ANATOMIC CONTEXT-AWARE SEGMENTATION OF ORGANS-AT-RISK IN THORAX COMPUTED TOMOGRAPHY SCANS

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

COMPUTER ENGINEERING

By

Haya Shamim Khan Khattak

December 2022

ANATOMIC CONTEXT-AWARE SEGMENTATION OF ORGANS-AT-RISK IN THORAX COMPUTED TOMOGRAPHY SCANS
By Haya Shamim Khan Khattak
December 2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Çiğdem Gündüz Demir(Advisor)

Selim Aksoy(Co-Advisor)

Alptekin Temizel

Can Alkan

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

ANATOMIC CONTEXT-AWARE SEGMENTATION OF ORGANS-AT-RISK IN THORAX COMPUTED TOMOGRAPHY SCANS

Haya Shamim Khan Khattak
M.S. in Computer Engineering
Advisor: Çiğdem Gündüz Demir
Co-Advisor: Selim Aksoy
December 2022

Organ segmentation plays a crucial role in disease diagnosis and radiation therapy planning. Efficient and automated segmentation of the organs-at-risk (OARs) requires immediate attention since manual segmentation is a time consuming and costly task that is also prone to inter-observer variability. Automatic segmentation of organs-at-risk using deep learning is prone to predicting extraneous regions, especially in apical and basal slices of the organs where the shape is different from the center slices. This thesis presents a novel method to incorporate prior knowledge on shape and anatomical context into deep-learning based organ segmentation. This prior knowledge is quantified using distance transforms that capture characteristics of the shape, location, and relation of the organ position with respect to the surrounding organs. In this thesis, the role of various distance transform maps has been explored to show that using distance transform regression, alone or in conjunction with classification, improves the overall performance of the organ segmentation network. These maps can be the distance between each pixel and the center of the organ, or the closest distance between two organs; such as the esophagus and the spine. When used in a single-task regression model, these distance maps improved the segmentation results. Moreover, when used in a multi-task network with classification being the other task, they acted as regularizers for the classification task and yielded improved segmentations. The experiments were conducted on a computed tomography (CT) thorax dataset of 265 patients and the organs of interest are the heart, the esophagus, the lungs, and the spine. The results revealed a significant increase in f-scores and decrease in the Hausdorff distances for the OARs when segmented using the proposed model compared to the baseline network architectures.

Keywords: Deep learning, medical image analysis, organ-at-risk segmentation, lung cancer, distance transforms, computed tomography segmentation.

ÖZET

TORAKS BİLGİSAYARLI TOMOGRAFİ TARAMALARINDA RİSK ALTINDAKİ ORGANLARIN ANATOMİK İÇERİK FARKINDALI SEGMENTASYONU

Haya Shamim Khan Khattak
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Çiğdem Gündüz Demir
İkinci Tez Danışmanı: Selim Aksoy
Aralık 2022

Organ segmentasyonu, hastalık teşhisinde ve radyasyon tedavisi planlamasında çok önemli bir rol oynar. Manuel segmentasyon zaman alıcı ve maliyetli olmasının yanı sıra gözlemciler arası değişkenliğe de yatkın olduğundan, risk altındaki organların (RAO'ların) verimli ve otomatik segmentasyonu daha önemli bir hale gelir. Risk altındaki organların derin öğrenme kullanılarak otomatik segmentasyonu, özellikle merkez kesitlerin şeklinin apikal ve bazal kesitlerden farklı olduğu organlarda, hata yaparak dış bölgeleri tahmin etmeye eğilimlidir. Bu tez, şekil ve anatomik içerik hakkındaki ön bilgileri derin öğrenmeye dayalı organ segmentasyonuna dahil etmek için yeni bir yöntem sunar. Bu ön bilgiyi nicelemek için, organın şekli, konumu ve organ pozisyonunun çevredeki organlara göre ilişki karakteristiklerini yakalayan mesafe dönüşümlerini kullanır. Bu tezde, mesafe dönüşümü regresyonunun tek başına veya sınıflandırma ile birlikte kullanılmasının organ segmentasyonu ağının genel performansını iyileştirdiğini göstermek için çeşitli mesafe dönüşümlerinin rolü araştırılmıştır. Mesafe dönüşümü olarak, her piksel ile organın merkezi arasındaki mesafe veya örneğin yemek borusu ve omurga gibi iki organ arasındaki en yakın mesafe kullanılmıştır. Bu tez, tek görev regresyon modelinde kullanıldığında, bu mesafe dönüşümlerinin segmentasyon sonuçlarını iyileştirdiğini göstermiştir. Ayrıca, sınıflandırmanın diğer görev olduğu çok görevli bir ağda kullanıldığında, sınıflandırma görevi için düzenleyici olarak çalışıp segmentasyonu iyileştirdiği gözlemlenmiştir. Deneyler, 265 hastanın bilgisayarlı tomografi (BT) toraks veri seti üzerinde gerçekleştirilmiştir. Kalp, yemek borusu, akciğerler ve omurga ilgi bölgeleri olarak seçilmiştir. Sonuçlar, önerilen model kullanılarak RAO'lar için segmentasyon yapıldığında, temel ağ

mimarilerine kıyasla, f-skorlarında önemli bir artış ve Hausdorff mesafelerinde azalma olduğunu göstermiştir.

Anahtar sözcükler: Derin öğrenme, medikal görüntü analizi, riskli organ segmentasyonu, akciğer kanseri, mesafe dönüşümleri, bilgisayarlı tomografi segmentasyonu.

Acknowledgement

This endeavour could not have been possible without the continuous support, mentorship, and invaluable feedback of my advisor Prof. Çiğdem Gündüz Demir. It has been my honor to work under her supervision. She has been extremely helpful and patient throughout my research, and at the time of writing this thesis. Her valuable guidance and immense knowledge has inspired me to continue my work in the field of machine learning. I am thankful to my co-advisor, Prof. Selim Aksoy, for his mentorship and help during the semester.

I would like to express my gratitude to Prof. Alptekin Temizel and Assoc. Prof. Can Alkan for agreeing to review this thesis and becoming part of the jury. I am also grateful to Prof. Durmuş Etiz, Assoc. Prof. Melek Çoşar Yakar, and Mr Kerem Duruer from the Radiation Oncology Department at Eskişehir Osmangazi University School of Medicine, for providing the dataset as well as medical expertise for this study.

Lastly, I would like to thank my parents, my dear husband, and my brothers for being my constant support. Without them, this accomplishment would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

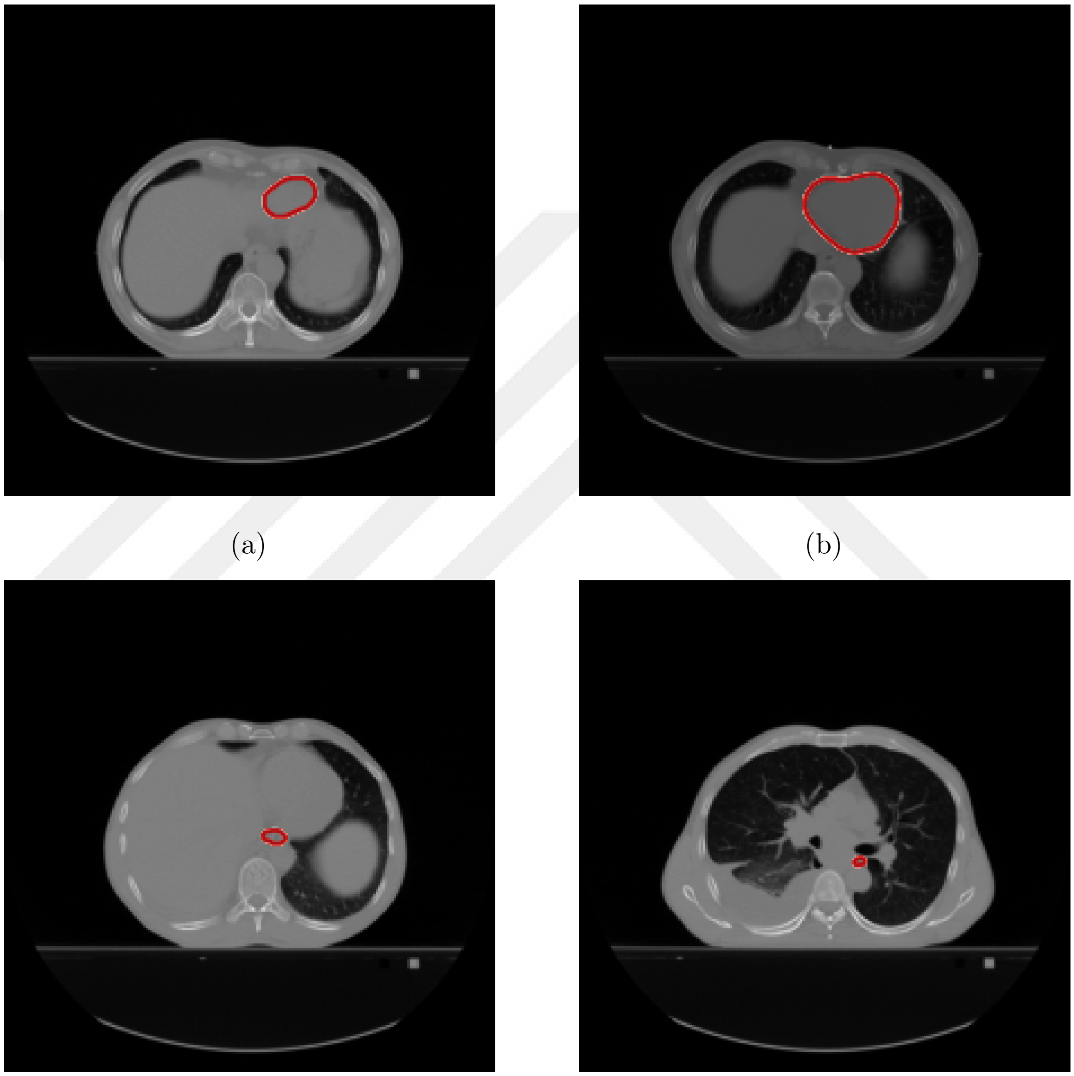
Introduction

Lung cancer patients are treated with radiation at one point during their prognosis, which includes careful radiation therapy to target only the cancerous tissues and protect the healthy organs. Computed tomography (CT) scan is the most commonly used cross sectional imaging modality used by radiation oncologists for radiation therapy planning. The procedure includes delineating all the organs in the thorax CT scans such as the heart, lungs, esophagus, and spine. Efficient and automated segmentation of these organs at risk (OARs) requires immediate attention since manual segmentation is a time consuming and a costly task and prone to inter-observer variability [5]. The aim of this thesis is to use manual segmentations to train a machine learning model that can accurately segment the organs in the thorax automatically. This information is then useful for disease diagnosis and radiation therapy planning.

However, this task comes with its challenges. Organs other than the lungs and the spine do not offer high contrasts in CT scan images, coupled with the fact that soft tissues and organs do not occupy the same position throughout the axial slices, it makes it extremely difficult to segment low contrast organs [6]. Most organs such as the heart and the esophagus also do not have a consistent shape throughout the axial slices, posing problems for generalized segmentation algorithms. The shape inconsistencies as well as variation in position in one of the patients can be seen in Figure 1.1.

Traditional fully convolutional network (FCN) based deep learning segmentation models such as the UNet [1] can perform organ segmentation of these thorax CT scans to some extent but they fail at apical and basal slices where the shape of the organ becomes very different. To address this problem, this thesis explored the use of distance transforms to embed more contextual and prior knowledge during the training of FCNs to achieve better segmentations in the difficult slices. The prior knowledge captures characteristics of the shape, location, and some domain information. Domain information refers to the anatomical facts that are true for all organs across all images such as that the heart is located always between the lungs. An axial slice may or may not contain all three organs i.e., the left and the right lung and the heart, but if the slices does contain all three then the heart will be located between the lungs. This thesis quantifies such domain information and uses it to boost the performance of traditional deep learning segmentation models.

To quantify this information, this thesis relies on making use of distance transform maps. Distance transform maps have been previously used to successfully boost segmentation networks [7]. A distance transform is a derived representation of a digital image; it receives a binary image as an input and outputs a distance map. The distance map is the same dimensions as that of the input image and contains the same number of pixels, however, each pixel contains the Euclidean distance to an obstacle pixel of choice such as the boundary, center, or any other point of interest. The distance maps can be signed or unsigned. In [8], signed distance maps are used in a deep regression network to obtain improved semantic segmentation. In the *DeepDistance* model [9], the distance maps calculated using the center of the cell, and the maps calculated using the boundary of the cell are used to successfully delineate boundaries of cells and achieve better cell detection in inverted microscopy images. This thesis makes use of the two kinds of distance maps: 1) inner-distance maps, which are calculated using the center of the organ, and 2) distance-to-landmark maps containing distance to a neighbouring organ. The landmark organ in this thesis is defined as a high contrast



(a) (b)

(c) (d)

Figure 1.1: Samples showing variation in shape and location of soft tissue organs across different slices in the thorax CT. The images in (a) and (b) show the heart contours in red; these images are obtained from the same patient but from different slice position of the thorax CT. The images in (c) and (d) show the same for the esophagus

easy-to-segment organ in the CT scan. This thesis introduces a novel method of utilizing the relationship between landmark organs and the low contrast organs using domain information to improve the segmentation of the low contrast, and thus, hard-to-segment organs.

