Deep Learning for Quantitative Image Analysis of Positron Emission Tomography

by

Saeedeh Afshari

M.Sc., Sharif University of Technology, 2014
B.Sc., Isfahan University of Technology, 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Spring 2019

Approval

Name:	**Saeedeh Afshari**
Degree:	**Master of Science (Computer Science)**
Title:	**Deep Learning for Quantitative Image Analysis of Positron Emission Tomography**
Examining Committee:	**Angelica Lim** Chair Assistant Professor
	Ghassan Hamarneh Senior Supervisor Professor
	Mark S. Drew Supervisor Professor
	Dr. Arman Rahmim Examiner Associate Professor University of British Columbia
Date Defended:	**March 5, 2019**

Abstract

Positron emission tomography (PET) is a popular imaging technique that produces a 3D image volume capturing functional processes within the body. In cancer studies, PET is increasingly used for diagnosis and evaluation of tumor extension, treatment planning and disease follow-up. Although PET has several limitations, including low spatial resolution and relatively low signal-to-noise ratio, it remains the modality of choice for its high sensitivity to tracer uptake in lesions. With the adoption of PET imaging, accurate segmentation and quantification of metabolic activities, specially tumor activities, is crucial and challenging because of the large variations of shape and intensity of tumor uptake patterns. In this thesis, we make two main contributions on automated tumor lesion detection and segmentation in PET. To automate segmentation, it is important to distinguish between normal active organs and activity due to abnormal tumor growth. In our first contribution, we propose a deep learning method to localize and detect normal active organs in 3D PET. Our method adapts the object detection deep convolutional neural network architecture of YOLO to detect multiple organs in 2D slices and aggregates the results to produce semantically labeled 3D bounding boxes. We evaluate our method on 479 18F-FDG PET scans and show promising results compared to the state-of-the-art organ localization methods. The second contribution addresses the challenge of creating accurate ground truth segmentation maps for training machine learning approaches for tumor delineation. We propose a fully convolutional network model to automatically delineate tumor regions in PET (i.e., indicates the border of cancerous lesions) while relying on weak bounding boxes annotations. To achieve this, we propose a novel loss function that dynamically combines a supervised component, designed to leverage the training bounding boxes, with an unsupervised component, inspired by the Mumford-Shah piecewise constant level-set image segmentation model. The model is trained end-to-end with the proposed differentiable loss function and is validated on a public clinical of 57 PET scans of head and neck tumors. Using only bounding box annotations as supervision, our model achieves results competitive with state-of-the-art supervised and semi-automatic segmentation approaches.

Keywords: Positron emission tomography, cancer, tumor, organ, objectd etection, object localization, Segmentation, Weak annotation, Deep Learning, Fully convolutional network

Dedication

I am dedicating this thesis to my beloved people who have meant and continue to mean so much to me.

Acknowledgements

I would like to thank Prof. Ghassan Hamarneh, my senior supervisor for his support, guidance and helpful suggestions during my study and completion of my thesis. I am also very grateful to my examiners, Dr. Drew, Dr. Rahmim and Dr. Lim, for their time and valuable feedback on this thesis.

I would further like to thank my lab members of the Medical Image Analysis Lab specially Dr. Aïcha BenTaieb, Zahra Mirikharaji, Saeed Izadi and Payam Ahmadvand for their support and sharing their experience and ideas with me.

Finally, I am very thankful to my family, my mother and my father for their endless support and unconditional love, and to my best friend and companion, Amirmasoud, who has given me all his inspiration and encouragement I needed during this step of my life.

