breast cases is very small. Slice thickness was varied depending on the variations of clinical protocols. Due to variation in its size, shape, location, image reconstruction protocol, and modality, medical-image segmentation is considered to be one of the most difficult tasks.

Considering all these factors, the deep learning based automated segmentation (DLAS) model should be chosen, trained, and evaluated. Heterogeneous and ambiguous images, such as those of glioblastoma from heterogeneous tumor regions, were trained through a patch-level approach. In addition, for small datasets with a tube-like or line shape, it may be a good idea to perform segmentation using a 2D-based method because it is better to train models with balanced datasets. When the slice thickness is dense, the learning method with 2D-based methods would have higher accuracies compared to 3D-based methods as a result of less anatomical integrity between the slices and a relatively larger number of training datasets. If

the boundary of the lesion is ambiguous or its size is small, the focal loss could improve the learning accuracy. In most cases, SE blocks could increase segmentation accuracy since it can preserve important information across channels. In the case of large size images and small target objects, the cascade method is recommended for increasing the accuracy due to the fine structure, with a high resolution, of the candidate area acquired at the first segmentation stage. These insights have been summarized in Figure ??.

Based on the results and the raised issues, these studies were extended to more advanced studies such as studies focusing on more balanced segmentation with different level labels, robust feature extraction of radiomics, smart labeling with a human in the loop, etc.



**Figure 4.1** Flowchart for strategy in medical image segmentation.

**Table 4.1** Summary of organ specific segmentation.

	<b>Glioblastoma</b>	<b>Infarct</b>	<b>Breast</b>	<b>Kidney</b>	<b>Pancreas</b>	<b>Dental multi-structures</b>
<b>Subjects (Scan)</b>	103 (103)	424	29 (58)	50	862	50
<b>Modality</b>	MRI(T1CE,FLAIR)	MRI(ADC,DWI)	MRI (Supine& prone, T1CE)	CT(Arterial phase)	CT(portal phase)	CBCT
<b>Contrast agent</b>	Enhanced	Non-enhanced	Enhanced	Enhanced	Enhanced	Non-enhanced
<b>Pre-processing</b>	Biasfield correction Normalization* Brain extraction	Bias field correction Normalization* Brain extraction	Normalization* Align to the supine direction	Normalization*	Normalization*	Normalization*
<b>Augmentation</b>	Random rotation, flipping, resizing, etc.	Random rotation, flipping, resizing, etc.	Random rotation, flipping**, resizing, etc.	Random rotation, flipping**, resizing, etc.	Random rotation, flipping, resizing, etc.	Random rotation, flipping***, resizing, etc.
<b>Slice thickness</b>	0.5–1mm	3–7mm	0.9mm	1–1.25mm	3–6mm	0.3–1mm
<b>CNN</b>	3D U-Net with SE	2D U-Net SE vs etc.	2D/3D U-Net nnU-Net	Cascaded 3D U-Net	Cascaded 2D/3D U-Net	3D U-Net, Cascaded 3D U-Net
<b>Loss</b>	DSC	DSC, boundary DSC, focal	DSC	DSC	DSC, focal	DSC
<b>External dataset</b>	BraTS 2017				KHUH	
<b>DSC(internal)</b>	0.81±0.11	0.86±0.19	0.95±0.08	0.71±0.21	0.76±0.13	0.91±0.07
<b>DSC(external)</b>	0.79±0.18				0.71±0.21	
<b>Issues</b>	Domain adaptation, ambiguous boundary	Thickness, scale-invariant due to disease progression, ambiguous boundary	Small dataset, variability of contrast agent uptake	Memory, high-cost labeling, small dataset	Thickness, domain adaptation, different pathologic lesions	Memory, high-cost labeling, small dataset
<b>Advanced study</b>	Robust feature extraction of radiomics	More balanced segmentation with different level labels	Domain adaptation	Smart labeling with a human in the loop	Domain adaptation, computer-aided diagnosis	Smart labeling with a human in the loop

*Note: \*mean and standard deviation normalization, \*\*(breast and kidney) flip to the right side, \*\*\*(mandible and mandibular canals) flip to the right side; DSC, Dice similarity coefficient; SE, Squeeze and excitation; BraTS, Brain tumor segmentation; KHUH, Kyung-Hee University Hospital*

## 4.2 Smart labeling with human in the loop

In stage I, five subclasses of 20 kidneys, including artery, vein, ureter, parenchyma, and RCC were manually delineated as GTs for initial training. After this stage, the GTs of the new data for the next stage was prepared by executing manual correction of the results from the CNN segmentation, which is referred to as CNN-corrected segmentation. In stage II, 16 kidneys from the previous stage were reused for training, along with new data pertaining to eight kidneys with RCC and eight normal kidneys. After stage II, the results of CNN segmentation for the new data were manually amended for the next stage, as was done in stage I. Finally, in stage III, 40 kidneys were used for training, while 10 kidneys were used for testing. The results of all the aforementioned stages were used to evaluate for accuracy. The manual and CNN-corrected segmentations were conducted using Mimics software (Mimics; Materialise, Leuven, Belgium).

To determine whether the performance of the network gets improved or not through active learning, we investigated the DSC at each stage and compared them with one another using the paired t-tests between stages 1 and 3 and stages 2 and 3 using the SPSS software (version 25.00; IBM). In addition, to evaluate the effect of the proposed method, we compared it with that of a more recent network—no-new-U-Net (nnU-Net) introduced by Isensee et al. This network won first place in the Kidney Tumor Segmentation Challenge (KiTS19) on Medical Image Computing and Computer-Assisted Intervention Society (MICCAI) 2019. We also validated the CNN-corrected segmentation for accuracy and total consumption time to evaluate labeling efficiency. Moreover, we converted the results of manual, CNN, and CNN-corrected segmentation to 3D models to compare their accuracies. The comparison was performed based on the points in the surface using quantitative root-mean-square (RMS) values in the 3-matic software (3-matic; Materialise, Leuven, Belgium). A total of 17,650 points were used to compare 3D models among manual and CNN segmentation and manual

and CNN-corrected segmentations. For comparison of CNN segmentation with CNN-corrected segmentation, 26,471 points were calculated. The calculation for RMS is the same as that in equation (3), where  $x$  is the difference between corresponding points in the two models, and  $n$  is the total number of points.

The results of the comparison of segmentation time for the five substructures between manual and CNN-corrected segmentation have been listed in Table 3. CNN-corrected segmentation decreased the time of artery segmentation by 19 minutes 8 seconds and that of the vein, ureter, parenchyma, and RCC by 12 minutes 1 second, 19 minutes 23 seconds, 8 minutes 20 seconds, and 17 minutes 8 seconds, respectively, with an overall segmentation time reduction of 76 minutes, which is more than half of the time required in manual segmentation. Apart from the time taken for the initial loading of the package, the CNN segmentation took less than 1 second per case. The results of CNN-corrected segmentation are observed to highly correspond with those of manual segmentation, while they do not with those of CNN segmentation.

#### *Evaluation and statistical analysis*

The average values of DSC for the five subclasses became increased with each stage's completion. Among the aforementioned subclasses, parenchyma segmentation has the highest DSC and the lowest SD values, while for RCC, the DSC value was the lowest, while the SD value was the highest. In addition, the final segmentation results at the last stage were found to be superior to those of the nnU-Net using our dataset.

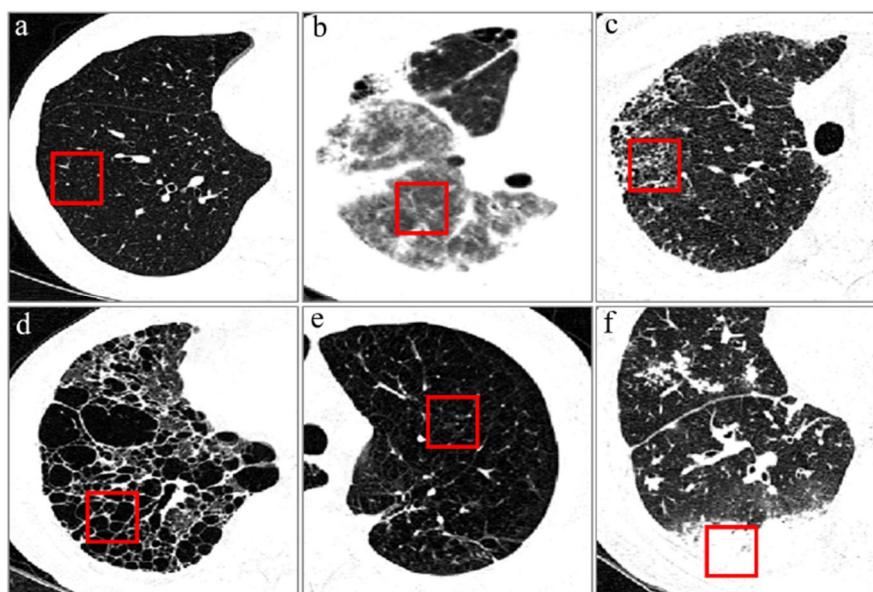
### **4.3 Fine-tuning with different level labels in imbalanced datasets**

Deep learning based automatic segmentation (DLAS) supports the reproducibility of radioactive functions, but its effect on radioactivity modeling is unknown, so DLAS extracted powerful anatomical and physiological MRI functions to evaluate whether it can accurately

assess the treatment response of glioblastoma patients.