The choice of landmark organs is made after a qualitative analysis of the CT scan images. The lungs and the spine offer a high contrast due to the role of air and bone density in tomography. The lungs filled with air occupy the darkest regions in the thorax whereas the spine is often the brightest object in a scan. This is consistent throughout the thorax scans. Therefore, this thesis proposes to use this as prior information to aid the segmentation model. More specifically, this work proposes the use of the lungs and the spine as landmark organs since the contrast of the lungs and the spine in the CT scans is more apparent than other soft tissues. The landmark organ will be used to create a distance map with respect to the target organ. This distance map contains more information than the simple ground truth maps where each pixel is labeled in the organ region. The distance map has values only in the region of the target organ, so it encompasses the shape and boundary information, and since these distance values are with respect to the landmark organ, the maps contain location information as well as information of its spatial proximity. The target organs chosen are the esophagus and the heart, and their landmark organ is the spine and the lungs, respectively. The relationship of these target organs to their landmark organs makes the distance maps very informative vis-'a-vis their location. Furthermore, this study uses inner-distance maps to improve the segmentation of the lungs and achieve state-of-the-art thorax OAR segmentation.

To the best of our knowledge, this is the first work that has incorporated domain knowledge by successfully quantifying it in the form of a distance transform that is then used to train an end-to-end multi-task FCN, resulting in a fully automated segmentation model with improved results. This thesis also shows how to use easier to obtain distance transforms such as distance to the center or to the boundary as regression tasks and obtain better segmentations.

1.1 Contributions

The contributions of this thesis are as follows:

- We proposed an automated framework for the segmentation of OARs in thorax CT scans that achieves high f-scores. The ablation study shows that this network achieves competitive results compared to the state-of-the-art models. This thesis focused on five organs of the thorax: the heart, the esophagus, the right and the left lung, and the spine and showed the most suitable network for each of these organs. This model can be applied to other similar datasets as well such as OAR segmentation in abdomen CT scans, head and neck CT scans or even whole body CT scans.

- The proposed model leverages shape information of the organ using inner-distance maps generated from a distance transform that measures the distance between the organ centroid and each pixel of the organ. These distance maps, when used in a deep regression network, produce more accurate segmentations. Since each pixel carries more information about the morphology of the organ, the network infers better semantics after training using these maps.

- We proposed the use of a *distance-to-landmark* map, which is generated from a distance transform that measures the distance of the organ of interest to another landmark organ that is anatomically related and provides relevant spatial and location information of the target organ. These distance maps embed shape and location information, and anatomical context. When used in conjunction with a segmentation task in a multi-task learning framework, we achieved improved semantic segmentation of the organs in the apical and basal slices where simple networks often have trouble predicting correct segmentations.

1.2 Outline

This thesis is organized as follows. In Chapter 2, a summary of the related work in the literature and the background information is provided. In Chapter 3, a detailed methodology is given with in-depth explanations about the distance transform maps, how they are calculated and how they are used in our network, followed by an overview of the network design used. In Chapter 4, the details of the experiments are given with information about the datasets used, the parameters for training, and the calculations of the metrics as well as the motivations why they were chosen. This chapter also provides the experimental results. In Chapter 5, where the thesis is concluded with discussions for future work.

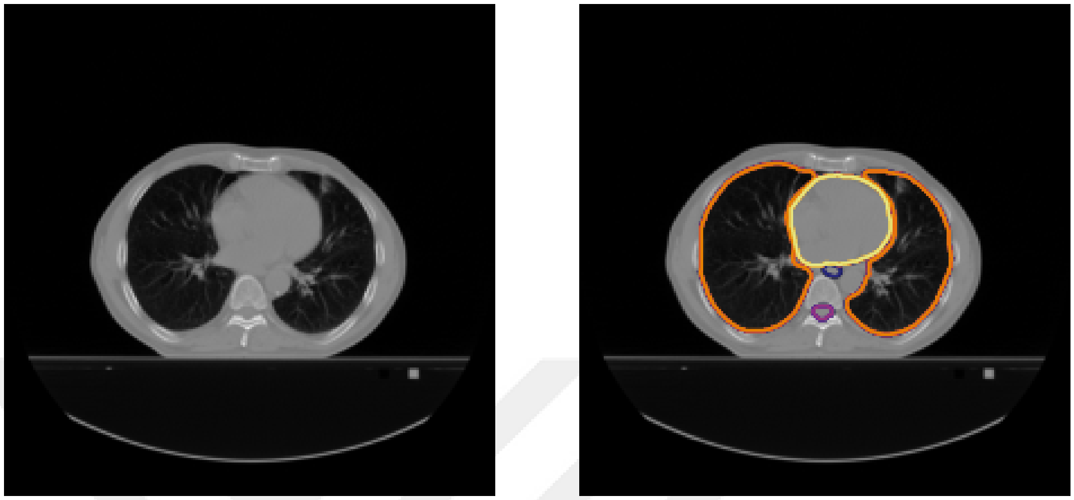
Chapter 2

Background

This chapter will provide background on computed tomography (CT) scans and the importance of automating organ delineation in CT scans. It will then briefly discuss traditional methods of automatic organ segmentation that are widely used in commercial tools used for planning radiation therapy. Moving on, this chapter will introduce deep learning methods and their application in medical image segmentation. Finally, it will outline some machine learning approaches that have been used for organ segmentation in CT images, as well as some other works that have leveraged distance transform maps or anatomical context to achieve improved performance of their models.

2.1 Radiation Therapy and Computed Tomography

One of the main causes of cancer deaths is lung cancer and according to American Cancer Society's projection, it causes 350 deaths per day in the United States [10]. Therefore, cutting edge research has been carried out to improve the mortality rate of lung cancer such as using low-dose computed tomography screening for



(a) (b)

Figure 2.1: A sample showing extremely low contrast in CT images for soft-tissues such as the heart and the esophagus in an axial slice from a thorax CT scan. (a)Original image and (b)Organ Contours.

early detection and non-invasive treatment and highly efficient treatment methods such as radiation therapy [11]. Radiation therapy relies on medical imaging techniques that make it possible for domain experts to view all the neighboring tissues and organs of the tumor and make accurate dose calculations. For now, CT is the primary three-dimensional imaging modality used for dose calculation in radiation oncology therapy [12]. However, CT does have some disadvantages. It offers suboptimal soft-tissue contrast, as shown in Figure 2.1. This makes it hard to delineate organs at risk and increases the chances of interobserver variations. Organs at risk (OAR) are the organs neighboring the tumor region that is to be targeted using the radiation beams. The ultimate goal of radiation therapy is to minimize any damage that can be caused by radiation to other soft-tissue by keeping the region of radiation as precise as possible. This makes it necessary that the organ and tumor segmentation is carried out accurately for efficient treatment and a highly precise organ segmentation method would reduce the risk involved in administering radiation therapy.

2.2 Organ Segmentation using Traditional Methods

Among other traditional organ segmentation methods, the most widely used are the statistical models and atlas-based methods.

A simple method for automatic segmentation that is included in a number of commercial products is the atlas-based method. With the images to be segmented, this approach registers atlas templates that contain precontoured structures, and the precontoured structures are propagated to the new images. The accuracy of image registration has a significant impact on the segmentation accuracy [13] of this method. Accurate registration is not always ensured due to variations in organ shape across slices and one patient differing from another. A large atlas dataset with enough variation can help solve this problem. However, it is challenging to incorporate all potential cases in the templates due to the unpredictable nature of tumor morphology.

For automated segmentation, the statistical model-based method [14] uses statistical shape models. The accuracy of these methods depends on the validity of the models. Although anatomical knowledge from existing datasets is used to build the models, the generalized models perform poorly on irregular pictures.

2.3 Deep Learning

2.3.1 Deep Learning for Organ Segmentation

Convolutional neural networks (CNNs) are the deep learning models that have shown remarkable performance in many computer vision tasks including semantic segmentation. Deep learning allows end-to-end training, and therefore, the need for meticulous feature extraction is skipped since the network is able to learn the feature representations automatically. One of the first works [15] performing automatic segmentation using deep learning proposed a fully convolutional network for 3D liver segmentation. An esophagus segmentation model [16] uses a random walker approach along with a CNN to incorporate neighboring pixel intensities during training to obtain more accurate segmentations. An early multi-organ segmentation using CNNs [17] segmented OAR in the head and neck CT images. It concluded that, although CNNs are potentially useful models for automatic segmentation, they can be improved using additional information.

OAR segmentation is important for radiation therapy, in which the tumor region, also called the clinical target volume (CTV), is delineated along with the OARs for effective and precise radiation. Most studies focus on OAR segmentation, except for [18] in which a deep dilated CNN was used to effectively segment the CTV, and five OARs for enhanced rectal cancer radiation therapy. The most popular model for medical image segmentation is UNet [1] since it not only labels each pixel but also salvages the location of the pixel by producing the output map in the same dimensions as the input. UNet uses the idea of defining skip connections between the corresponding encoder and decoder layers, in which features learned earlier on are carried to the denser layers, hence there is less loss of information in the pooling layers. This model has outperformed all other types of CNNs, particularly for medical image segmentation, as oftentimes these datasets are smaller. UNet does have some limitations, such as inaccurate boundary region segmentations, and so improved versions of this model are being introduced.

For instance, the UNet-GAN [19] model combines UNet in a GAN inspired model. Generative adversarial networks (GAN) are successful generative image modeling networks that use a generator and discriminator framework to produce results. UNet-GAN uses UNet as a generator and CNN as a discriminator to achieve segmentation results. However, there is some pre-processing involved where the region of interests (ROIs) have to be obtained using the lung volume and the center of the esophagus is chosen as the center of the total lung which may not be true for apical and basal slices depending upon the start and end points of the thoracic region. Another similar work [20] uses a 3D UNet to segment the OARs in the thorax. This is a two stage network in which the input is first passed through a 3D UNet for localization and then based on the predictions, the regions for each organ are cropped and then passed through another 3D UNet particular for the organ. Such a network can be costly and time consuming to train. A similar approach has been used by [21] for OAR segmentation on whole body CT images. This work attempts to delineate 50 organs in the head, thorax, and abdomen region and the UNet model performance decreases as labels are increased. A lot of cropping, pre-processing for localization, and post processing for merging the results is part of that approach. This work by Tran et al. [6] improves the esophagus segmentation using an attention module. To make their network more spatially aware, they deploy an atrous spatial pyramid pooling module to better capture the features of the esophagus, as it is one of the hardest organs to segment. In all the thorax CT image segmentation works, the lungs and the spine have consistently emerged with higher Dice scores whereas the esophagus has always been the most difficult organ to segment.

2.3.2 Deep Learning using Distance Maps

There have been some works where distance transforms have been taken advantage of for medical image segmentation. These include [22] that uses distance maps and multi-task learning to improve segmentation of heart chambers in cardiac cine magnetic resonance images (MRIs) in a UNet inspired 3D multi-task network. They use distance transforms as auxiliary tasks to refine the semantic segmentation results, unlike [23], which uses signed distance map prediction as their main task and achieves improvements for 3D organ segmentation. In another 3D segmentation work, Wang et al. [24] derived their distance transform for a tubular structure using the pixel distance to the surface of the tubular structure. They use this distance transform in 3D segmentation networks and have achieved increased performance for tubular segmentation. However, the transform used embeds shape information of the organ.

This work [25] combines the concept of distance transform maps as well as adversarial networks for multi-organ segmentation. They perform segmentation of the heart, esophagus, trachea, and aorta by generating multi-class distance maps for a deep regression GAN network in which the input to the discriminator is the distance maps. The use of anatomical context is even rare; the one shot PACS [26] introduces this concept by using a recurrent segmentation network but it requires planned CT images; i.e., CT images of each patient taken weekly. To the best of our knowledge, a model leveraging distance maps to embed shape, location, and anatomical context at the same time has not been developed for thorax CT images.

Chapter 3

Methodology

In this chapter, the details of the distance transforms and the consequent distance maps are provided. This proposed network relies on using these distance maps to obtain results. Therefore, the different kinds of distance maps used are explained thoroughly, along with the motivations for using each kind. Then, this chapter discusses the network architecture used for the experiments carried out for this thesis. The network architecture varies for different organs, and their distance map, hence the motivations for this choice are also explained.

3.1 Distance Transforms

The distance transforms used in this thesis are the inner-distance and the distance-to-landmark maps, details of which are discussed below.