Table of Contents

Approval ii

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Tables viii

List of Figures ix

List of Acronyms xi

List of Notations xii

1 Introduction 1

1.1 Background and Motivation 1

1.2 Thesis Contributions 4

2 Automatic Localization of Normal Active Organs in 3D PET Scan 6

2.1 Overview 6

2.2 Related Work 6

2.3 Method 8

2.4 Material 10

2.5 Experiments 10

2.6 Results 13

2.7 Discussion 16

2.8 Conclusion 18

3 Weakly Supervised Segmentation of Tumor Lesions in 3D PET Scan 19

3.1 Overview 19

3.2 Related Work 20
3.3 Method 20
3.4 Material 23
3.5 Experiments 24
3.6 Results 25
3.7 Discussion 26
3.8 Conclusion 26

4 Future Work **29**

Bibliography **31**

List of Tables

Table 2.1 Per-organ detection accuracy for the original YOLO model, 2D ConvNets [1], and our proposed extension to YOLO. P, R, and F1 stand for Precision, Recall and F1-score. 14
Table 2.2 Classification performance (F1-score) for different organs reported in previous works. LOO stands for Leave One Out cross validation approach. 14
Table 2.3 Average and standard deviation of 3D Euclidean distances between the centroid and walls of predicted and reference 3D bounding boxes for each organ. 14
Table 2.4 Results reported in previous works for organ localization accuracy on different datasets and various metrics. 15

List of Figures

Figure 1.1 An example of a 3D whole-body PET scanner [2]. 2

Figure 1.2 2D visualization of a 3D 18 F-FDG PET image. The arrows refer to thorax tumor mass and metastases to left supraclavicular lymph nodes. 3

Figure 2.1 Proposed method overview. 2D coronal PET slices (left) are input to the YOLO model. * indicate convolutional blocks with different kernel sizes. Organs detected in successive 2D slices with confidence score C>0.5 are aggregated into an organ-specific 3D bounding box with dimensions W×H×D and centroid coordinates (x,y,z). 9

Figure 2.2 Reference bounding boxes drawn overlaid on coronal (top) and sagittal (bottom) maximum intensity projection (MIP) views of 3 PET scans of 3 patients (3 columns). The boxes surround the brain (red), heart (green), left and right kidneys (blue and white) and bladder (purple). 11

Figure 2.3 Examples of detected (yellow) and reference 3D bounding boxes (other colors) for 3 PET scans from the coronal (top) and sagittal (bottom) plane in MIP views. The predicted class label is shown at the top right corner of the box. RK and LK correspond to the right and left kidneys. 12

Figure 2.4 Examples of detected 3D bounding boxes for different organs with the corresponding 3D IOU scores. A coronal (top row) and (bottom row) sagittal view is shown. 16

Figure 3.1 U-Net architecture. The inputs are 3D PET volume (Head and Neck area) and the single bounding box ground truth mask with the size of $160 \times 160 \times 128$ voxels. The output is the segmented lesions mask in the same size of the inputs. 23

Figure 3.2 An illustration of the proposed multi-loss function value during the training in different situations. The upward arrows and downward arrows demonstrate high and low values of the specified loss terms, respectively. 24

Figure 3.3 Mean Dice and Mean unsigned distance error (UDE) for different methods. The error bars represent the standard deviation of the metrics for each method. 27
Figure 3.4 2D (coronal and sagittal views) and 3D rendering of segmented lesions in two test cases. Green and purple colors, in order, show the lesions segmented by the proposed method and the average of ground truth lesions segmented by experts. In the first row, the proposed method successfully segments the lesions (Dice similarity of 83%). In the second row, our method fails to segment the lesions (Dice similarity of 31%) 27

List of Acronyms

PET	Positron Emission Tomography
CNN	Convolutional Neural Network
FCN	Fully Convolutional Network
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
FDG	Fluorodeoxyglucose
SUV	Standard Uptake Value
ROI	Region Of Interest
YOLO	You Only Look Once
ML	Machine Learning
MIP	Maximum Intensity Projection
LK	Left Kidney
RK	Right Kidney
IOU	Intersection Over Union
HU	Hounsfield Unit
CRF	Conditional Random Field
GMM	Guassian Mixture Model
QIN	Quantitative Imaging Network
LS	Level-Set
FS	Fully Supervised
SA	Semi-Automatic
UDE	Unsigned Distance Error