### *Dataset*

The IRB of AMC for human investigations approved the study protocol, removed all patient identifiers, and waived informed consent requirements due to the retrospective design of this study. At AMC, patients with various types of DILD, including cryptogenic organizing pneumonia (COP, 218), usual interstitial pneumonia (UIP, 196), and nonspecific interstitial pneumonia (NSIP, 498), were retrospectively enrolled and scanned with the Siemens CT scanner (Sensation 16, Siemens Medical Solutions, Forchheim, Germany) for the HRCT and volumetric CT. All the CT images were obtained using typical HRCT protocol parameters, such as slice thicknesses of 1–2 mm, intervals of 5–10 mm, 220 mAs, and 120–140 kVp, with an enhanced reconstruction kernel (B70f in the Siemens scanner and a sharp kernel in the GE scanner), and a considering of volumetric CT protocol parameters, including a sub-millimeter thickness without intervals. The radiographer had instructed the patients to hold their breaths while conducting the CT.



**Figure 4.2** Typical ROIs in six image patterns of diffuse interstitial lung disease. ((a) Normal parenchyma, (b) Ground-glass opacity, (c) Reticular opacity, (d) Honeycombing, (e)

Emphysema, and (f) Consolidation.)

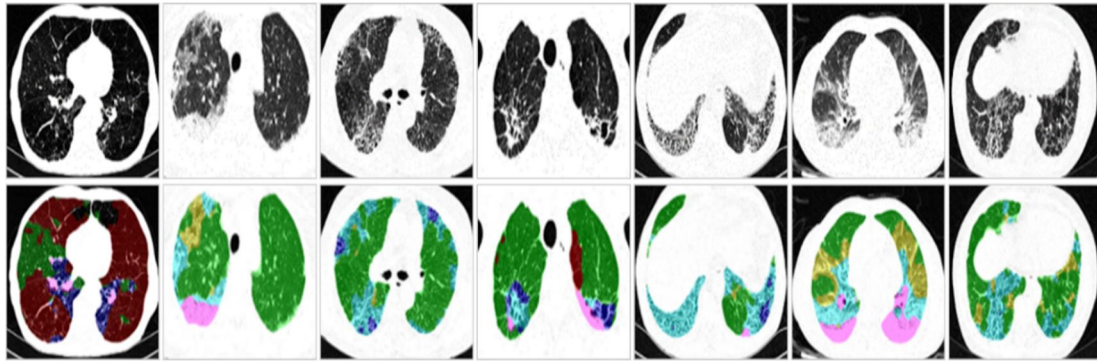
*The gold standard of patch-level for six diseased image patterns and image-level manual drawings in HRCT*

The following six types of image characteristics were evaluated: normal parenchyma (NL) (Figure 4.2.a), ground-glass opacity (GGO) (Figure 4.2.b), reticular opacity (RO) (Figure 4.2.c), honeycombing (HC) (Figure 4.2.d), emphysema (EMPH) (Figure 4.2.e), and consolidation (CONS) (Figure 4.2.f). Ground-glass opacity is characterized by an abnormally hazy focus on the lungs that is not associated with obscured underlying vessels. A similar observation associated with obscured underlying vessels is termed as consolidation. Increased reticular lung opacity is the product of a thickened interstitial fiber network of the lung that results from fluid, fibrous tissue, or cellular infiltration. In emphysema, there are focal areas of very low attenuation that can be conveniently contrasted with the surrounding normal parenchyma of higher attenuation. Emphysema can typically be distinguished from honeycombing by the fact that while the former has areas of emphysematous destruction that lack a visible wall, in the latter, honeycombing cysts have thick walls of fibrous tissue. Honeycombing is also characterized by extensive fibrosis with lung destruction that results in a cystic reticular appearance. However, there is still a problematic middle area where more than two regional characteristics are displayed simultaneously or when a regional characteristic is too ambiguous for radiologists to arrive at an agreement. One hundred patches with  $30 \times 30$  pixels of a typical image pattern were marked independently at AMC and NJHC. A total of 1200 ROIs were labeled by patch levels. A patch-level conventional classifier using an SVM was developed [70-72] and evaluated to classify each disease according to its subtype.

These patch-level image patterns were used to train a conventional classifier using an SVM, a conventional machine-learning technique with human-engineered features, including texture and shape features [70]. This classifier was used to categorize pixel-by-pixel DILD image patterns in the 1200 DILD scans of the whole lung area in an automated manner. Figure 4.3

presents examples of HRCT images and their corresponding classification results on six typical image patterns of DILD as per the gold standard.

In addition, to evaluate for image-label accuracy, 92 independent HRCT images were selected randomly from the HRCT scans of DILD patients. These were manually delineated by two expert thoracic radiologists with over 10 and 20 years of experience. Fig. 4.2 provides examples of HRCT images and their corresponding manual drawing results of six image regions by the two radiologists included in this study. Evidently, there are significant differences in the manual drawing results for the two expert radiologists.



**Figure 4.3** Examples of HRCT images (upper row) and their corresponding classification results on typical six image patterns of DILD by the gold standard (lower row). (Normal, green; ground-glass opacity, yellow; reticular opacity, red; honeycombing, light blue; emphysema, dark blue; consolidation, pink)

#### *Pre-processing*

As there are different therapies and prognoses for DILDs depending on their type, it is important to distinguish these types. Segmenting a diseased organ is critical while analyzing the images of the organ using a CAD model. This step generally precedes the main image analysis [70, 73]. An incorrect setting of the organ's borders due to segmentation errors is highly likely to affect the subsequent analysis. The automation of organ segmentation is a challenging task as it might be difficult to identify the lung borders of patients with pulmonary diseases as these diseases further reduce the distinction between the lung tissue and the



surrounding structures. We have developed an accurate and robust U-Net-based DILD lung segmentation method in this study, which was used to segment the lung region of additional 1200 DILD slices with different clinical protocols, including HRCT and volumetric CT.

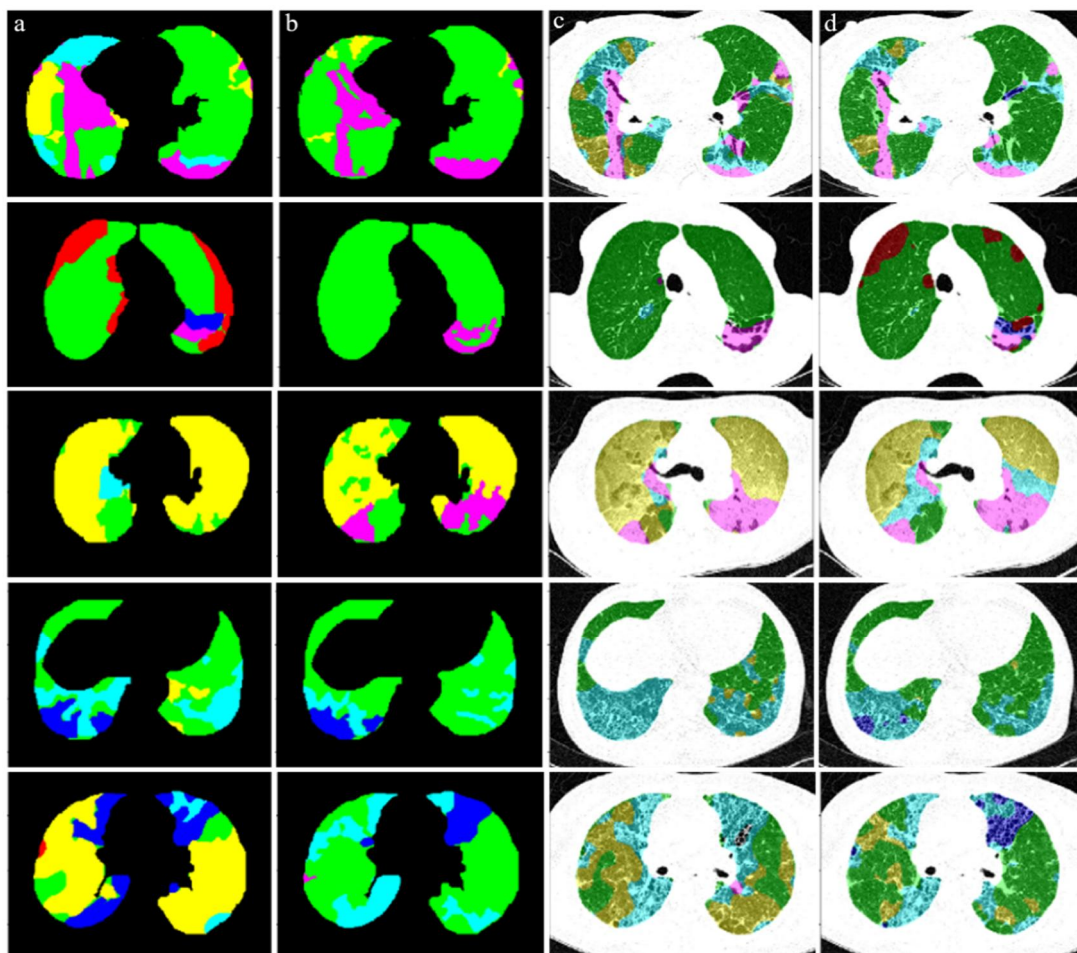
Although chest CT images consist of distinct 3D spatial data, 3D approaches such as 3D U-Net or V-Net may be inappropriate as HRCT images have variable intervals and weak connectivity. Therefore, we extracted randomly selected axial images from the training dataset and evaluated the test data by stacking adjacent 2D results. To ensure a fair and accurate evaluation, the data were divided into training, validation, and test sets on a per-patient basis, and the comparisons were performed using a test set that was not used during the training process. Of the total 1200 HRCT scans, 80%, 10%, and 10% scans were used for training, validation, and test, respectively. The training data consisted of 17,857 axial images, which were randomly shuffled for each epoch.