3.1.1 Inner-Distance Map

The inner-distance is defined as the distance of every organ pixel to its centroid. Since anatomical organs are irregularly shaped with inconsistent and variable morphology throughout the CT slices, we calculate the centroid for each organ as part of the distance transform. First, the contours of the binary image containing the annotation of the organ is generated. This contour map is used to calculate image moments M, which are then used to find the centroid using; $C_x = \frac{M_{10}}{M_{00}}$ and $C_y = \frac{M_{01}}{M_{00}}$. This is similar to calculating the centroid by taking the average of the contours of the organ. However, for better approximations in the case of irregular shapes, the method of OpenCV [27] using moments was preferred. Then, for each pixel q of the organ, the Euclidean distance to the centroid was calculated, and the inner-distance map was defined as,

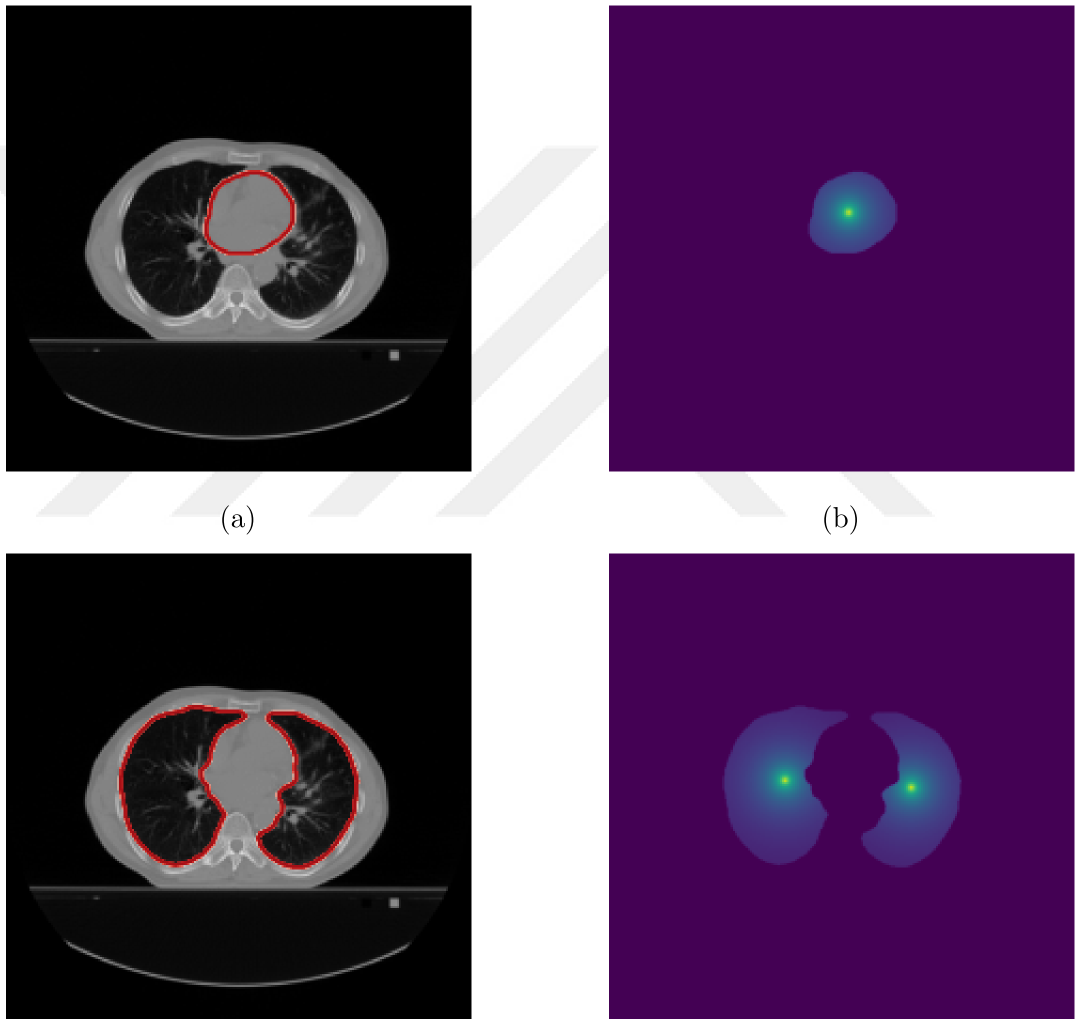
$$d_{\text{inner}}(q) = \begin{cases} \frac{1}{1+\alpha\|q-\mathcal{C}\|^2} & \text{if } q \in \mathcal{A} \\ 0 & \text{if } q \in \text{ background} \end{cases}$$

where α is the decay ratio chosen as 0.1, $\mathcal{A}$ is the set of annotated pixels of the organ, and $\mathcal{C}$ is its centroid pixel. This normalized distance map emphasizes the shape information of the organ since its highest value is towards the center of the organ and it describes the relation of each pixel with respect to the center.

It can be seen in Figure 3.1 and Figure 3.2 that the inner-distance maps are more descriptive for the heart and the lungs in comparison with the esophagus and the spine. This can be attributed to the differences in the shape and size of the organs. The inner-distance maps of the heart and the lungs, which are larger organs in size, encompass shape information by representing the relationship of each pixel value to the centroid. Even though it is the same for the esophagus and the spine, the map is not as descriptive for two reasons; firstly, the unsymmetrical shape, as in the case of the esophagus, does not provide a good centroid for calculation, secondly, the small size of both organs makes the normalized values of the distance very close to the annotation values, making this map less informative.

3.1.2 Distance-to-Landmark Organ Map

To address the limitations of the inner-distance map, this thesis proposes the distance-to-landmark organ map, which embeds more contextual information of



(a) (b)

(c) (d)

Figure 3.1: Example showing (a) the heart and (c) the lungs from a typical thorax CT scan slice, followed by their inner-distance transforms in (b) and (d). The image dimensions are the original dimensions that are 512×512.

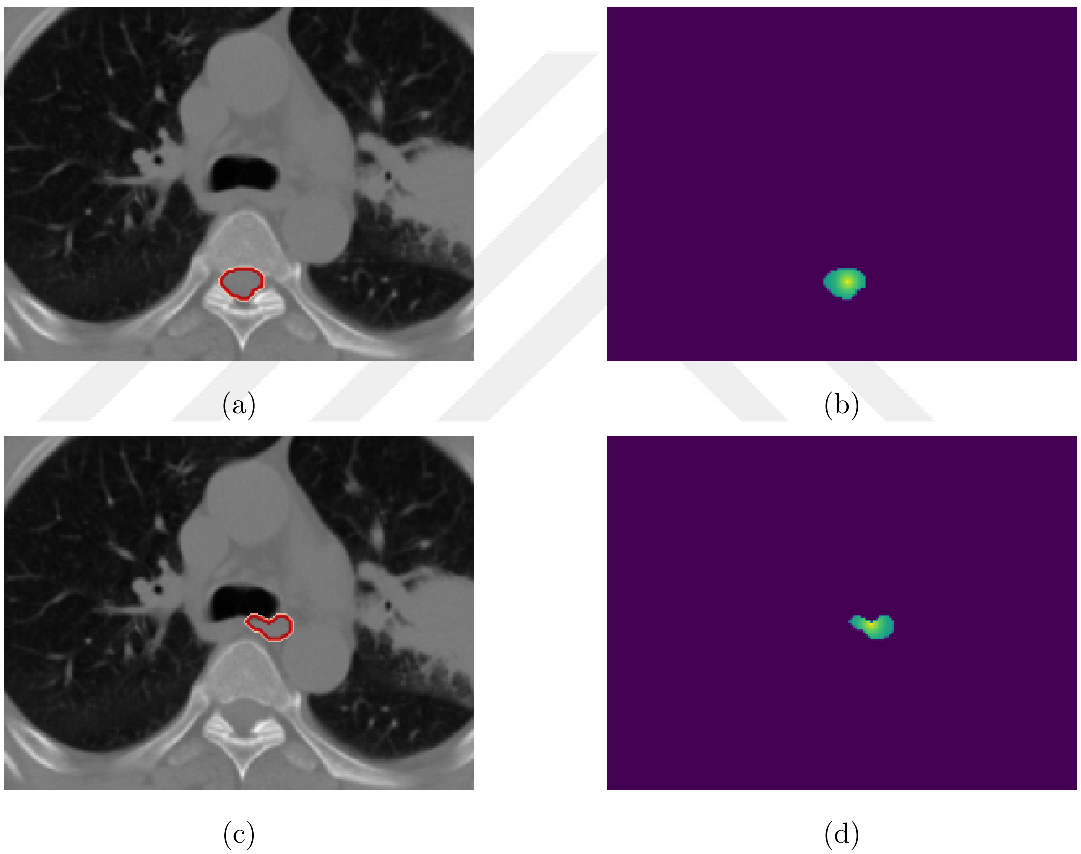


Figure 3.2: Annotation of (a) the spine and (b) its inner-distance map. (c) An irregularly shaped esophagus and (d) its inner-distance map. The image dimensions have been cropped to show the spine and the esophagus more clearly since they are small organs.

an organ. Additional spatial, location, and contextual priors can help improve the performance of machine learning models. The aim of this thesis is to use anatomical context to increase network performance. After consultation with domain experts and detailed evaluation of the dataset, certain anatomical facts were considered as potentials for context priors for the network. This included the relationship between neighboring organs in the thorax and their existence throughout the thorax, as well as their visibility in CT images. Therefore, a distance map incorporating the relationship of the organs is devised, called the distance-to-landmark map. For the two chosen organs to improve, i.e., the heart and the esophagus, the landmark organs are the lungs and the spine, respectively. The calculation of these distance maps also used the Euclidean distance. For each pixel of the target organ, its distance from the closest pixel of the landmark organ is calculated. This generates a distance map that holds information about the pixel-wise relationship of the target and landmark organ. Let $\mathcal{L} = \{x_i\}$ be the set of pixels belonging to the landmark organ then for each pixel q belonging to the organ of interest, the distance-to-landmark is calculated as follows:

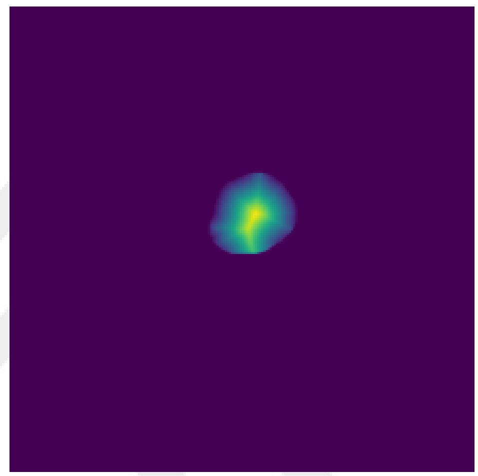
$$d_{\text{Landmark}}(q) = \begin{cases} \frac{\min_{x_i \in \mathcal{L}} \|q - x_i\|^2}{\max \min_{x_i \in \mathcal{L}} \|q - x_i\|^2} & \text{if } q \in \mathcal{A} \\ 0 & \text{if } q \in \text{ background} \end{cases}$$

where the denominator is the normalization factor. This ensures that very large distances do not skew the distance transform map so that during training an accurate relationship of the pixels is learned by the model. An example of these maps is given in Figure 3.5.

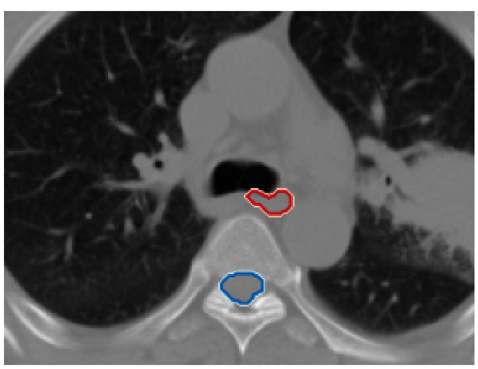
Due to the high contrast of the lungs and the spine in the thorax CT, they were chosen as landmark organs. The lungs are suitable landmark organs for the heart since the heart is always located between the two lungs, however, as the lungs begin to obfuscate in basal slices as shown in Figure 3.4, the existing networks find it harder to segment the heart. In this scenario, the distance-to-lungs map of the heart, emphasize the location of the heart (see Figure 3.5) with respect to the lungs, hence improving the inference ability of the network.

The landmark organ chosen for the esophagus is the spine. The esophagus

(a)



(b)



(c)

(d)

Figure 3.3: (a) Annotation of the heart and the lungs and (b) the corresponding heart-to-lungs distance map. (c) Annotation of the esophagus and the spine along with (d) its esophagus-to-spine map. The image dimensions have been cropped to show the spine and the esophagus more clearly since they are small organs.

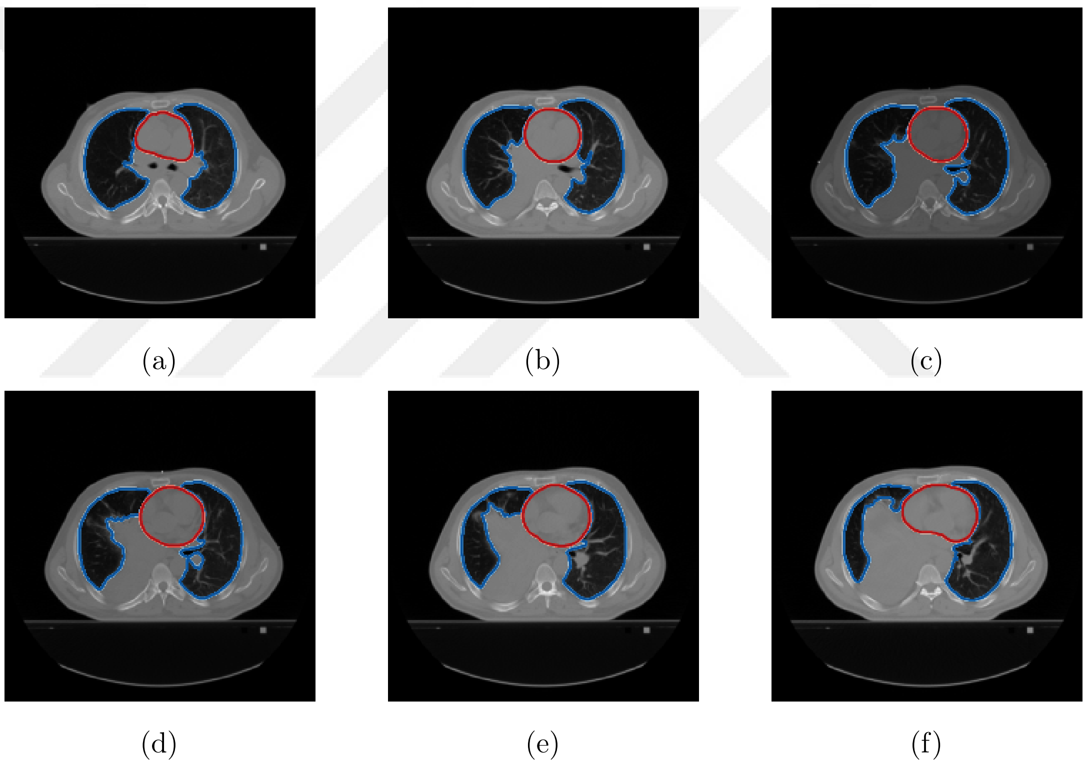
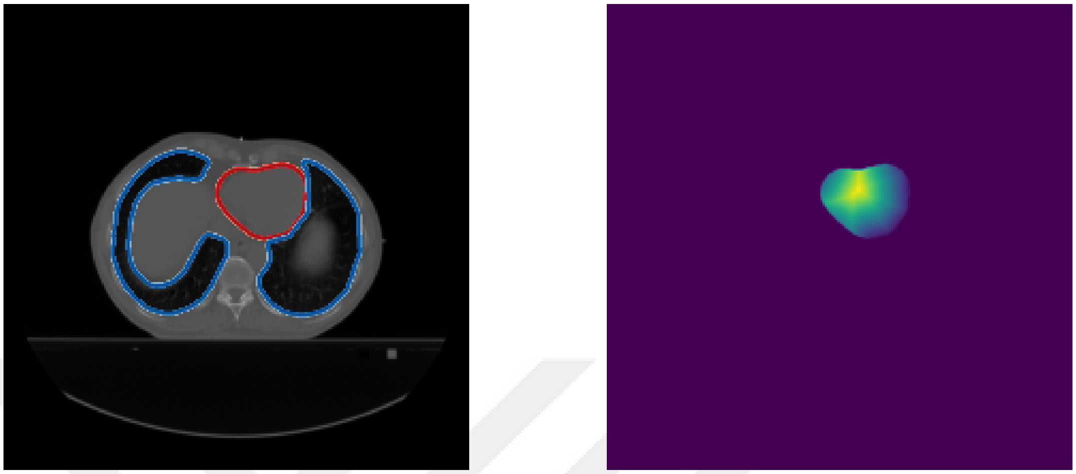


Figure 3.4: Annotations of the heart and the lungs in consecutive basal slices of the thorax CT. In order from (a) to (f), these slices show how the lungs and the heart change their morphology as they get closer to the thoracic boundaries.