List of Notations

$Pr(x)$	Probability of x
C	Confidence score
W	Width of the bounding box
H	Height of the bounding box
D	Depth of the bounding box
L	Proposed loss function
L_{Dice}	Dice loss function
L_{MDice}	Modified Dice function
L_{MS}	Mumford-Shah inspired loss function
$\hat{p_i}$	Prediction voxel label
p_i	Reference voxel label
H_ϵ	Approximated Heaviside function
C_i	Average intensity of voxels
$\hat{P}$	Predicted segmentation volume
P	Ground-truth segmentation volume
$d(x, y)$	Distance between x and y
S_i	Surface area enclosing the volume i
$\bar{d}$	Average distance

Chapter 1

Introduction

1.1 Background and Motivation

PET is a nuclear medicine imaging modality that provides functional information for the purpose of diagnosis, staging and monitoring cancerous lesions in radiation therapy [3]. In contrast to magnetic resonance imaging (MRI) and computed tomography (CT), imaging modalities that capture the body's structural or morphological properties, in PET image, properties of biological function and processes within the body is captured. Hence, it improves cancer treatment and early disease detection.

PET is a minimally invasive imaging procedure that has a wide range of practical and research applications [4]. In therapy response assessment, after injecting radiotracers into the patient's body, different tumor features, such as the semi-quantitative parameter standard uptake value (SUV), are measured in the context of radiotracer analysis. Radiotracers are biologically relevant material, like glucose, which are labelled with rapidly decaying positron emitting radioisotopes (or radionuclides), such as carbon-11, nitrogen-13, oxygen-15 and fluoride-18. Fluorodeoxyglucose (FDG) is a glucose analogue that is a very common radiotracer used in clinical practice. FDG is taken up by cells that are requiring much glucose such as the brain, kidneys and rapidly growing tumor cells. For this reason, it is widely used for the study of cancer metabolism in the body [5].

A schematic of a whole-body 3D PET scanner is shown in Figure 1.1 [2]. PET imaging scanners usually are integrated with low dose x-ray CT images. The trade-off between image quality, acquisition time and injected radiation dose vary for different PET scanner. Figure 1.2 shows the 2D visualization of a 3D PET image in cancer screening applications. A thorax tumor mass and metastases to left supraclavicular lymph nodes are visible in head and neck area because of the 18 F-FDG radiotracer uptake in cells that shows the level of tissue heterogeneity within this area. [6].

Image intensity is defined by SUV, resampled discrete quantities of FDG concentration corrected by the injected dose and patient weight. Therefore, intensity resolution in PET images is dependent on the SUV range. Even with high-resolution scanners, PET data often