During the training, each input image was transformed by adding Gaussian noise with an SD of 0.1 and rotating it randomly by  $-10^\circ$  to  $+10^\circ$ , zooming it by 0–20%, and flipping it horizontally for regularization. The model was implemented using Keras (2.0.4) with Theano (0.9.0) backend in Python 2.7. Adam optimizer<sup>19</sup> was used for stochastic gradient descent with learning rates of  $10^{-5}$ . The DSC was used as a loss function and was calculated in mini-batch units with at least one sample containing lungs. This is because the DSC may display different behaviors for positive samples with lung and negative samples without lungs.

#### *Fine-tuned model*

The image patterns of the classification results generated by the SVM classifier are severely imbalanced because the ratios of image patterns generally depend on the severity and progress of the DILD. Therefore, the semantic segmentation network that was trained with these image pattern maps could display an imbalance of image patterns, a bias of the SVM classifier, the progress of DILD, and others. Therefore, we fine-tuned the U-Net-based semantic

segmentation network with the image-level labels of 92 HRCT images by two radiologists (total 184 images). These datasets were split equally and randomly between the training set and test set. Images from the same patient were not used in both the training set and test set simultaneously.



**Figure 4.4** Examples of HRCT images with manual drawings of reader 1 and reader 2 and the results of semantic segmentation before fine-tuning and after fine-tuning in the same. Gold standard masks are drawn by (a) Reader 1 and (b) Reader 2 and semantic segmentation results (c) without fine-tuning and (d) with fine-tuning. (Normal, green; ground-glass opacity, yellow; reticular opacity, red; honeycombing, light blue; emphysema, dark blue; consolidation, pink)

## Results

### *Semantic segmentation of six image patterns*

The agreement between the two radiologists regarding the test set of image-level labels was

0.71 in terms of the DSC. The agreements between the SVM classifier and the two radiologists were 0.59 and 0.62, respectively, which were lower than those between the two radiologists. This could have been caused by the limitation of the patch-level labels and the imbalance of the sub-classes. The U-Net-based semantic segmentation was trained based on the classification results of the SVM classifier. The agreements between the U-Net-based semantic segmentation and the two radiologists were 0.57 and 0.54, respectively, which were lower than those between the two radiologists and the SVM classifier. This could have been also caused by the limitation of the image-level imbalance of the sub-classes.

#### *Fine-tuning the semantic segmentation with image-level gold standards*

The U-Net-based semantic segmentation was fine-tuned using image-level labels from 46 HRCT images of the training set. The agreements between the fine-tuned semantic segmentation and the radiologists were 0.66 and 0.72, respectively. These agreements were significantly higher than those of the SVM classifier and the non-fine-tuned U-Net-based semantic segmentation. In addition, these agreements are similar to those between the two radiologists. Fig. 4.3 presents typical examples of HRCT images, manual drawing of reader 1 and reader 2, and the results of semantic segmentation before fine-tuning and after fine-tuning of the same HRCT images.

#### *Statistical analysis*

Table 4.1 presents the ratio of the six classes evaluated by the two radiologists. The difference between the two readers' RO patterns was the largest, with reader 1 being 0.061 and reader 2 0.235. We performed the non-parametric Wilcoxon signed-rank test to determine the evaluation statistical significance between the models.

**Table 4.2** Ratios of six image patterns

Pattern	NL	GGO	RO	EMPH	CONS	HC
Reader 1	0.647	0.034	0.061	0.042	0.188	0.028
Reader 2	0.453	0.048	0.235	0.057	0.187	0.020

In Table 4.2 the performance of the fine-tuned model in terms of six pattern classification is significantly higher than that of the before-fine-tuned model. As presented in this table, the fine-tuned model achieved the highest value of DSC 0.72.

**Table 4.3** Agreement among radiologists, SVM, and deep learning with and/or without fine-tuning

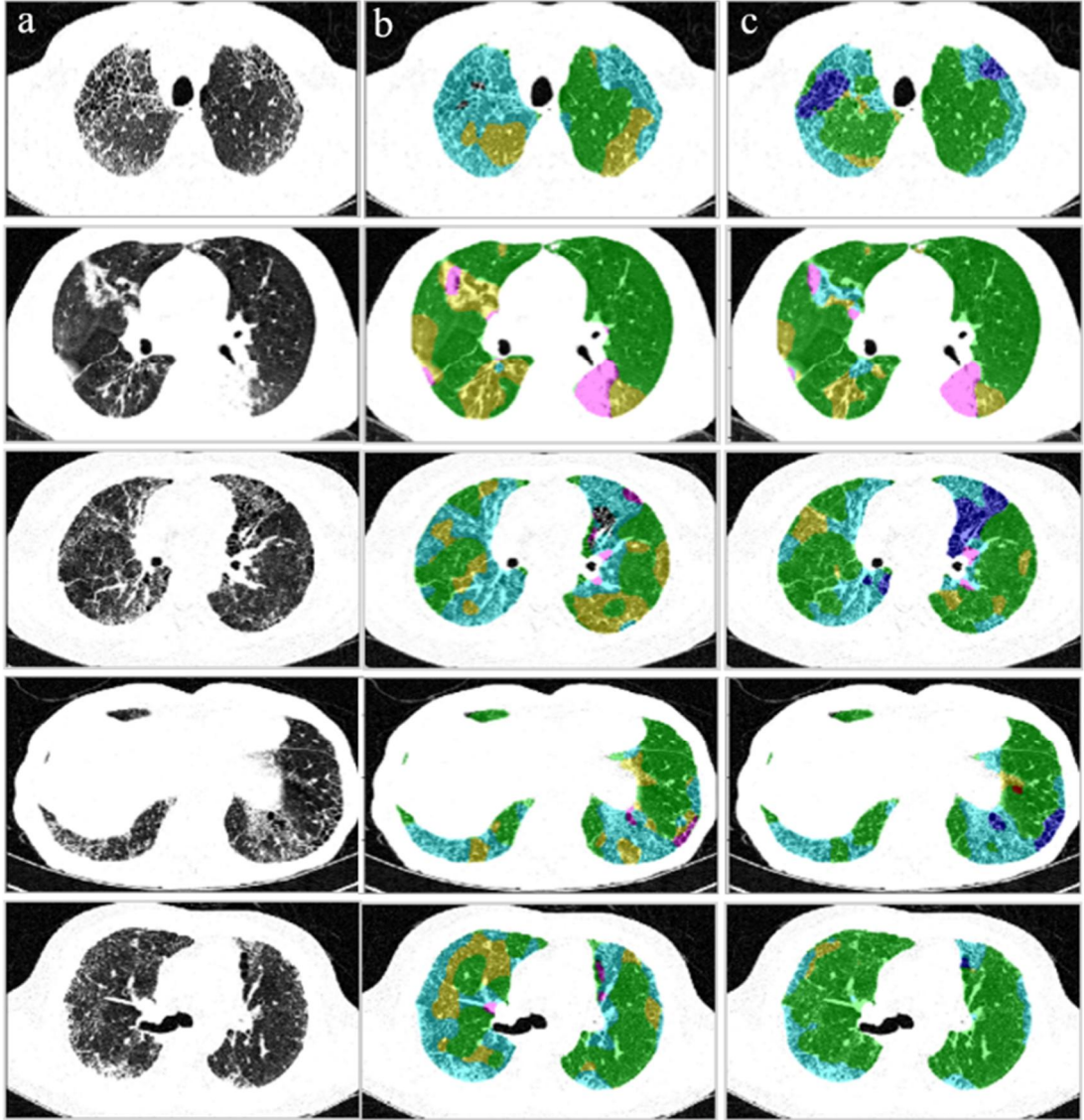
	Agreement of two readers	SVM	DL1	DL2
Reader 1	0.71	$0.523 \pm 0.146$	$0.531 \pm 0.211^*$	$0.643 \pm 0.105^{***}$
Reader 2		$0.540 \pm 0.144$	$0.664 \pm 0.185^*$	$0.721 \pm 0.124^{***}$
Time(min)	10~20	$1.431 \pm 1.432$	$0.055 \pm 0.005$	$0.051 \pm 0.004$