(a) (b)

Figure 3.5: (a) An example showing a basal slice of the thorax CT where the left lung is obfuscated as the abdomen organs are becoming visible. (b) In such a scenario, the heart-to-the lungs distance transform map holds valuable anatomical context.

is one of the most difficult to segment organs in CT scans [3][26][23]. It is extremely prone to inter- and intra-observer variability due to its poor contrast in CT images and its variation in shape throughout the slices. The esophagus also extends beyond the thorax in both directions, i.e., towards the head and neck from the apex of the thorax, as well as towards the abdomen from the base of the thorax. This provided another motivation to choose the spine as the esophagus' landmark organ, since the spine also extends beyond the thorax, thus it does not obfuscate in the edge slices, unlike the lungs. Careful analysis of the dataset showed that the esophagus and the spine maintain a consistent distance between each other even as the morphology shifts slice by slice. The distance between the two organs is also more than the distance of the esophagus' pixels to its centroid, and, the normalized distance map values are more informative than the pixel-wise annotations.

3.2 Network Architecture

In this section, we will talk about the two networks that we used for OAR segmentation: single-task regression network and multi-task network. Both networks have UNet inspired architectures with an encoder path and a decoder path for feature extraction and then consequent reconstruction of the image to output a 512×512 map (see Figure 3.6). The encoder path of the UNet architecture is responsible for high level feature extraction, and in recent times, residual learning approaches such as ResNet [2] have gained popularity for maintaining gradient flow in deep architectures that are designed for feature extraction by having a residual connection that carries the element-wise sum, as shown in Figure 3.7. This ensures that even in the deep layers in the network, the gradient does not reach zero. This thesis combines these two powerful models to obtain a ResNet backbone UNet, *Res-UNet*, that performs regression analysis to predict the distance transform maps. In the Res-UNet, the encoder blocks of the UNet are replaced with encoder and skip connection blocks, making the architecture more robust.

The flow of the single task regression is shown in Figure 3.8. This network is trained using the distance maps. Hence, it is a regression model and so it uses mean squared error (MSE) as the loss function. The network produces the distance maps as the output of the linear layer. Due to the encoder-decoder structure of this network. The output is the same dimension as the input. Since the goal is to obtain organ segmentation, the single post-processing step for this network is using a threshold for the distance maps to obtain the pixel-wise segmentation map.

The second architecture is leveraging multi-task learning to use the distance maps not only as carriers of contextual and prior information, but also as regularizers for the segmentation task. This network contains a shared encoder, again with a ResNet backbone, and two decoders, one being a segmentation task that has a softmax layer at the end to output the probabilities of the segmentation and uses categorical cross-entropy loss during training, the second being

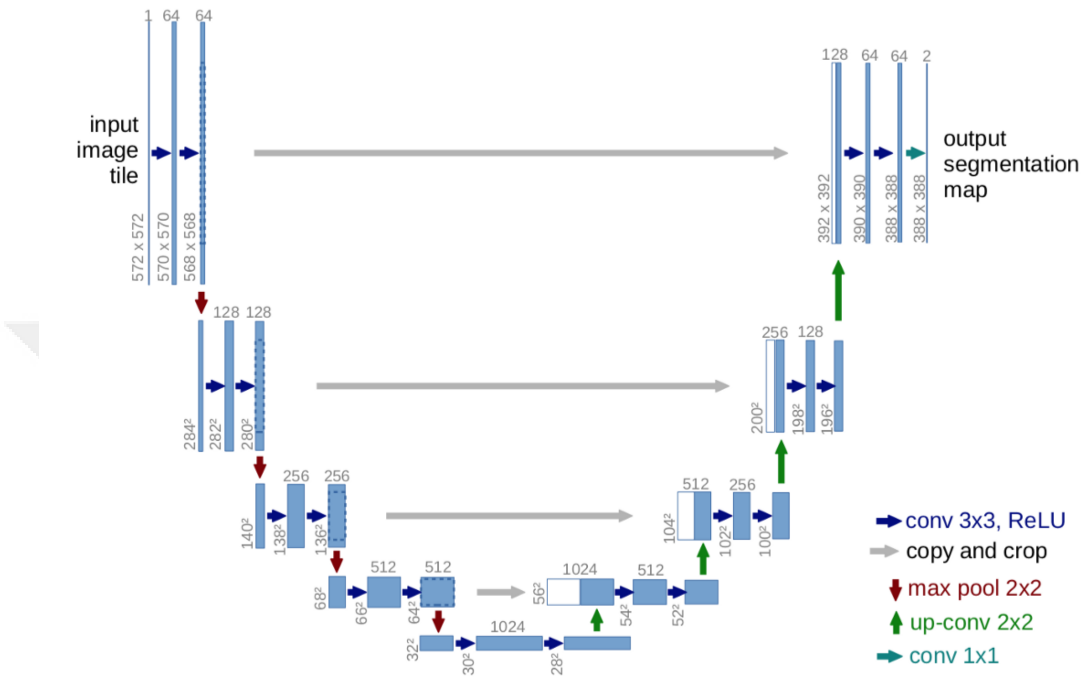


Figure 3.6: The original UNet architecture [1].

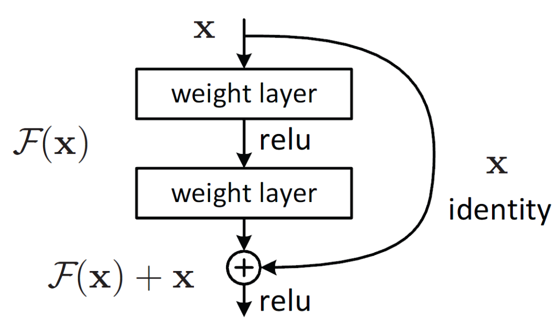


Figure 3.7: A residual learning building block [2].

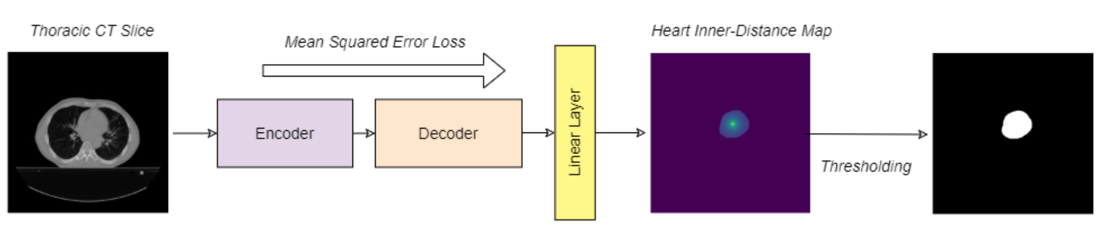


Figure 3.8: Illustration of the network design for the single-task deep regression model that uses an encoder and decoder to obtain reconstructed prediction maps, that can then be used to obtain the organ segmentation after thresholding.

a regression task that uses the mean squared error loss function during training. This is shown in Figure 3.9. Apart from conventional hyperparameters, this network has an additional setting of weights of the branches which is chosen after experimentation.

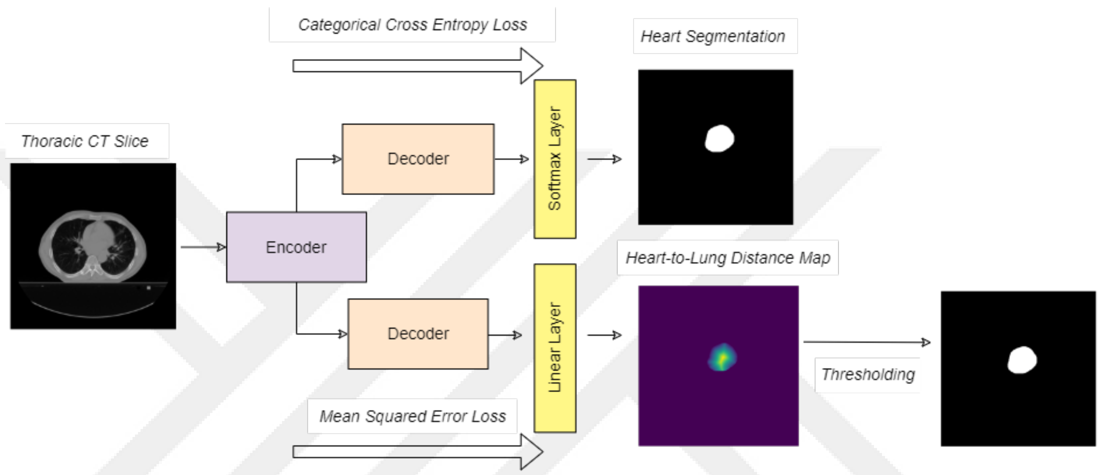


Figure 3.9: Illustration of the network design for the multi task framework that uses a shared encoder and two decoders to obtain the reconstructed segmentation maps as one output and the prediction maps as the second output.

The architectures use four ResNet encoder blocks and then four decoder blocks for each of the decoders. The ResNet chosen for the backbone is the ResNet-34 architecture, and its pretrained weights on the ImageNet were used for the network [28]. Experiments that trained from scratch offered no difference compared to the pretrained backbone therefore, following convention, the pretrained weights were used. The feature maps extracted and then reconstructed using skip connections are shown in Figure 3.10 and Figure 3.11.

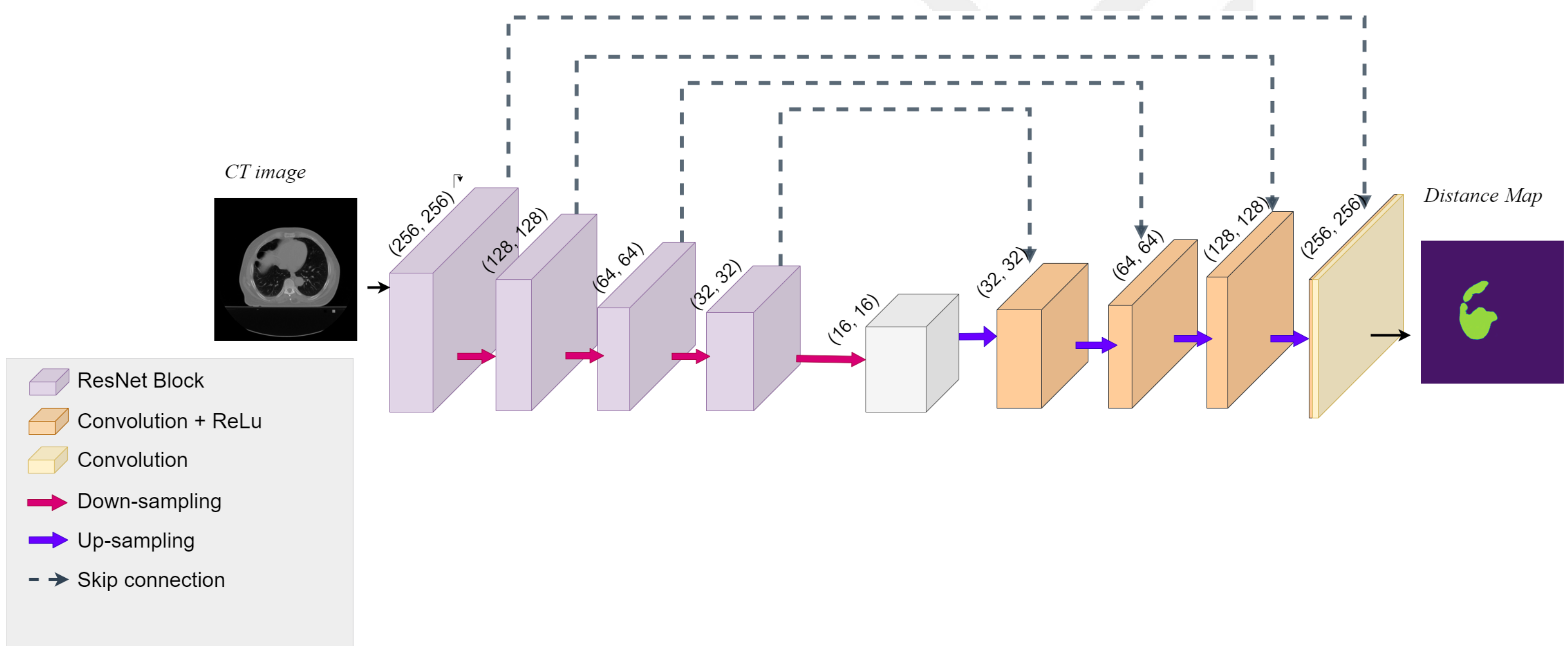


Figure 3.10: Res-UNet single-task architecture. This is the UNet inspired single-task architecture where the encoder blocks have been replaced with ResNet blocks that are being taken from a pretrained ResNet-34 architecture, followed by regular decoder blocks.

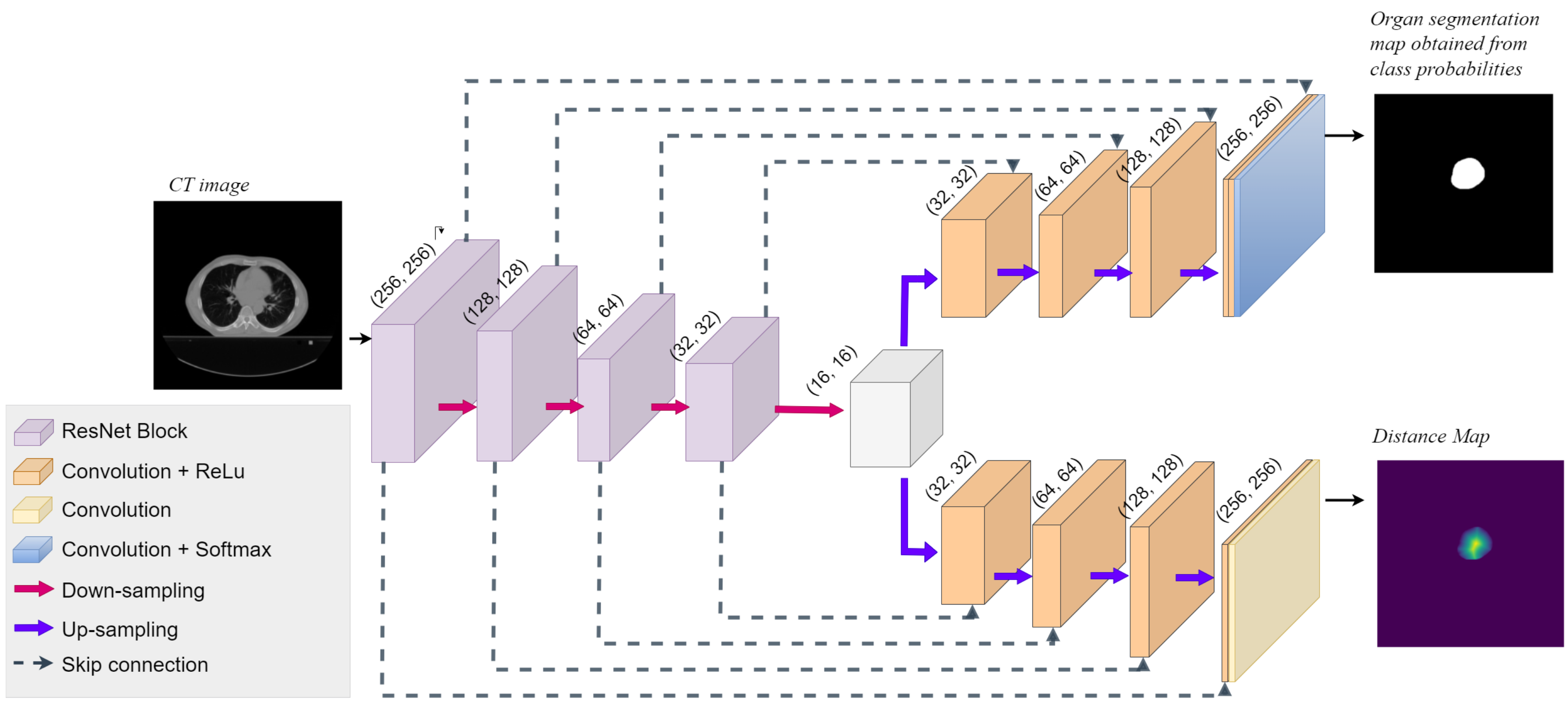


Figure 3.11: Res-UNet multi-task architecture. This is the UNet inspired multi-task architecture where the encoder blocks have been replaced with ResNet blocks that are being taken from a pretrained ResNet-34 architecture, followed by regular decoder blocks making two decoders. The first decoder path is a segmentation branch, and the second is a regression branch.

Chapter 4

Experiments and Results

4.1 Dataset

The dataset used in this study was obtained from the Radiation Oncology Department at Eskişehir Osmangazi University School of Medicine. This dataset has computed tomography (CT) images from the medical records of 265 patients. The CT scans were carried out using a Siemens SOMATOM Definition AS. The CT scans are 512×512 pixels in size, with an in-plane resolution varying between 0.76 to 0.98 mm per pixel. The number of slices varies from 80 to 115 with a z-resolution of 5 mm. The most common resolution is $0.98 \times 0.98 \times 5$ mm^3. This dataset (265 patients) was randomly split into a training set of 50 patients, a validation set of 15 patients and a test set of 200 patients. The ground truth for OARs was delineated by a team of experienced radiation oncologists. Due to the nature of these images, no augmentation techniques were used.

4.2 Network Training

The networks have been implemented using PyTorch [29]. They use Adam optimizer with a step size of 0.0001 for training. The epochs are arbitrary for each run since early stopping is used to prevent overfitting. The early stopping criterion monitors the validation loss and finishes the training if the validation loss is unchanged for 30 epochs. For the multi-task network, the weight of each task is kept equal. The trainings utilized a Tesla T4 GPU.

4.3 Evaluation Metrics

The results of these segmentations are reported for the 200 patient test set. The quantitative metrics used to report the results are average precision, average f-score, average HD95, and mean surface distance. These metrics are calculated for each of the test set images and then the mean is reported. Metrics such as precision and f-score are chosen since they are better at reporting the model performance for datasets with class imbalances. In this case, the background pixels are higher in number than the organ pixels, and therefore, these metrics show how successfully the model was able to predict the relevant pixels, i.e., the organ pixels. The definition of precision, recall and f-score are given below where TP, FP, and FN denote the number of true positive, false positive and false negative pixels respectively.

$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F-score} = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

However, these metrics still do not fully grasp the model performance for these tasks, especially since apart from segmentation overlap, the location of the estimated pixel is also necessary for a fair assessment. Therefore, two metrics inspired by the metrics used in the American Association of Physicists in Medicine (AAPM) segmentation challenge are also reported. First is the HD95, which utilizes the directed Hausdorff distance (HD). HD measures the distance of a pixel in segmentation A to the closest pixel in segmentation B. We perform this in both directions, i.e., estimated segmentation to ground truth and vice versa. Then we use the maximum of HD reported in either direction to make it a symmetric measurement. Since HD is extremely sensitive to outliers, we report the 95th percentile of the distances calculated.

$$\text{HD 95} = \frac{\vec{d}_{H,95}(A,B) + \vec{d}_{H,95}(B,A)}{2}$$

The second metric is the mean symmetric surface distance that uses directed mean surface distance in both directions and then takes the average of the obtained distances.

$$\text{MSD} = \frac{\vec{d}_{H,avg}(A,B) + \vec{d}_{H,\text{ avg}}(B,A)}{2}$$

where directed mean surface distance is $\vec{d}_{H,avg}(A,B) = \frac{1}{|A|}\sum_{a\in|A|}\min_{b\in|B|} d(a,b)$, calculating the average distance of a point in A to its nearest neighbor in B.

4.4 Comparisons

The proposed model for heart and esophagus segmentation is the multi-task network that uses the annotations and distance-to-landmark organ maps for training. This proposed model is compared to the standard, segmentation Res-UNet and regression Res-UNet. The proposed model for the lungs is the single task regression Res-UNet that is trained using inner-distance maps. It is compared against standard Res-UNet and the multi-task network. Similar results for the spine have also been reported.

The experiments performed for the lungs and the spine are as follows:

- Single-task segmentation: Res-UNet trained using annotations.
- Single-task regression: Res-UNet trained using inner-distance maps.
- Multi-task segmentation: Multitask Res-UNet trained using annotations and inner-distance maps.

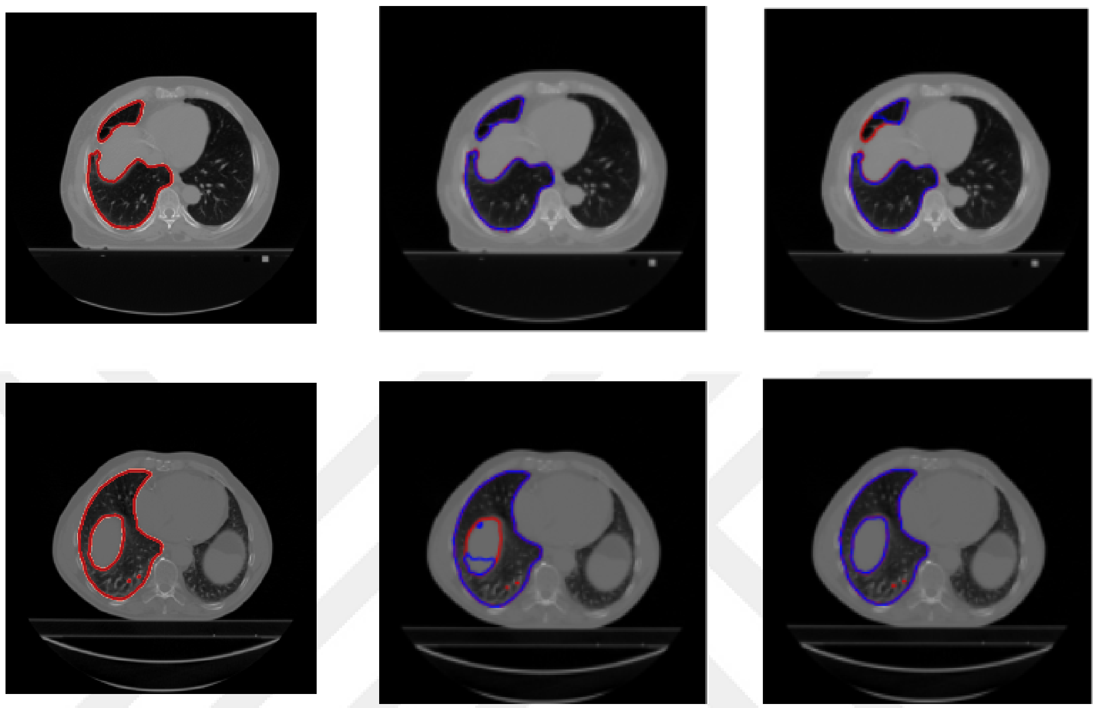
The experiments performed for the heart and the esophagus are as follows:

- Single-task segmentation: Res-UNet trained using annotations.
- Single-task regression: Res-UNet trained using inner-distance maps.
- Single-task regression: Res-UNet trained using distance-to-landmark organ maps.
- Multi-task segmentation: Multitask Res-UNet trained using annotations and distance-to-landmark organ maps.

Furthermore, the dataset from the 2017 AAPM Thoracic Auto-segmentation Challenges was used to train and test the proposed models as well as the comparison models to illustrate that use of distance maps yields improved results regardless of the dataset. The dataset has annotations for five OARs; the heart, esophagus, spine, right lung, and left lung. The details of the dataset can be found in [3].

4.5 Results and Discussion

The results of the performed experiments are reported in Table 4.1 and Table 4.2 for the five OARs of the thorax. As shown in this quantitative analysis, the proposed model is superior in f-score and distance metrics compared to the baseline models for all organs except for the spine.



(a) (b) (c)

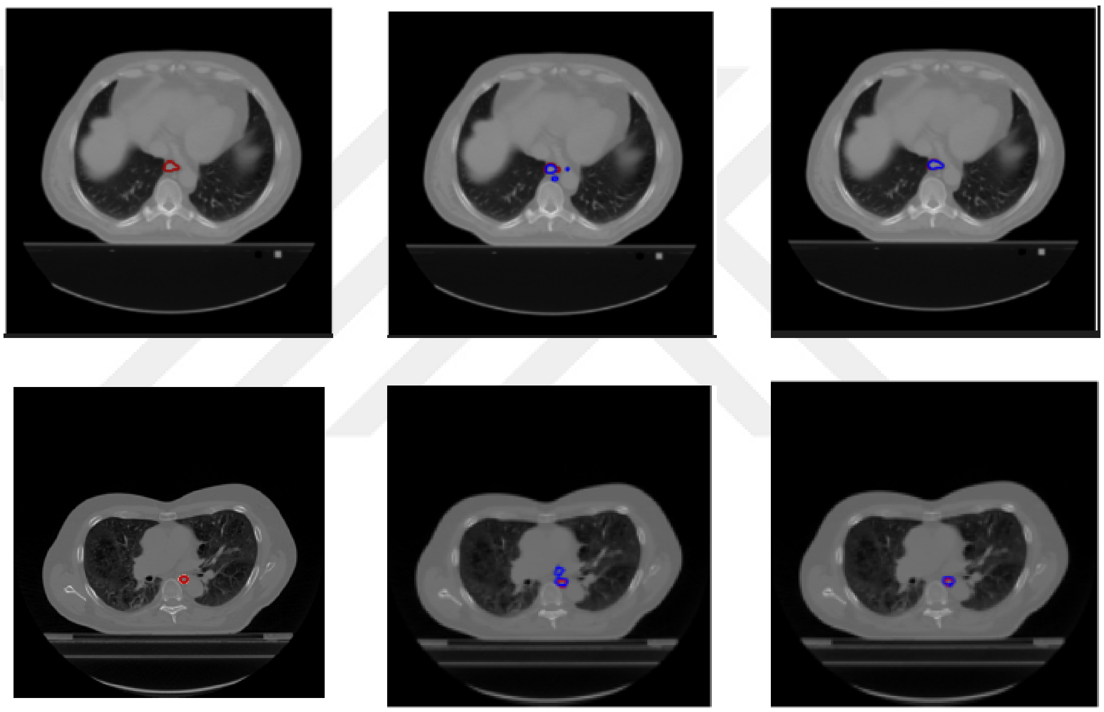
Figure 4.1: (a) An example showing the gold standard of the right lung along with (b) the segmentation prediction of the single-task segmentation Res-UNet in and (c) the segmentation map obtained from single task regression using the inner-distance map. These test results demonstrate that distance maps are particularly helpful in the slices where the organ is partially present. The first row image is from our dataset and the second row image is from the AAPM dataset [3].