Table 4.3 presents the ratios of each image pattern determined with the help of the fully automated fine-tuned semantic segmentation model in three typical types of DILD: COP, UIP, and NSIP, having different clinical characteristics, treatment strategies, and survivals. In the HRCT images of COP patients, the consolidation ratio is seen to be dominant. In the case of UIP, honeycombing covers the significant regions of the HRCT of the lung.

**Table 4.4** Ratios of each image patterns by the fine-tuning model in three typical DILDs

	NL	GGO	RO	EMPH	CONS	HC
--	----	-----	----	------	------	----

COP	0.506 (0.118)	0.053 (0.045)	0.070 (0.058)	0.071 (0.044)	0.278 (0.055)	0.022 (0.006)
UIP	0.507 (0.130)	0.059 (0.076)	0.071 (0.045)	0.066 (0.029)	0.269 (0.042)	0.029 (0.016)
NSIP	0.502 (0.109)	0.073 (0.066)	0.070 (0.041)	0.063 (0.024)	0.268 (0.042)	0.024 (0.006)



**Figure 4.5** Further examples of HRCT images (middle row) and the results of semantic segmentation before fine-tuning and after fine-tuning in the same HRCT: (a) Original images, (b) without fine-tuning, and (c) with fine-tuning.

#### 4.4 Comparison between deep learning based and human segmentations in radiomics

Deep learning based automatic segmentation (DLAS) supports the reproducibility of

radioactive functions, but its effect on radioactivity modeling is still unknown; therefore, DLAS facilitated powerful anatomical and physiological MRI functions to evaluate whether it can accurately assess the treatment response of glioblastoma patients.

### *Dataset*

The training dataset consisted of a population of adult patients who presented with histologically proven glioblastoma at Seoul National University Hospital (SNUH) from March 2014 to August 2018. It contained 238 preoperative MRI acquisitions from initial diagnoses. The validation set was comprised of two groups: The first is an internal validation set, which included 40 pre-treatment glioblastoma patients and 53 post-treatment glioblastoma patients. The second is an external validation set with 91 patients from SNUH, which was included to test the generalizability and the true performance of the algorithm. This set included 58 pre-treatment glioblastoma patients and 33 post-treatment glioblastoma patients. In this study, we reported reproducibility and accuracy in the radioactive features acquired from DLAS and focused on CE-T1w imaging.

### *Image pre-processing and reference segmentation*

MRI examinations included the following sequences: unenhanced T1- and T2-weighted imaging, FLAIR, contrast-enhanced T1-weighted imaging, diffusion-weighted imaging, and dynamic susceptibility contrast imaging. The pre-processing of the MRI examinations, which include resampling to  $1 \times 1 \times 1$  mm and co-registration was performed. Bias field correction and intensity normalization were applied to the CE-T1w images using ANTsR and WhiteStripe packages [74, 75] in the R software package (R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>, 2016). Calculated ADC and CBV maps were then co-registered to the CE-T1w images using the SPM software ([www.fil.ion.ucl.ac.uk/spm/](http://www.fil.ion.ucl.ac.uk/spm/)). The co-registration process includes the generation of a brain mask from a CE-T1w and

transformation to ADC and CBV maps for each patient. Images were registered on the brain-extracted CE-T1w volume using rigid transformation with 6 degrees of freedom.

For generating a reference mask, segmentation was performed in the enhancing tumor region by a neuroradiologist (with four years of experience in neuro-oncological imaging), who semi-automatically defined the 3D CE-T1w using a segmentation threshold and region-growing segmentation algorithm that was implemented using a software (MITK, [www.mitk.org](http://www.mitk.org) German Cancer Research Center, Heidelberg, Germany) [76]. All segmented images were validated by a neuroradiologist (with 18 years of experience in neuro-oncologic imaging).

#### *Segmentation network architecture*

For the training of the glioblastoma dataset, we used the deep learning method in 3D U-Net with SE building blocks.

#### *Feature extraction*

Radiomics features from CE-T1w—our radiomics feature groups—were extracted using Matlab R2016a (The Mathworks, Natick, MA): volume and shape features, first-order features, texture features, and wavelet-transformed features, in accordance with a previously described automated process [77]. The radiomic features used here adhered to the standards set by the Imaging Biomarker Standardization Initiative (IBSI) [78]. For each segmented mask, there were seven volumes and shape features, 17 first-order features, 162 texture features, and 1432  $([17 + 162] \times 8)$  wavelet features, which resulted in 1618 radiomics features for each imaging sequence. Radiomics features were obtained from the CE-T1w images. The DLAS accuracy was assessed using DSC.

#### *Feature reproducibility*

The reproducibility of the radiomics features was assessed by calculating concordance

correlation coefficients (CCCs). The CCC cutoff of 0.80 was applied, and those features with a CCC above the cutoff were considered to be reproducible and comparable. The reproducibility of the ADC and nCBV histogram parameters was calculated using intraclass correlation coefficients (ICC) and a two-way mixed effect model. Reproducibility was found to be excellent for ADC and CBV features (ICC, 0.82–0.99) and the first-order features (pre- and post-treatment, 100% and 94.1% remained, respectively), but lower for texture (pre- and post-treatment, 79.0% and 69.1% remained, respectively) and wavelet-transformed (pre- and post-treatment, 81.8% and 74.9% remained, respectively) features of CE-T1w. DLAS-based radiomics showed to have a performance level similar to that in human-performed segmentations in internal validation (AUC, 0.81 (95% CI, 0.64–0.99) vs. AUC, 0.81 (0.60–1.00),  $P = .80$ ) but slightly lower performance in external validation (AUC, 0.78 (0.61–0.95) vs. AUC, 0.65 (0.46–0.84),  $P = .23$ ).



## 5 Discussion

The medical image segmentation with deep learning clearly delineated the boundaries of the organs or tumors in diagnostic medical images, such as CT or MRI images, acquired using various types of medical imaging equipment. Segmentation through deep learning can hierarchically extract from simple features to structural and complex features, and it has been found to perform at a high level compared to the traditional methods. In addition, deep learning accelerates segment processing by automating complex medical image segmentation.

In this study, we attempted several ways to further improve the segmentation performance. First, several experiments for automatic segmentation were conducted using the data consisting of MRI images of the brain and breast, abdominal CT, and dental CBCT. The detection and segmentation of GBM from brain MRI in a fast, accurate, and reproducible manner is a challenging task. Many segmentation techniques were applied in accordance with the characteristics that help distinguish tumors from normal tissues. In case tumors can be distinguished using MRI, 3D U-Net techniques were employed. Therefore, we designed refined and fully convolutional neural networks with a hierarchical DL to segment sub-regions of GBM. Therefore, we provided various cascaded segmentation methods based on deep learning, for instance, initial segmentation to localize the brain tumor and fine segmentation to segment the GBM region precisely. For better and robust segmentation, transfer learning, and random cropping for data augmentation were used to reduce learning time and improve learning efficiency in relation to the models. Random cropping makes effective use of a limited number of data. Moreover, we found that the optimal cut-off value for data cropping was 16. The extra-validation experiment on BraTS 2017 data proved that our learning model is robust. With their reported high performances, deep learning methods can be considered the current state-of-the-art for GBM segmentation. In the traditional automatic GBM segmentation methods, translating former knowledge into probabilistic maps or selecting substantially representative features for classifiers is a challenging task. However, semantic segmentation

of CNN has the advantage of automatically enabling learning of representative complicated features for both healthy brain tissues and tumor tissues directly from the multi-modal MRI images. Future improvements and modifications to CNN architectures and the addition of complementary information from other imaging modalities, such as positron emission tomography (PET), magnetic resonance spectroscopy (MRS) and diffusion tensor imaging (DTI) may improve the current methods, which would eventually lead to the development of clinically acceptable automatic glioma segmentation methods for better diagnosis.