The inner-distance maps computed for the lungs have successfully increased the f-score of lungs by 1 percent. Visual results shown in Figure 4.1 demonstrate that using distance maps is particularly more helpful in segmenting the oddly shaped lungs where a normal UNet fails. Using a multi-task network has not been particularly helpful for lung segmentation. Therefore we propose the single-task regression model to keep the parameters, and hence, the cost to a minimum.

Moving on to soft tissue organs; it can be seen that our model outperforms the baseline model for both the esophagus and the heart. These results are statistically significant with a $p < 0.05$ verified using a student's t-test on the f-scores of the test set samples. In the case of the esophagus, the most successful model is a multi-task architecture that uses the annotations and the distance-to-spine maps. The increase in the f-score of the esophagus is particularly meaningful since it is one of the lowest contrast organs therefore it is harder to achieve higher segmentation scores. The distance-to-spine map has outperformed the inner-distance map, thus verifying that greater contextual information is embedded in the distance-to-landmark organ map, as evident from the predicted segmentations in Figure 4.2. The spine as a landmark organ to the esophagus makes sense for many reasons. Firstly, both organs exist beyond the scope of the thorax. It is not useful to use the lungs as a landmark organ for the esophagus because about 50 percent of the slices containing the esophagus also have the lungs whereas 100 percent of all esophagus containing slices have the spine. The spine consistently emerges as one of the easiest to segment organs and a visual analysis of the slices with the esophagus and the spine shows that both are in close proximity to each other, and therefore, using the distance-to-spine can provide valuable information about the location of the esophagus.

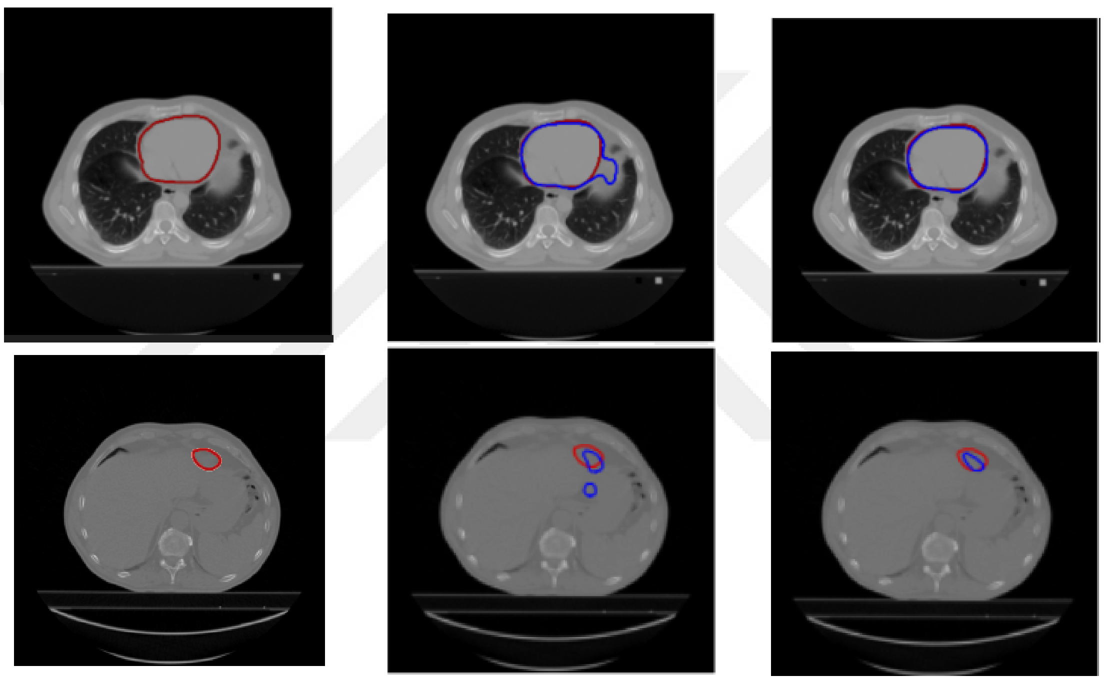
For the heart, the obvious choice of a landmark organ is to use the two lungs since anatomically the heart is always between the two lungs. However, as shown in Figure 4.3, some basal slices of the thoracic cavity may have the heart but not the lungs or may have a very obfuscated lung. The performance of a simple UNet is particularly low in these slices as it is prone to predicting large extraneous regions beyond the true boundary. Slices with the heart but without the lungs are less than 10 percent and in those cases, the distance-to-lungs map has been replaced by the inner-distance map during training. The use of this regression auxiliary task contributes to the regularization of the main segmentation task. As seen in Figure 4.4, it is apparent that the proposed model learns context since it performs a better reconstruction of the 3D volume using the predicted axial slices. The proposed model also eliminates any secondary components that the baseline model is prone to predict. Although, in theory, they can be removed after post processing, the aim of the deep learning model is to be fully automatic end-to-end, and therefore, we try to minimize any uses of pre- or post-processing.

The spine is a small organ and has a higher contrast in CT images due to



(a) (b) (c)

Figure 4.2: (a) An example showing the gold standard of the esophagus along with (b) the segmentation prediction of the single-task segmentation Res-UNet in and (c) the segmentation map obtained from the multi-task network that used distance-to-spine. Even though the secondary components can be removed by post processing, it is relevant to notice that the predicted shape of the largest component in (b) is not as similar to the annotation as in (c). The first row image is from our dataset and the second row image is from the AAPM dataset [3].



(a) (b) (c)

Figure 4.3: (a) An example showing the gold standard of the heart along with (b) the segmentation prediction of the single-task segmentation Res-UNet in and (c) the segmentation map obtained from the multi-task network that used distance-to-lungs. Due to the discrepancy in the shape of the lung the, simple UNet is inclined to predict an incorrect boundary but such a problem is solved by the proposed network that leverages the heart-to-lungs relationship. The first row image is from our dataset and the second row image is from the AAPM dataset [3].

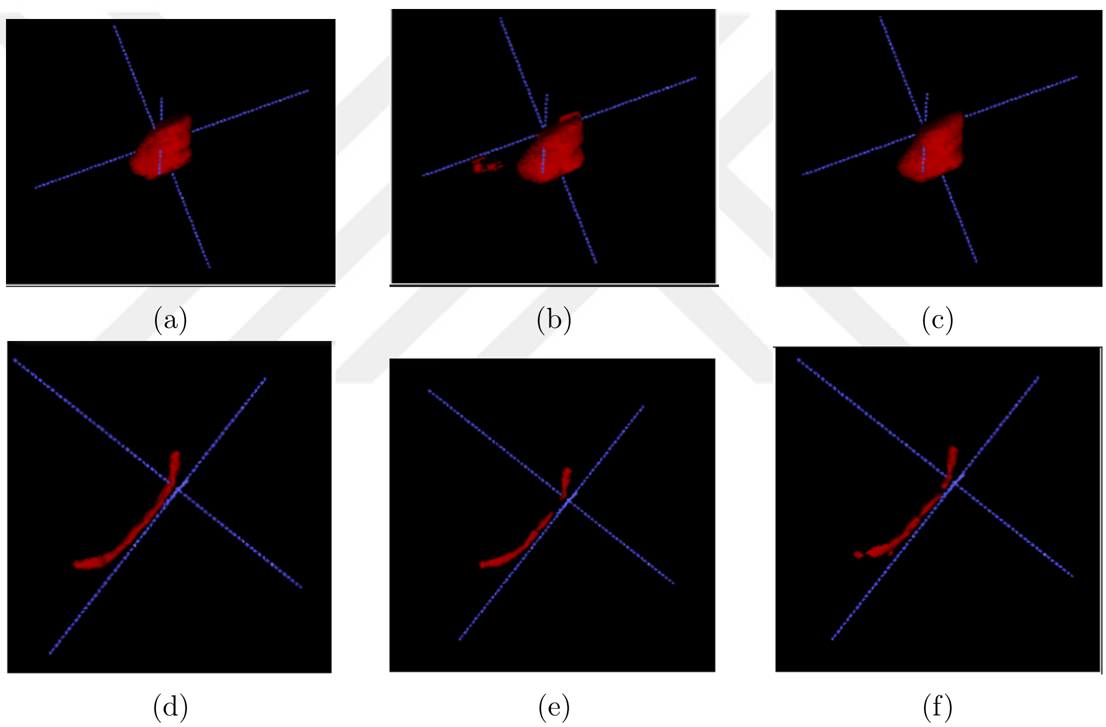


Figure 4.4: 3D reconstruction of the heart and the esophagus from the axial slices using ITK-Snap [4]. Here (a) and (d) correspond to the gold standard of the heart and the esophagus respectively, (b) and (e) are their predictions from the UNet model, and (c) and (f) are their predicted segmentation maps using the proposed distance-to-landmark organ maps.

bone density. Thus, it is already an easy to segment organ as reported in many previous studies [3] [19]. The inner-distance maps for the spine did not provide a considerable improvement owing to the size of the organ. Since it is a small round organ, the normalized inner-distance map is very close quantitatively to the annotation thus both networks have similar results. This shows that inner-distance maps are not suitable for small organs. However, for large organs such as the heart and the lungs, they can prove to be very informative since they contain shape information. Since the spine spans more regions of the body than just the thorax, it can be used in the same distance-to-landmark organ transform for other organs in the head and neck or abdomen region CT scans.

Table 4.1: Quantitative results obtained by the proposed models and the comparison algorithms. These are averages of three runs and the standard deviation is reported. These are the test set results of our thorax CT dataset.

(a) Heart

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	90.23 ±0.44	90.35 ±0.78	89.31 ±0.14	4.08 ±.03	3.66 ±0.06
Single-task (inner-distance)	**90.27 ± 0.37**	90.80 ±0.56	89.68 ±0.18	**3.98 ± 0.04**	3.79 ±0.02
Single-task (distance-to-lungs)	90.04 ±0.49	91.31 ±0.72	89.69 ±0.53	4.02 ±0.04	3.54 ±0.05
Multi-task (distance-to-lungs)	89.94 ±0.47	**91.38 ± 0.58**	**89.79 ± 0.05**	4.02 ±0.04	**3.42 ± 0.05**

(b) Esophagus

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	79.65±0.82	71.87±0.21	73.80±0.48	2.243±0.003	2.084±0.004
Single-task (inner-distance)	**79.85±0.15**	70.69±0.65	73.02±0.81	2.235±0.002	2.228.002
Single-task (distance-to-spine)	78.88±0.60	72.30 ±0.51	74.10±0.21	2.256±0.011	2.272 ±0.001
Multi-task (distance-to-spine)	78.99±0.57	**72.33±0.46**	**74.30±0.46**	**2.231±0.011**	**1.341±0.002**

(c) Right lung

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	95.80 ±0.03	93.09 ±0.45	93.95 ±0.26	2.91 ±0.03	2.01±0.01
Single-task (inner-distance)	**95.91 ± 0.15**	**94.58 ± 0.36**	**94.80 ± 0.13**	**2.69 ± 0.08**	**1.52± 0.12**
Multi-task (inner-distance)	94.12±0.05	91.12±0.45	91.98±0.15	3.17±0.10	2.99±0.12

(d) Left lung

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	95.81±0.23	93.68±0.47	94.12±0.33	2.67±0.06	2.30±0.20
Single-task (inner-distance)	**95.88±0.12**	**94.71±0.22**	**94.87±0.05**	**2.59±0.01**	**1.48±0.15**
Multi-task (inner-distance)	95.21±0.10	94.52±0.32	94.83±0.10	2.60±0.05	1.92±0.12

(e) Spine

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	95.06.12	**91.26±0.35**	**92.84±0.22**	1.636±0.014	0.611±0.110
Single-task (inner-distance)	**95.27±0.10**	90.86±0.32	92.72±0.12	**1.618±0.023**	**0.561±0.120**
Multi-task (inner-distance)	95.35±0.20	90.78±0.34	92.73±0.11	1.625±0.022	0.584±0.120

Table 4.2: Quantitative results obtained by the proposed models. These metrics are reported for a single run. These are the results obtained on the 2017 AAPM Thoracic Auto-segmentation Challenge dataset [3].