In infarct segmentation on DWI images, we evaluated for the performance enhancement of a semantic segmentation model in DWI images. Our method possesses the advantage of enhancing channel-wise information with feature maps in each semantic segmentation model. We compared encoder-decoder, basic U-Net, and nnU-Net, Dense U-Net with and without SE blocks. In addition, various loss functions, such as focal and DSC loss, were applied to 2D images. Every network with SE blocks and focal loss showed significantly better performance than those without SE blocks and DSC loss of conventional image-processing algorithms. Therefore, applying SE blocks and focal loss to segmentation networks improved the performance level of several brain segmentations. Moreover, this method is more robust in segmenting small infarct lesions, which is very important for clinical applications.

Through breast and other tissue segmentation of breast MRI images, we investigated the use of U-Net-based deep learning methods in two datasets and ways, prone, and supine with 2D U-Net and 3D U-Net. According to the DSC, JCS, and HD values obtained in this study, the U-Net-based segmentation method surpassed the existing methods that incorporate a traditional approach. It has also been shown that the U-Net-based segmentation method is relatively more stable than the conventional method for breast MRI images. In MRI with a dynamic range, the CNN-based segmentation model works well with small datasets and can greatly reduce time consumption. Since the contrast enhancement pattern was different from the prone and supine postures images, there were no significant benefits even by combining

the datasets to training. In addition, the parenchyma with a cancer region was mainly influenced by the training models. The pre-processing method of nnU-Net instead lowered the learning efficiency. It was possible to facilitate enough training in the basic U-Net architecture. From the results of comparing 2D U-Net- and 3D U-Net-based methods, it was found that the former performed better in breast and other tissue segmentation. The accuracy of this type of segmentation was high enough to facilitate the registration of prone and supine.

There are several limitations to this study. First, we studied only a small number of breast cancer patients (29 patients had 116 breast cancer tumors). Thus, there is a need for further studies, which would include a larger patient population. Second, although we used MRI scans with different MRI protocols, we made use of only one MRI unit of the same hospital. In future research works, we aim to apply and evaluate datasets from multiple centers.

For the segmentation of CT images of the kidney and sub-structures with RCC, we used a cascaded 3D U-Net with a smart labeling framework for semantic segmentation of RCC and the fine sub-structures of the kidney. The segmentation accuracy increased with the progress of each stage, and the overall performance was found to be reasonable compared with other state-of-the-art segmentation networks. Furthermore, it was successfully reduced the effort required to create new GTs from scratch. Only the modifications from the CNN segmentation were more efficient and timesaving, as well as less variable, compared to manual annotation. In this study, the authors used cascaded 3D U-Net architecture for coarse region detection, followed by fine region segmentation, and trained this architecture with smart labeling. This cascaded network showed superior performance compared to a single 3D U-Net. To validate the performance of the proposed method, we compared it with the recent competitive network—nnU-Net. This network had achieved excellent performance in the KiTS19 challenge with the ability to dynamically adapt to the details of the datasets. However, the segmentation results of nnU-Net, when applied to our dataset, were to be inferior to ours. Some of the reasons might be explained by the inadequate pre-processing of a dataset and by the fact

that the fine-tuning process of training data for nnU-Net was far different in our dataset.

In the pancreas segmentation, the automatic segmentation of the pancreas and its lesions is important and very useful in clinical practice, although it can be challenging due to the low contrast in boundary and variability in location, shape, and size. To improve the accuracy and efficiency of the pancreas and focal lesions segmentation, we implemented the cascaded U-Net method. In the first step, we trained with 3D U-Net, but we, in the second step, applied 2D U-Net models to each lesion. The pancreas segmentation was as high as an expert, and the pancreas area was a large voxel in the CT image, which resulted in a considerable amount of time to be consumed to draw the label manually; however, the deep learning segmentation method significantly reduced time consumption. In the CT image, the cystic lesion was seen to have relatively clear boundaries, and the segmentation accuracy was higher than that for the cancer lesions. However, in cancer lesions, the intensity level of the NET lesion was brighter than in the case of other cancers. Therefore, the cancer segmentation quality of NET lesion was significantly lower than that of other solid lesions. Nevertheless, the total segmentation accuracy was not affected as the number of the NET was small. However, the accuracy results of the external test set were not as good as the internal test results in Kyung-Hee Hospital was not the same as the image quality and scan machines in ours. This study's limitation is that even though the total number of datasets obtained in this study is sufficient, the labels for each lesion are unbalanced. Therefore, in order to solve such data imbalance, we need to learn by training weights proportional to the number of imbalances for each label or apply an augmentation method. In addition, we have to learn small lesions at the patch level or apply several different networks for another study of the common pancreas and focal lesions.

In addition, we proposed an automatic 3D segmentation method, demonstrated very fast segmentation results for multi-facial structures in CBCT with reasonable accuracies; these structures include hard tissues, maxillary sinus, mandibular, and mandibular canals. Furthermore, since the mandibular canal was a fairly small area, the mandibular canal could

be learned more easily by the segmentation of the mandibular canal from the segmented mandible image. In dental CBCT, we showed that a fully convolutional, deep neural network model has a higher performance level than hard tissue, maxilla sinus, mandible, and mandibular canal segmentation. In fact, hard tissue, maxilla, and mandible canals are quite large at the voxel level; therefore, DSC is bound to be generated adequately. Consequently, we had to confirm the results related to the CBCT image and other evaluation metrics. As a result, it appears that the condyle segmentation performance was poor in the mandible area. In the future, we will obtain the data labeled around the condyle region. If this model is applied to the mandible test data, it is believed that the accuracy of the whole mandibular segmentation will improve. There are still some limitations to the findings of our study. First, our experiments can lead to overfitting. In general, the main reason for overfitting is that the training dataset is small in size and that it is only imported by some hospitals. Therefore, we plan to improve performance by evaluating the method using a wider range of training datasets, such as deeper networks and low contrast data. Ultimately, our objective is to be able to handle a wider range of medical data. Second, the segmentation network used is limited. In the future, we must plan to improve the performance of segmentation by applying wider and more diverse networks, including deep networks. To reduce the time required to apply deep learning algorithms in the clinical environment and increase accuracy, additional segmentation research is needed to be applied to various clinical environments.

In advanced issues of medical image segmentation, smart labeling was used for the new GTs to be segmented before manual correction and the network to be iteratively trained with limited data, instead of suggesting the most effective annotation areas. Our model performed well in the later stages. With the completion of each step, the DSC value increased, which means that the network improved with iterative learning and the use of additional labels. CNN-corrected segmentation has been shown to be more effective than manual segmentation as the former can be carried out much easier and faster than manual segmentation. Considering that

labeling is fundamental but extremely labor-intensive, which makes initiating deep learning difficult, this model can be considered a useful alternative in this regard. In addition, human labels are not always constant in the segmentation process due to intra-human and inter-human variabilities. Active learning with smart labeling frameworks can reduce this uncertainty by boosting accuracy by increasing collaboration with the help of deep learning algorithms.

As a radiometric study, we researched whether deep learning based segmentation could ensure the robustness of anatomical and physiological MRI functions and make accurate assessments of the response to treatment of patients with glioblastoma. Then, deep learning based segmentation showed high reproducibility in the primary features of anatomical and physiological MRI. We tested the effects of DLAS on the robustness of radiomics features and quantitative ADC and CBV parameters of tumor cellularity and vascularity across different centers and the diagnostic performance of the extracted features in the cases of post-treatment glioblastomas. Quantitative analysis based on DLAS has proven to be feasible, producing high segmentation accuracy and feature reproducibility in various medical centers, in particular with the first-order quantitative imaging parameters of ADC and CBV and first-order radiomics features extracted from CE-T1w. In our study, the deep learning segmentation method used was 3D U-Net, which developed more technically with the recent works on simultaneous volume segmentation through VNet and simultaneous image registration. The difference between our study and previous segmentation studies lies in our study's objective of testing the reproducibility of DLAS-derived radiomics features from anatomical imaging and quantitative parameters from physiologic imaging, especially in an independent validation set. The automatization of image registration and segmentation will further enhance the reproducibility of imaging features, and future studies on testing reproducibility using the abovementioned methods are needed. This work provided a segmentation mask but did not account for the differences in radiomics features according to segmentation. Our study emphasizes the importance of DLAS as well as the possibility of considering the optimization

of DLAS for quantitative feature extraction.

For the DILD pattern segmentation study, we fine-tuned deep learning with convolutional neural net architecture through manual drawing by two radiologists to enhance the automatic six image pattern segmentation of DILDs in HRCT through a conventional machine-learning method with a patch-based dataset. In detail, semantic segmentation with U-Net was trained based on the classification results of the previous SVM classifier. In addition, this semantic segmentation has been fine-tuned through 92 images manually drawn by the two radiologists. The whole lung quantification of CNN with the fine-tuned model exhibited a significantly higher accuracy than that of the SVM and deep learning before-fine-tuning models. In addition, the proposed fine-tuned model provides 10 times higher computational efficiency as well as the potential to overcome the inherent imbalance in the six image patterns. The final semantic segmentation displayed a similar accuracy to the agreement of the two radiologists. This study used previously developed semantic lung parenchyma segmentation on HRCT of patients with DILD [79] and patch-level conventional classifier using an SVM and fine-tuned post-processing [72] in order to generate a large-scale whole lung classification result. This large-scale dataset was used to train a more robust deep learning based semantic segmentation. However, training these large amounts of datasets could cause an extreme imbalance in the DILD image patterns. Therefore, we fine-tuned the semantic segmentation using image-level manual drawings created by the two radiologists to overcome these imbalances and enhance the semantic segmentation model.

From a clinical perspective, the method could be used to develop an image biomarker in HRCT for treating novel anti-fibrotic agents in patients with DILD, which might be currently available. The accuracy of the U-Net-based fine-tuned model could be used as an imaging biomarker for making treatment decisions on novel anti-fibrotic agents. Making use of CNN with the fine-tuned method effectively improved the performance; however, this method displays several limitations. First, the HRCT data were acquired from only one center.

Therefore, in further studies, multi-center data would be required. Second, the agreements between the U-Net-based semantic segmentation and the two radiologists were found to be quite low. Although they are higher than those of the SVM results and semantic segmentation before fine-tuning, it is necessary to enhance the accuracy of semantic segmentation by label curation through approaches such as human-in-loop correction, etc. Third, there are excessively strong imbalances in the six image patterns in the dataset; to overcome these problems, we need more innovative augmentation methods such as Perlin noise and GAN-augmentation.

Overall, we evaluated and tried various kinds of experiments to find adequate deep learning based semantic segmentation models with a medical image with high accuracy and relevance to the clinical context. Thus, we proposed a fully automatic segmentation network with various kinds of CNN models considering organ-specific, image modality-specific, and image reconstruction-specific variations. Basically, in the case of slices with dense thickness, 2D semantic segmentation has a better performance. In addition, pre-processing is sensitive to the development of robust segmentation, which needs image normalization and various types of augmentation. However, since modern GPUs lack memory for 3D semantic segmentation, cascaded semantics segmentation or patch-based semantic segmentation shows better results. An anatomic variation could be easily trained by semantic segmentation, but disease variations of cancer are hard to train. In addition, size-invariant semantic segmentation could be one of the important issues in medical image segmentation. Variations in contrast agent uptake could be vulnerable to the overall performance of semantic segmentation. For multi-center evaluation, subtle variations including variations of vendors, image protocols, and noise levels between centers could be problematic in training robust semantic segmentation. In addition, the labeling of semantic segmentation is too tedious and time-consuming. Therefore, deep learning based smart labeling would be needed. Given these issues, various applications with semantic segmentation in medical images, including active learning with smart labeling,



robust radiomics analysis, and disease pattern segmentation, automated segmentation were developed and evaluated. Semantic segmentation in medical images has been already shown to have great potential in detecting and analyzing tumors in clinical images, and this trend will undoubtedly continue in the future.

## 6 Conclusion

In this paper, we have found that adequate semantic segmentation with deep learning in medical images can improve the segmentation quality. Such an improvement can be helpful in computer-aided diagnosis, image quantification, and surgical planning in actual clinical settings. In addition, we provided an overview of the main technologies, discussed technical details, the major challenges of deep learning based medical image segmentation, and its applications. Most importantly, we have found an effective solution to deal with the major challenges of medical image segmentation, including robust segmentation for smart labeling, radiomics analysis, and imaging biomarker development of the DILD progress. We believe that this paper can help present and future researchers choose the right network structure and handle major issues related to medical image segmentation. Finally, deep learning based medical image segmentation and its application may be sufficient to provide practical utility to many physicians and patients who do not need to learn sectional anatomy.

## References

1. Maulik, U., *Medical image segmentation using genetic algorithms*. IEEE Trans Inf Technol Biomed, 2009. **13**(2): p. 166-73.
2. Sharma, N. and L.M. Aggarwal, *Automated medical image segmentation techniques*. J Med Phys, 2010. **35**(1): p. 3-14.
3. Rekik, I., et al., *Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal*. Neuroimage Clin, 2012. **1**(1): p. 164-78.
4. Cai, L., J. Gao, and D. Zhao, *A review of the application of deep learning in medical image classification and segmentation*. Ann Transl Med, 2020. **8**(11): p. 713.
5. Chen, X., et al., *Anatomy-Regularized Representation Learning for Cross-Modality Medical Image Segmentation*. IEEE Trans Med Imaging, 2020. **PP**.
6. Doi, K., *Computer-aided diagnosis in medical imaging: historical review, current status and future potential*. Computerized medical imaging and graphics, 2007. **31**(4-5): p. 198-211.
7. Gao, J., et al., *Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview*. Math Biosci Eng, 2019. **16**(6): p. 6536-6561.
8. Kubota, K., et al., *Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images*. Surg Endosc, 2012. **26**(5): p. 1485-9.
9. Hesamian, M.H., et al., *Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges*. J Digit Imaging, 2019. **32**(4): p. 582-596.
10. Lundervold, A.S. and A. Lundervold, *An overview of deep learning in medical imaging focusing on MRI*. Z Med Phys, 2019. **29**(2): p. 102-127.
11. Yamanakkanavar, N., J.Y. Choi, and B. Lee, *MRI Segmentation and Classification of Human Brain Using Deep Learning for Diagnosis of Alzheimer's Disease: A Survey*. Sensors (Basel), 2020. **20**(11).
12. Farmaki, C., et al., *Spatially adaptive active contours: a semi-automatic tumor segmentation framework*. Int J Comput Assist Radiol Surg, 2010. **5**(4): p. 369-84.
13. Li, H., et al., *Erratum: "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours" [Med. Phys., - (2008)]*. Med Phys, 2008. **35**(12): p. 5958.
14. Fritscher, K.D., et al., *Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours*. Med Phys, 2014. **41**(5): p. 051910.
15. Chen, X., et al., *Medical image segmentation by combining graph cuts and oriented active appearance models*. IEEE Trans Image Process, 2012. **21**(4): p. 2035-46.
16. Sjoberg, C. and A. Ahnesjo, *Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures*. Comput Methods Programs Biomed, 2013. **110**(3): p. 308-19.
17. Wang, H., et al., *Multi-Atlas Segmentation with Joint Label Fusion*. IEEE Trans Pattern Anal Mach Intell, 2013. **35**(3): p. 611-23.

18. Gholipour, A., et al., *Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly*. Neuroimage, 2012. **60**(3): p. 1819-31.
19. Sun, L., L. Zhang, and D.Q. Zhang, *Multi-Atlas Based Methods in Brain MR Image Segmentation*. Chin Med Sci J, 2019. **34**(2): p. 110-119.
20. Zhao, Z.Q., et al., *Object Detection With Deep Learning: A Review*. Ieee Transactions on Neural Networks and Learning Systems, 2019. **30**(11): p. 3212-3232.
21. Luongo, F., et al., *Deep learning-based computer vision to recognize and classify suturing gestures in robot-assisted surgery*. Surgery, 2020.
22. Black, K.M., et al., *Deep learning computer vision algorithm for detecting kidney stone composition*. BJU Int, 2020. **125**(6): p. 920-924.
23. Villalba-Diez, J., et al., *Deep Learning for Industrial Computer Vision Quality Control in the Printing Industry 4.0*. Sensors (Basel), 2019. **19**(18).
24. Voulodimos, A., et al., *Deep Learning for Computer Vision: A Brief Review*. Comput Intell Neurosci, 2018. **2018**: p. 7068349.
25. Chi, W., et al., *Deep learning based medical image segmentation with limited labels*. Phys Med Biol, 2020.
26. Wang, G., et al., *Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning*. IEEE Trans Med Imaging, 2018. **37**(7): p. 1562-1573.
27. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology*. Insights Imaging, 2018. **9**(4): p. 611-629.
28. Kim, M., et al., *Deep Learning in Medical Imaging*. Neurospine, 2020. **17**(2): p. 471-472.
29. Lee, J.Y., et al., *Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets*. Sci Rep, 2020. **10**(1): p. 8379.
30. Lau, M.M. and K.H. Lim, *Convolutional and Fully Connected Layer in DFN*. Journal of Information Science and Engineering, 2020. **36**(5): p. 1069-1078.
31. Jordao, A., M. Lie, and W.R. Schwartz, *Discriminative Layer Pruning for Convolutional Neural Networks*. Ieee Journal of Selected Topics in Signal Processing, 2020. **14**(4): p. 828-837.
32. Uchida, K., M. Tanaka, and M. Okutomi, *Coupled convolution layer for convolutional neural network*. Neural Networks, 2018. **105**: p. 197-205.
33. Yu, Z., et al., *Convolutional networks with cross-layer neurons for image recognition*. Information Sciences, 2018. **433**: p. 241-254.
34. Shelhamer, E., J. Long, and T. Darrell, *Fully Convolutional Networks for Semantic Segmentation*. IEEE Trans Pattern Anal Mach Intell, 2017. **39**(4): p. 640-651.
35. Ma, X., et al., *U-Net based deep learning bladder segmentation in CT urography*. Med Phys, 2019. **46**(4): p. 1752-1765.
36. Livne, M., et al., *A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease*. Front Neurosci, 2019. **13**: p. 97.
37. Falk, T., et al., *U-Net: deep learning for cell counting, detection, and morphometry*. Nat Methods, 2019. **16**(1): p. 67-70.