(a) Heart

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	**93.34**	87.84	89.56	4.27	3.56
Multi-task (distance-to-lungs)	92.68	**89.09**	**90.13**	**4.22**	**3.48**

(b) Esophagus

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	72.55	57.14	60.82	**2.343**	2.372
Multi-task (distance-to-spine)	**74.13**	**59.55**	**63.06**	2.356	**2.226**

(c) Right lung

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	94.16	95.22	94.09	3.02	2.13
Single-task (inner-distance)	**94.88**	**96.16**	**94.65**	3.02	**1.76**

(d) Left lung

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	**96.73**	92.30	93.82	2.78	1.55
Single-task (inner-distance)	94.58	**97.08**	**95.51**	**2.71**	**1.16**

(e) Spine

	Precision	Recall	F-score	HD95	MSD
Single-task (segmentation)	**81.92**	91.12	**88.34**	**1.51**	**0.81**
Single-task (inner-distance)	86.19	**92.68**	88.33	1.52	0.95

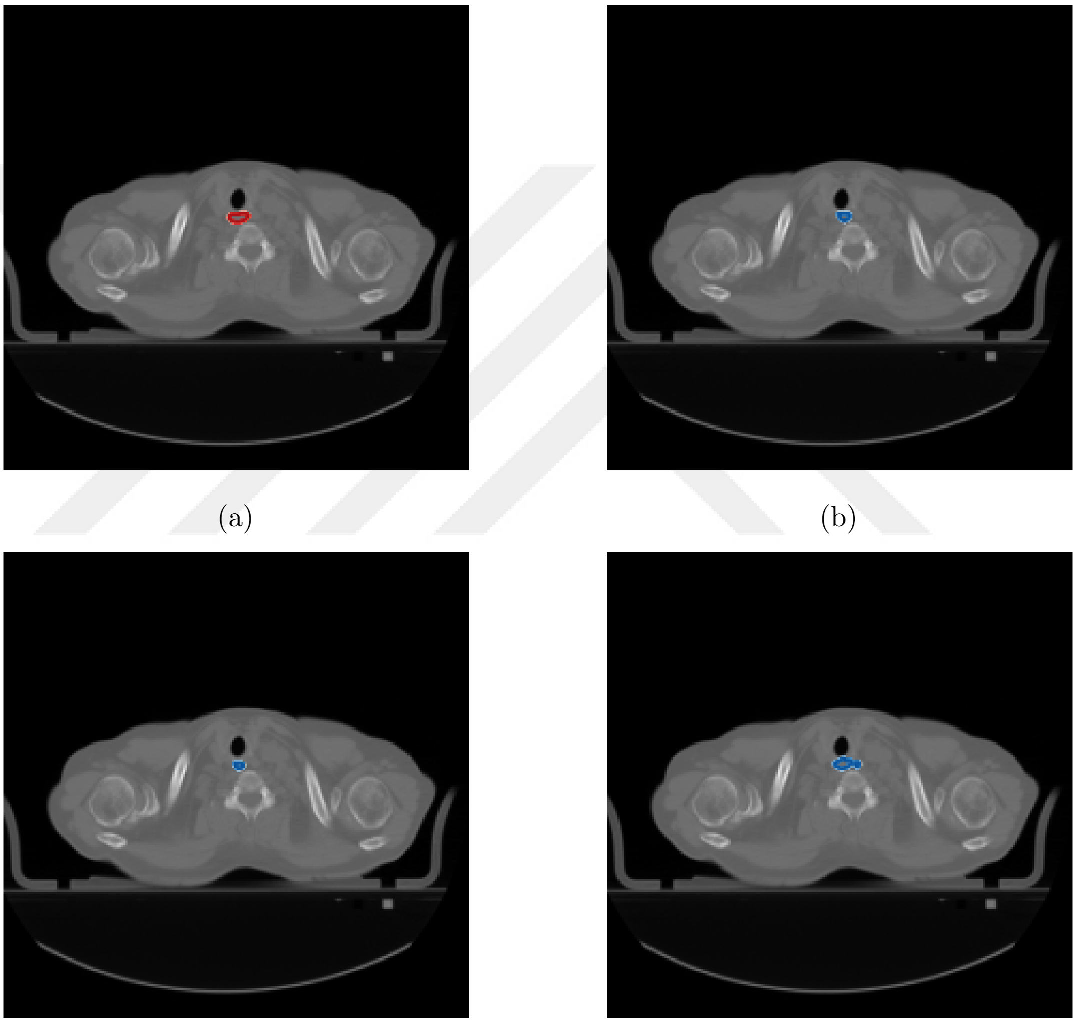
Chapter 5

Conclusion

This thesis provides a novel combination that embeds location, shape, and anatomic context to train a successful segmentation model for thoracic CT images. The quantitative and qualitative analyses show an increase in performance compared to the baseline models.

The distance-to-landmark maps are carriers of shape and anatomical context information for organs of interest and the usefulness of this is shown particularly in the apical and basal slices where the shape of the organs becomes unconventional and so using the location of the neighboring organs, the shape and location of the organ of interest is improved. The distance-to-landmark organ map is superior to the inner-distance map as evident in Figure 5.1 and Figure 5.2.

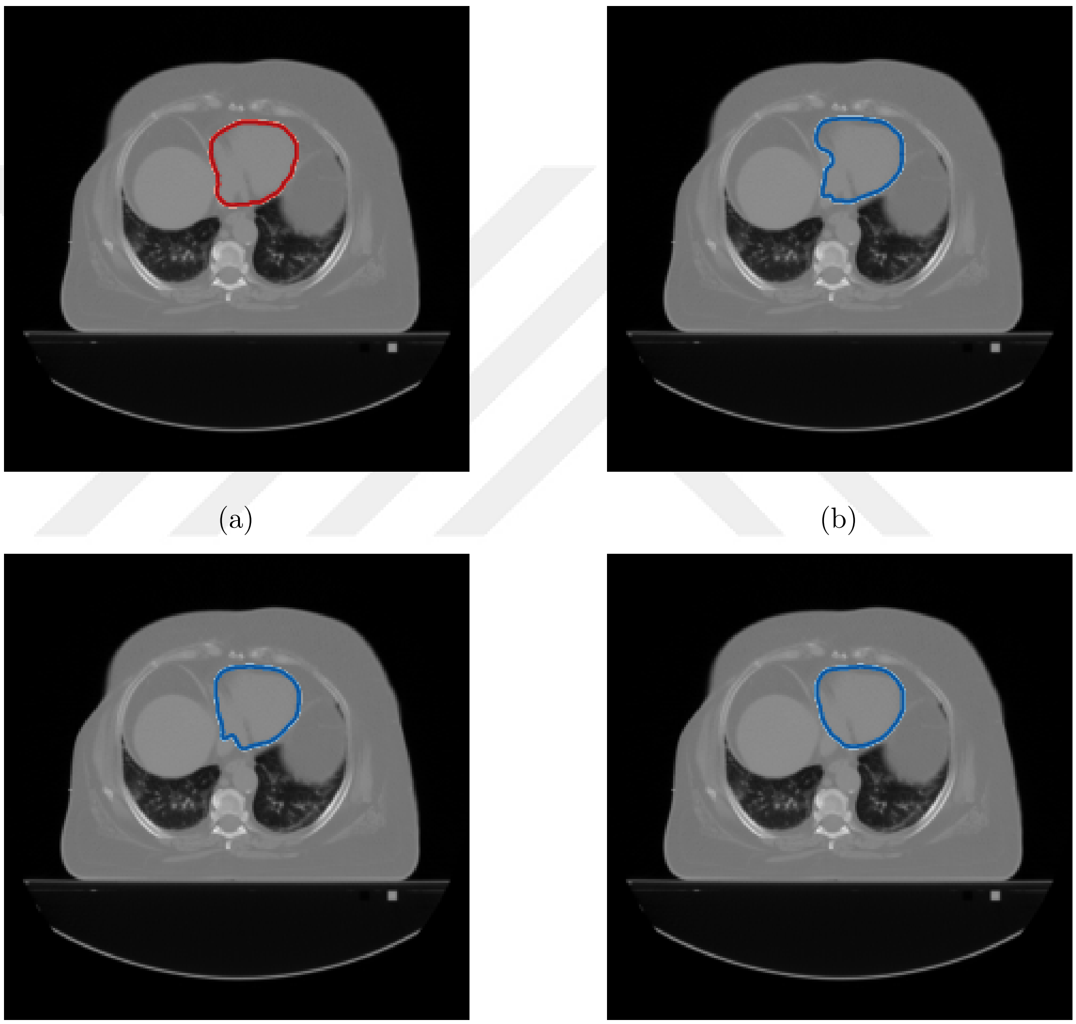
This thesis proposed the distance-to-landmark transform to make use of anatomical information as a prior for the machine learning model, to get improved results. The distance-to-landmark organ requires careful consideration of anatomical facts, and thorough analysis of the dataset, to make the right choice of the landmark organ for the required target organ. As shown in Figure 5.3, the choice of lungs as landmarks for the heart has improved the performance of the simple segmentation network for the unconventional slices where the shape of the organs deviates from their shape in the rest of the dataset. This disrupts



(a) (b)

(c) (d)

Figure 5.1: (a) Test set visualizations showing the annotation, (b) the result of the Res-UNet segmentation, (c) the result of the Res-UNet regression with inner-distance map and (d) the result of multi-task network that used distance-to-spine maps.



(a) (b)

(c) (d)

Figure 5.2: (a) Test set visualizations showing the annotation in, (b) the result of the Res-UNet segmentation, (c) the result of the Res-UNet regression with inner-distance map and (d) the result of multi-task network that used distance-to-lungs maps.

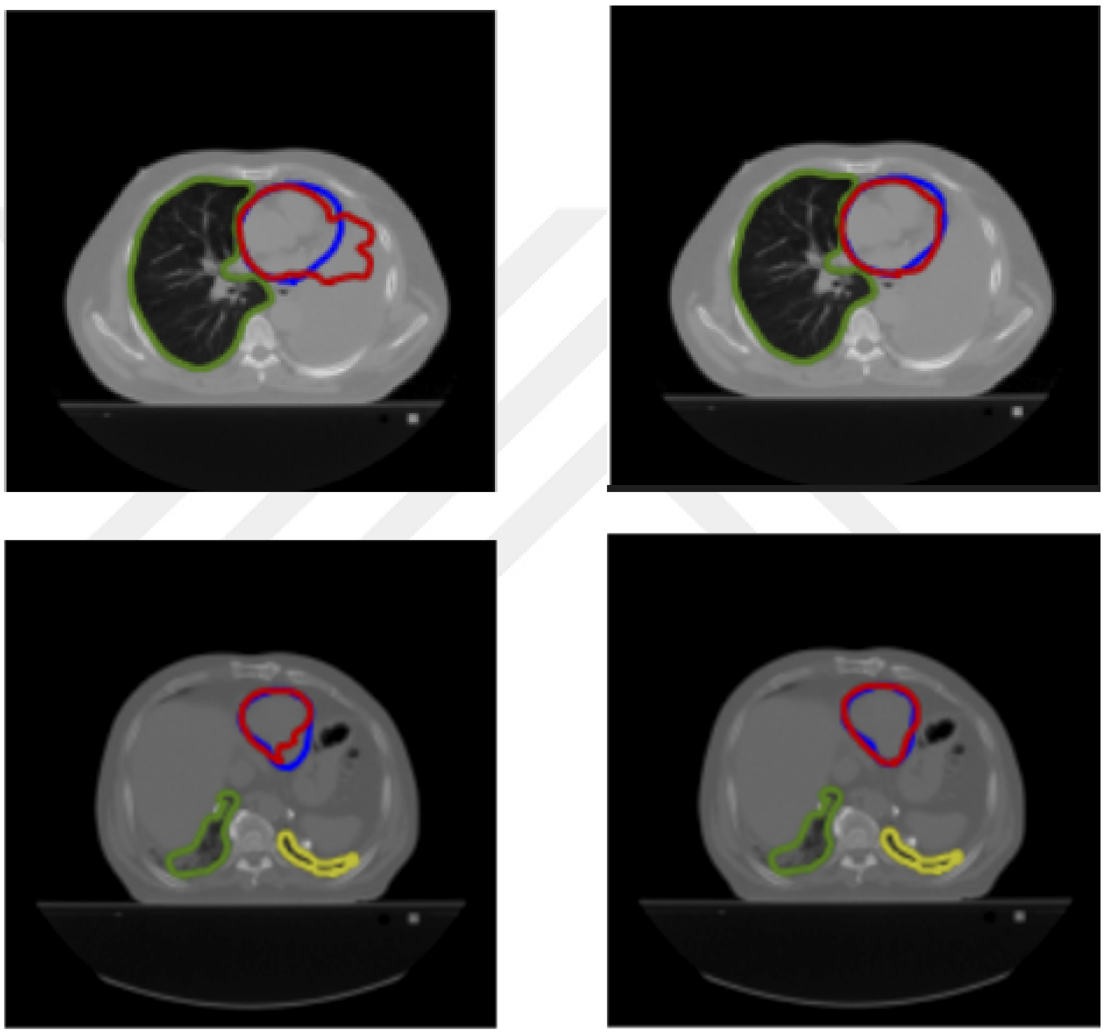
the inference capabilities of the network in such a way that the obfuscation of the lungs is causing incorrect heart segmentation. The main motivation of using a distance-to-lungs map is to give a prior to the network so that it can counteract the effect of the changing organ shapes throughout the CT scan slices.

Subsequently, the inner-distance map contains only the shape information of the larger organs, but is equally useful since it is more universal and can be generated for any organ without constraints unlike distance-to-landmark maps, which require more than one annotated organ and a relevant relationship between them.

The inner-distance map is not suitable for the esophagus for similar reasons as discussed above for the spine. However, inner-distance maps are effective for the heart segmentation in the case when the boundaries of the thorax are defined in such a way that all the CT slices used contain the heart and the lungs. In the dataset used in this thesis, the thorax is not limited to any slice of the body. In fact, these segmentations are carried out for the organ even if it spans outside the region of the thorax. This is where the distance-to-lungs map performs better as it is carrying information about the lungs' position from the previous slices and guiding the segmentation even when there is no lung. This paves way for many potential applications of this contextual information for future networks.