38. Al-Masni, M.A., et al., *A Two Cascaded Network Integrating Regional-based YOLO and 3D-CNN for Cerebral Microbleeds Detection*. Annu Int Conf IEEE Eng Med Biol Soc, 2020. **2020**: p. 1055-1058.
39. Raju, M., et al., *Multi-class diagnosis of Alzheimer's disease using cascaded three dimensional-convolutional neural network*. Phys Eng Sci Med, 2020.
40. Liu, J., et al., *A Cascaded Deep Convolutional Neural Network for Joint Segmentation and Genotype Prediction of Brainstem Gliomas*. IEEE Trans Biomed Eng, 2018. **65**(9): p. 1943-1952.
41. Viola, P. and M.J. Jones, *Robust real-time face detection*. International Journal of Computer Vision, 2004. **57**(2): p. 137-154.
42. Yeghiazaryan, V. and I. Voiculescu, *Family of boundary overlap metrics for the evaluation of medical image segmentation*. J Med Imaging (Bellingham), 2018. **5**(1): p. 015006.
43. Havaei, M., et al., *Brain tumor segmentation with Deep Neural Networks*. Med Image Anal, 2017. **35**: p. 18-31.
44. Kwon, D., et al., *Combining generative models for multifocal glioma segmentation and registration*. Med Image Comput Comput Assist Interv, 2014. **17**(Pt 1): p. 763-770.
45. Bauer, S., et al., *A survey of MRI-based medical image analysis for brain tumor studies*. Physics in Medicine and Biology, 2013. **58**(13): p. R97-R129.
46. Havaei, M., et al., *Within-brain classification for brain tumor segmentation*. Int J Comput Assist Radiol Surg, 2016. **11**(5): p. 777-88.
47. Menze, B.H., et al., *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)*. IEEE Trans Med Imaging, 2015. **34**(10): p. 1993-2024.
48. Zikic, D., et al., *Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR*. Med Image Comput Comput Assist Interv, 2012. **15**(Pt 3): p. 369-76.
49. Kamnitsas, K., et al., *Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation*. Med Image Anal, 2017. **36**: p. 61-78.
50. Wheeler, H.M., et al., *Early diffusion-weighted imaging and perfusion-weighted imaging lesion volumes forecast final infarct size in DEFUSE 2*. Stroke, 2013. **44**(3): p. 681-5.
51. Muir, K.W., et al., *Imaging of acute stroke*. Lancet Neurol, 2006. **5**(9): p. 755-68.
52. Yoo, A.J. and T. Andersson, *Thrombectomy in Acute Ischemic Stroke: Challenges to Procedural Success*. J Stroke, 2017. **19**(2): p. 121-130.
53. Lee, H., et al., *Fully Automated and Real-Time Volumetric Measurement of Infarct Core and Penumbra in Diffusion- and Perfusion-Weighted MRI of Patients with Hyper-Acute Stroke*. J Digit Imaging, 2020. **33**(1): p. 262-272.
54. Dafni, U., Z. Tsourti, and I. Alatsathianos, *Breast Cancer Statistics in the European Union: Incidence and Survival across European Countries*. Breast Care (Basel), 2019. **14**(6): p. 344-353.
55. DeSantis, C.E., et al., *Breast cancer statistics, 2019*. CA Cancer J Clin, 2019. **69**(6): p. 438-451.
56. Boyd, N.F., et al., *Mammographic density and the risk and detection of breast cancer*.

- N Engl J Med, 2007. **356**(3): p. 227-36.
57. Youlden, D.R., et al., *The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality*. Cancer Epidemiol, 2012. **36**(3): p. 237-48.
  58. Mann, R.M., C.K. Kuhl, and L. Moy, *Contrast-enhanced MRI for breast cancer screening*. J Magn Reson Imaging, 2019. **50**(2): p. 377-390.
  59. Chitalia, R.D. and D. Kontos, *Role of texture analysis in breast MRI as a cancer biomarker: A review*. J Magn Reson Imaging, 2019. **49**(4): p. 927-938.
  60. Zhang, Y., et al., *Automatic Breast and Fibroglandular Tissue Segmentation in Breast MRI Using Deep Learning by a Fully-Convolutional Residual Neural Network U-Net*. Academic Radiology, 2019. **26**(11): p. 1526-1535.
  61. Melloni, P. and R. Valls, *The use of MRI scanning for investigating soft-tissue abnormalities in the elbow*. European Journal of Radiology, 2005. **54**(2): p. 303-313.
  62. Niukkanen, A., et al., *Quantitative Volumetric K-Means Cluster Segmentation of Fibroglandular Tissue and Skin in Breast MRI*. J Digit Imaging, 2018. **31**(4): p. 425-434.
  63. Nie, K., et al., *Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI*. Med Phys, 2008. **35**(12): p. 5253-62.
  64. Gubern-Merida, A., et al., *Breast segmentation and density estimation in breast MRI: a fully automatic framework*. IEEE J Biomed Health Inform, 2015. **19**(1): p. 349-57.
  65. Lin, M., et al., *Template-based automatic breast segmentation on MRI by excluding the chest region*. Med Phys, 2013. **40**(12): p. 122301.
  66. Milenkovic, J., et al., *Automated breast-region segmentation in the axial breast MR images*. Comput Biol Med, 2015. **62**: p. 55-64.
  67. Dalmis, M.U., et al., *Using deep learning to segment breast and fibroglandular tissue in MRI volumes*. Med Phys, 2017. **44**(2): p. 533-546.
  68. Zhang, L., et al., *Automated deep learning method for whole-breast segmentation in diffusion-weighted breast MRI*. J Magn Reson Imaging, 2020. **51**(2): p. 635-643.
  69. Wang, C.B., et al., *Breast tumor movements analysis using MRI scans in prone and supine positions*. Sci Rep, 2020. **10**(1): p. 4858.
  70. Jun, S., et al., *Development of a Computer-Aided Differential Diagnosis System to Distinguish Between Usual Interstitial Pneumonia and Non-specific Interstitial Pneumonia Using Texture- and Shape-Based Hierarchical Classifiers on HRCT Images*. J Digit Imaging, 2018. **31**(2): p. 235-244.
  71. Lim, J., et al., *Regional context-sensitive support vector machine classifier to improve automated identification of regional patterns of diffuse interstitial lung disease*. J Digit Imaging, 2011. **24**(6): p. 1133-40.
  72. Chang, Y., et al., *A support vector machine classifier reduces interscanner variation in the HRCT classification of regional disease pattern in diffuse lung disease: comparison to a Bayesian classifier*. Med Phys, 2013. **40**(5): p. 051912.
  73. Kim, G.B., et al., *Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease*. J Digit Imaging, 2018. **31**(4): p. 415-424.
  74. Nolden, M., et al., *The Medical Imaging Interaction Toolkit: challenges and*

- advances*. International Journal of Computer Assisted Radiology and Surgery, 2013. **8**(4): p. 607-620.
75. Avants, B.B., et al., *A reproducible evaluation of ANTs similarity metric performance in brain image registration*. Neuroimage, 2011. **54**(3): p. 2033-44.
  76. Louis, D.N., et al., *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary*. Acta Neuropathol, 2016. **131**(6): p. 803-20.
  77. Shinohara, R.T., et al., *Statistical normalization techniques for magnetic resonance imaging*. Neuroimage-Clinical, 2014. **6**: p. 9-19.
  78. Kang, D., et al., *Diffusion radiomics as a diagnostic model for atypical manifestation of primary central nervous system lymphoma: development and multicenter external validation*. Neuro-Oncology, 2018. **20**(9): p. 1251-1261.
  79. Park, B., et al., *Lung Segmentation on HRCT and Volumetric CT for Diffuse Interstitial Lung Disease Using Deep Convolutional Neural Networks*. J Digit Imaging, 2019. **32**(6): p. 1019-1026.

## Abstract (In Korean)

의료 영상 분할은 영상의 정량화, 자동 진단, 수술 계획 등 다양한 의료영상을 이용한 응용에 필수적인 역할을 하고, 장기 혹은 병변의 모양과 부피에 대한 중요한 정보를 제공한다. 그러나, 사람이 수동으로 분할을 한다면 일관된 분할이 어렵고 시간도 많이 소요된다. 따라서, 고전적인 영상 처리 혹은 머신 러닝 방법 등을 이용하여 반자동 또는 자동으로 영역을 분할하는 다양한 연구가 진행되었다. 그러나 기존의 영상처리 및 머신 러닝을 통한 분할 방법들은 영상의 해부학적 모양 및 질병 상태 등에 따라 영상의 특징이 매우 다양하여, 최적의 분할 방법을 찾기가 어렵다. 따라서, 많은 연구자들이 보다 안정되고 효율적인 의료 영상 분할 방법을 모색하고 있는 중이다.

최근 몇 년 동안 딥러닝 모델은 컴퓨터 비전 분야에 성공적으로 적용되어서 의료 영상 분야에도 빠르게 적용되고 있다. 특히, 의료 영상 분할에서도 다양한 해부학적 정보와 질병에 대해 정확하고 견고한 분할 결과를 보여주며, 비약적인 발전을 이루었다. Residual Net, Fully Convolutional Network(FCN) 및 U-Net 과 같은 몇 가지 심층 합성곱신경망(Deep Convolution Neural Net; CNN) 모델이 제안되었다. 이러한 모델은 이미지 분류, 세분화, 물체 감지 및 추적 작업을 위한 최첨단 성능을 제공할 뿐만 아니라, 이미지 처리에 대해 새로운 관점을 제공한다. 따라서, 딥러닝은영상의학 전문의와 외과외과가 컴퓨터 단층 촬영(CT) 또는 자기 공명 영상(MRI) 이미지 등에서 다양한 해부학적 구조를 분할하는 데 도움을 줄 수 있다.

본 논문에서는 정확성과 임상적 관점에서 딥러닝을 기반으로 적절한 의료 영상 분할 모델을 찾고 적용하기 위해 다양한 실험을 평가하고 시도하였다. 이 연구의 목적은 두 가지이다. 1) 임상적 가치와 영상의 특성에 기반하여 딥러닝을 사용한



분할 모델을 찾고 개발한다. 2) 스마트 레이블링을 통한 능동적 학습(Active Learning using smart labeling), 강건한 레디오믹스(Radiomics) 분석, 질병 패턴 분할 등 의료영상 분할을 이용한 다양한 응용에 생기는 문제를 해결하는 분할 기법을 개발하고 적용한다. 이를 위해 장기의 위치 및 해부학적 특성, 이미지 모달리티 등을 고려한 다양한 종류의 병변에 대해 완전 자동 분할이 가능한 네트워크를 제안했다. 딥러닝 기반의 분할 네트워크 모델은 뇌 MRI의 신경 교종, 급성 뇌졸중 검색, MRI의 유방 및 기타 조직, 채장암이 포함된 채장 분할, CBCT의 하악골 및 상악동, 등과 같은 의료 영상 데이터를 가지고 수행되었다.

일반적으로 두께가 두꺼운 슬라이스의 경우 2D 영상 분할이 더 나은 성능을 보여주었다. 또한, 영상 특성에 맞는 전처리가 강인한 영상 분할 기법을 개발하는데 민감하다. 이는 딥러닝의 특성상 이미지 표준화 중요하다는 것을 알수 있다. 또한, 컴퓨터 그래픽 메모리는 3차원의 영상을 자동 분할을 하기에 메모리가 부족하기 때문에, 계단식(Cascaded) 분할 또는 패치 기반 분할이 더 나은 결과를 보여주었다. 자연스런 해부학적 변이는 영상 분할(Semantic Segmentation)에 의해 상대적으로 쉽게 훈련될 수 있지만, 암의 질병 변이는 쉽게 훈련되지 않는다. 암 등 크기가 변하는 병변은 크기 불변(Scale invariant) 영상 분할이 중요한 문제 중 하나이다. 또한, 조영제에 따른 영상의 변화는 영상 분할의 전반적인 성능을 낮춘다. 멀티센터 평가의 경우 장비업체, 이미지 프로토콜 및 센터 간의 영상 수준 등의 변화에 따라 강인한 영상 분할을 학습하는데 문제가 된다. 또한 영상 분할의 레이블링은 시간이 상당히 많이 걸린다. 따라서 딥 러닝 기반으로 한 효율적인 스마트 라벨링이 필요하다. 이러한 문제들을 바탕으로 능동적 학습, 강인한 레디오믹스 분석, 영상 바이오마커 (imaging biomarker)를 위한 질병 패턴 분할 등 의료 영상 분할을 통한 다양한 응용을 위한 연구를 하였다.

결론적으로 우리는 의료 영상에서 딥러닝을 통한 최적의 영상 분할이 분할 품질을 개선할 수 있다는 것을 발견했으며, 이는 실제 임상 환경에서 컴퓨터 지원 진단, 영상 정량화 및 수술 계획에 도움이 될 수 있게 되었다. 의료 영상 분할과 그 적용은 해부학을 배울 필요가 있는 많은 의사와 환자에게 실질적인 유용성을 제공하기에도 충분할 것이라 사료된다.

## Acknowledgements

### 감사의 글

두려움 반 설레임 반으로 시작한 박사 학위 과정이 4 년만에 끝을 맺었습니다. 박사 과정을 시작해야 할지 말지에 대한 선택의 기로에 있을 때 이 길을 선택하고 확신을 갖게 도와준 모든 분들이 주마등처럼 스쳐 지나갑니다. 그 분들이 아니었다면 이 순간을 맞보지 못했을 것입니다. 처음에 그저 추상적으로 다가왔던 의료 인공지능 연구를 어떻게 시작해야 하는지, 어떤 사고를 가져야 하는지, 왜 하는지 등을 박사 과정을 통해 구체적으로 하나씩 깨우쳐 갔습니다. 많은 시행착오를 겪고 다양한 기회와 배움의 장이 된 연구실 생활이 제 인생에서 정말 중요한 시기였음을 다시한번 깨닫습니다.

제일 먼저 시간과 노력을 쏟아 지도해 주신 김남국 교수님께 감사의 말씀을 드립니다. 바쁘신 와중에도 논문을 지도하고 심사해 주신 김호성, 서준범, 윤지혜, 홍헬렌 교수님께도 감사를 드립니다.

함께 생활했던 여러 선생님과 교수님들께도 감사의 마음을 전합니다. 처음 같이 연구를 시작했던 박지은 교수님, 교수님의 열정에 저도 자극받고 더욱 열심히 할 수 있는 원동력이 되었습니다. 전반적인 연구실 생활에 정신적 지주가 되어주신 박사님이란 호칭이 더 익숙한 윤지혜 교수님, 깔끔하고 직관적인 조언으로 연구의 퀄리티를 높여주신 이현나 교수님, 무슨 일에 있어 진심어린 조언을 아끼지 않는 이경화 선생님, miccai 학회에서 같은 방을 쓰며 많은 추억을 쌓은 이아름 선생님, 보기만해도 미소 짓게 되는 귀여운 열정가 권진희 선생님, 아무리 힘들어도 웃음으로 승화시키는 밤샘 동료 김민지 선생님, 어떠한 요청도 싫은 내색 없이 똑딱 완성해서 주시는 금손 디자이너 김민경 선생님, 박사 심사 받는동안 도움 주신 신기원 선생님 등 이 외에도 학위 과정 중에 지혜와 마음을 나누어준 모든 분들께 감사 인사드립니다.

또한 지금의 제가 있기까지 물심양면으로 지원을 아끼지 않은 든든한 후원자이자 저의 버팀목이 된 자랑스러운 엄마, 무한 긍정의 아빠, 친구 같은 언니.. 우리 식구 사랑합니다. 귀엽고 사랑스러운 우리 똑순이 할머니, 오래오래 건강하세요! 필요한 것 똑딱 만들어주시는 분위기 메이커 이모, 저의 대학원 문을 열어 주신 지성과 인품을 겸비한 큰고모 존경하고 감사합니다.

늘 그렇듯 어떤 일을 마무리할 때 후회와 아쉬움이 남지만 유독 아쉬움이 많이 남는 것을 보면 그만큼 애정을 가진 연구들이었나 봅니다. 앞으로 더욱 의미 있는 연구로 많은 사람들에게 도움이 되기를 바라며, 이만 감사의 글을 마칩니다.