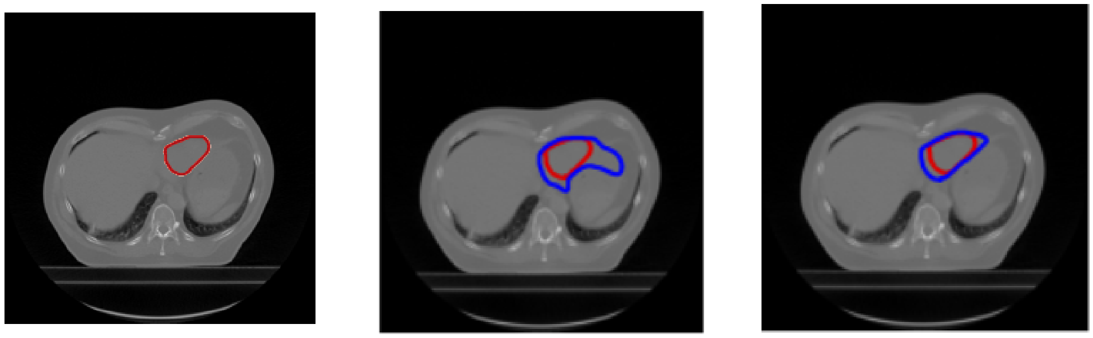
5.1 Future Work

In this thesis, for each OAR, a separate network was trained. Therefore, an apparent improvement for this thesis is to define this as a multi-class problem and perform multiple organ segmentation simultaneously. This can be done by generating distance maps for each organ and combining them together to be used in the network. The multi-class distance map can be used in the same manner as in this thesis, where it acts as shape, location, and anatomical context prior or, it can be used as part of the loss function. Since the organs occupy their own space in the image, the distance map can contribute as weights for the loss function



(a) (b)

Figure 5.3: Two examples showing annotated heart (blue) and the predicted segmentation of heart obtained (red) to show how the distance-to-lungs transform enhances the segmentation in the slices on the boundary of the thorax. The variation in shapes of the left lung (yellow) and the right lung (green) is also shown. (a) Segmentation map from the UNet predictions and (b) the segmentation maps predicted from the proposed model that uses the heart-to-lungs distance map.



(a) (b) (c)

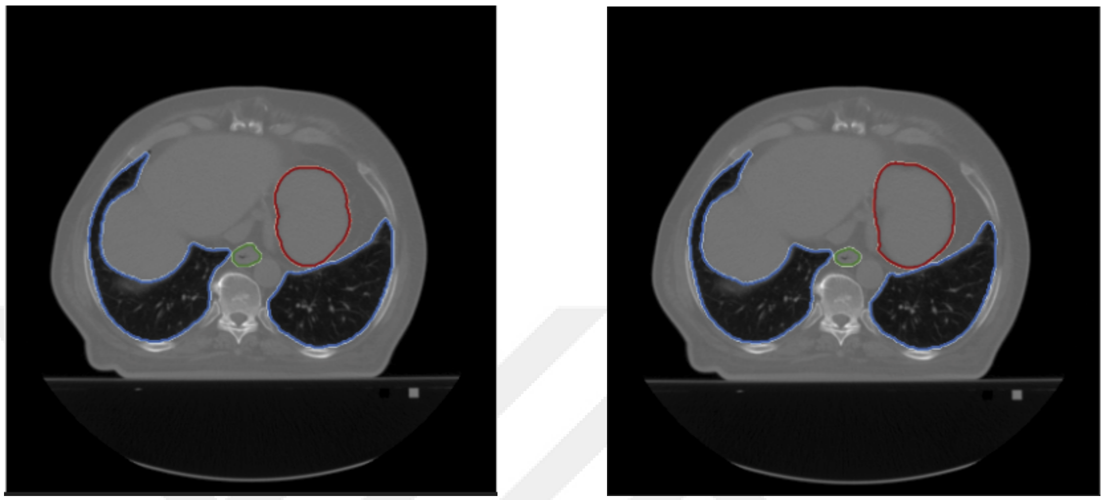
Figure 5.4: An example showing (a) the gold standard of the heart along with (b) the segmentation prediction of the single task segmentation Res-UNet and (c) the segmentation map obtained from the multi-task network that used distance-to-lungs. This image is from the AAPM dataset [3].

during training.

This work can also be extended into a cascaded network, in which the predicted segmentation maps of landmark organs can be used to calculate the distance maps that are consequently used in the second part of the cascaded network designed to segment the more difficult organs.

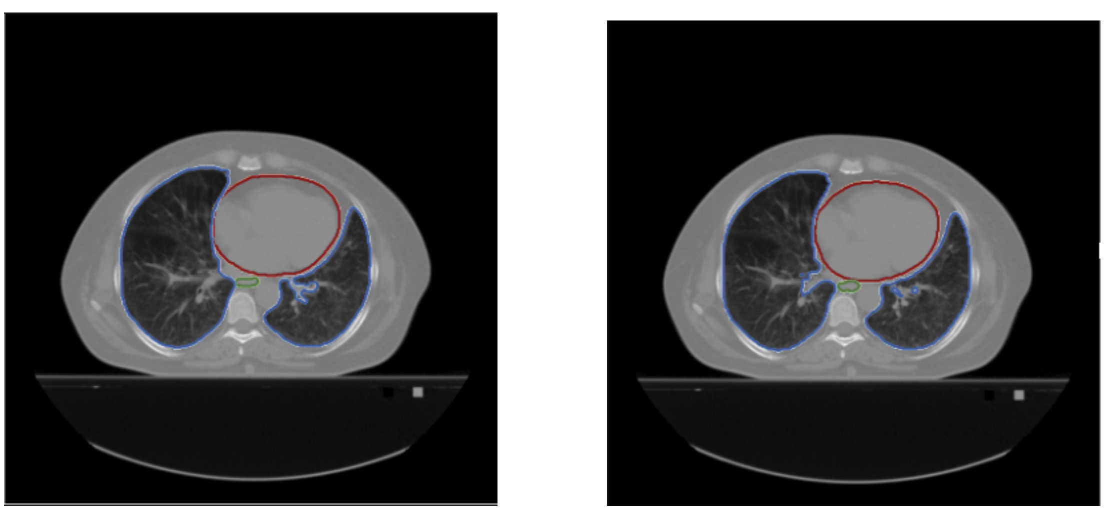
In Figure 5.4, it is shown that the distance map does indeed improve the segmentation, however, the motivation of OAR segmentation is to achieve extremely precise segmentation for accurate dosage planning during radiation therapy. With such a risk involved, the segmentation of the heart in Figure 5.4 and Figure 5.5, and the segmentation of the left lung in Figure 5.6 is still in need of improvement. This can be achieved by embedding more anatomical context by considering the relationship of each organ to its shape and location in the preceding and succeeding slices. This can be achieved by extending this thesis to the 3D space. The distance maps will have to be calculated with the respect to the voxel space occupied by the organ, and the surfaces of the 3D organs. A 3D UNet architecture can be altered to perform regression analysis for the obtained 3D distance maps to achieve more accurate segmentations.

This thesis proposed the use of distance transforms to generate meaningful



(a) (b)

Figure 5.5: (a) Ground truth annotations and (b) the predicted segmentations by the proposed models for each of the thoracic organs. While mostly the delineations are accurate, the heart is not exactly the same as the annotation.



(a) (b)

Figure 5.6: (a) Ground truth annotations and (b) the predicted segmentations by the proposed models for each of the thoracic organs. While mostly the delineations are accurate, the right lung is not exactly the same as the annotation.

auxiliary maps. These distance maps have the potential to increase a deep learning network's inference capabilities by embedding prior knowledge that can guide the training of a network. This thesis explored this hypothesis in the scope of organs-at-risk segmentation in thorax CT scans. We successfully showed that distance transforms are useful tools to quantify contextual information, which makes them meaningful to be used by a machine learning model during training. The inner-distance map showed that shape information can be embedded in the distance map, which when used in a deep learning segmentation network, produces better segmentation results than using just the annotations. More importantly, this thesis proposed the distance-to-landmark organ map, which not only embeds shape information, but contains anatomical context by quantifying the relationship between organs. This map proved to be successful specifically for cases in which the ordinary segmentation algorithms failed. The experiments were performed on our dataset and then verified on an open source dataset as well. Therefore, the use of these distance maps has the potential to further enhance all kinds of segmentations.

Bibliography

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Springer International Publishing, 2015.

[2] H. Kaiming, Z. Xiangyu, R. Shaoqing, S. Jian, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 34, pp. 770–778, 2016.

[3] J. Yang, H. Veeraraghavan, S. G. Armato III, K. Farahani, J. S. Kirby, J. Kalpathy-Kramer, W. van Elmpt, A. Dekker, X. Han, X. Feng, P. Aljabar, B. Oliveira, B. van der Heyden, L. Zamdborg, D. Lam, M. Gooding, and G. C. Sharp, "Autosegmentation for thoracic radiation treatment planning: A grand challenge at aapm 2017," *Medical Physics*, vol. 45, no. 10, pp. 4568–4581, 2018.

[4] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[5] M. Mukesh, R. Benson, R. Jena, A. Hoole, T. Roques, C. Scrase, C. Martin, G. A. Whitfield, J. Gemmill, and S. Jefferies, "Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?," *The British Journal of Radiology*, vol. 85, no. 1016, pp. e530–e536, 2012. PMID: 22815423.

[6] R. Trullo, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen, "Segmentation of organs at risk in thoracic ct images using a sharpmask architecture and conditional random fields," *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017.

[7] J. Ma, Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, C. Gaoxiang, J. Liu, C. Peng, L. Wang, Y. Wang, and J. Chen, "How distance transform maps boost segmentation cnns: An empirical study," in *Proceedings of the Third Conference on Medical Imaging with Deep Learning* (T. Arbel, I. Ben Ayed, M. de Bruijne, M. Descoteaux, H. Lombaert, and C. Pal, eds.), vol. 121 of *Proceedings of Machine Learning Research*, pp. 479–492, PMLR, 06–08 Jul 2020.

[8] N. Audebert, A. Boulch, B. Le Saux, and S. Lefèvre, "Distance transform regression for spatially-aware deep semantic segmentation," *Computer Vision and Image Understanding*, vol. 189, p. 102809, 2019.

[9] C. F. Koyuncu, G. N. Gunesli, R. Cetin-Atalay, and C. Gunduz-Demir, "Deepdistance: a multi-task deep regression model for cell detection in inverted microscopy images," *Medical Image Analysis*, vol. 63, p. 101720, 2020.

[10] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, 2022.

[11] A. A. Adjei, "Lung cancer worldwide," *Journal of Thoracic Oncology*, vol. 14, no. 6, p. 956, 2019.

[12] G. C. Pereira, M. Traughber, and R. F. Muzic, "The role of imaging in radiation therapy planning: Past, present, and future," *BioMed Research International*, vol. 2014, p. 1–9, 2014.

[13] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.

[14] S. Sun, C. Bauer, and R. Beichel, "Automated 3-d segmentation of lungs with lung cancer in ct data using a novel robust active shape model approach," *IEEE transactions on medical imaging*, vol. 31, no. 2, pp. 449–460, 2011.

[15] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3d deeply supervised network for automatic liver segmentation from ct volumes," in *International conference on medical image computing and computer-assisted intervention*, pp. 149–157, Springer, 2016.

[16] T. Fechter, S. Adebahr, D. Baltas, I. Ben Ayed, C. Desrosiers, and J. Dolz, "Esophagus segmentation in ct via 3d fully convolutional neural network and random walk," *Medical physics*, vol. 44, no. 12, pp. 6341–6352, 2017.

[17] B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks," *Medical physics*, vol. 44, no. 2, pp. 547–557, 2017.

[18] K. Men, J. Dai, and Y. Li, "Automatic segmentation of the clinical target volume and organs at risk in the planning ct for rectal cancer using deep dilated convolutional neural networks," *Medical physics*, vol. 44, no. 12, pp. 6377–6389, 2017.

[19] X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W. J. Curran, T. Liu, and X. Yang, "Automatic multiorgan segmentation in thorax ct images using u-net-gan," *Medical Physics*, vol. 46, no. 5, pp. 2157–2168, 2019.

[20] X. Feng, K. Qing, N. J. Tustison, C. H. Meyer, and Q. Chen, "Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3d images," *Medical Physics*, vol. 46, no. 5, pp. 2169–2180, 2019.

[21] X. Chen, S. Sun, N. Bai, K. Han, Q. Liu, S. Yao, H. Tang, C. Zhang, Z. Lu, Q. Huang, G. Zhao, Y. Xu, T. Chen, X. Xie, and Y. Liu, "A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy," *Radiotherapy and Oncology*, vol. 160, pp. 175–184, 2021.

[22] S. Dangi, C. A. Linte, and Z. Yaniv, "A distance map regularized cnn for cardiac cine mr image segmentation," *Medical Physics*, vol. 46, no. 12, pp. 5637–5651, 2019.

[23] Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang, "Shape-aware organ segmentation by predicting signed distance maps," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 12565–12572, 04 2020.

[24] Y. Wang, X. Wei, F. Liu, J. Chen, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Deep distance transform for tubular structure segmentation in ct scans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3833–3842, 2020.

[25] R. Trullo, C. Petitjean, B. Dubray, and S. Ruan, "Multiorgan segmentation using distance-aware adversarial networks," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014001, 2019.

[26] J. Jiang and H. Veeraraghavan, "One shot pacs: Patient specific anatomic context and shape prior aware recurrent registration-segmentation of longitudinal thoracic cone beam cts," *IEEE Transactions on Medical Imaging*, vol. 41, pp. 2021–2032, 2022.

[27] Itseez, "Open source computer vision library." https://github.com/itseez/opencv, 2015.

[28] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.