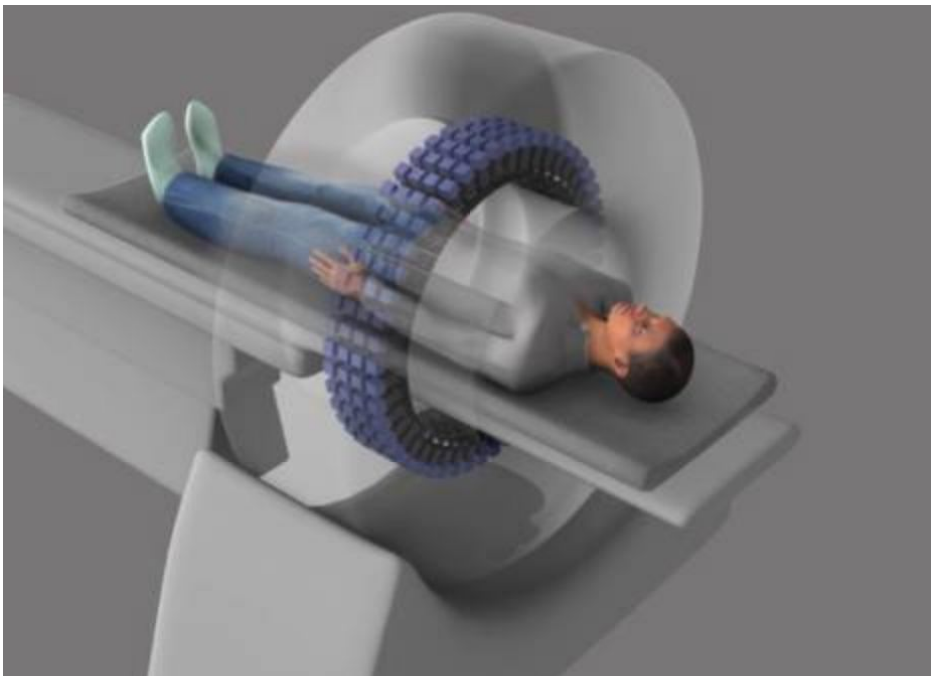

Figure 1.1: An example of a 3D whole-body PET scanner [2].

do not contain sufficient resolution to extract features and meaningful patterns for data analysis purposes. Quantitative analysis of SUV facilitate distinguishing between normal and abnormal levels of uptake such that the SUV value above a particular threshold is considered to be suggestive of malignancy, whereas lesions with SUV less than this threshold are considered to be benign.

In recent years, many studies proposed the use of FDG-PET image analysis techniques to detect and segment cancerous lesions due to quantification of tumor metabolic activity from the segmented tumor volume and radiation therapy treatment plans. This, in turn, requires more reliable, accurate and reproducible PET image tumor delineation [7]. As it is mentioned above, the important challenge in designing lesion segmentation techniques from PET images is handling the low signal to noise ratio and the low resolution of these images.

Different methods have been proposed for delineating lesions in PET images and can be categorized into five major groups. (i) The methodology of the first group is based on selecting either fixed or adaptive thresholds of the maximum standardized uptake values (SUV_{max}). However, this approach often performs poorly when applied to real clinical data as lesions generally have irregular shapes, values may vary depending on the reconstruction method used to acquire the PET image, and the images generally suffer from heterogeneous radio-tracer distributions [8, 9]. (ii) The second group consists of graph optimization methods which leverage user input to guide the segmentation algorithm or refines its result. Seed-guided region-growing methods belong to this group. Relying on manual input not only burdens users with time-consuming tedious interactive steps, but also gives rise

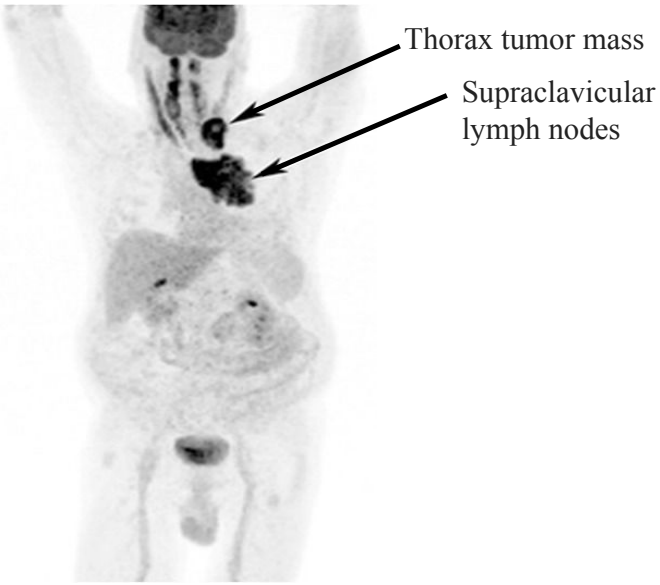

Figure 1.2: 2D visualization of a 3D 18 F-FDG PET image. The arrows refer to thorax tumor mass and metastases to left supraclavicular lymph nodes.

to inter- and intra-rater variability in the analysis results [10, 11]. (iii) The third category are energy-minimization segmentation methods, which are designed to locate and identify the boundaries of the target objects in PET images [12, 13]. Active contour-based models based on the Mumford-Shah functional [14] are examples of this category. These models typically require an initial contour or surface to start the optimization-based segmentation and the setting of different parameters to define the objective function. In addition to sensitivity to initialization and parameter setting, the execution time of these method on novel images tend to be slow to take into account the computational optimization steps. (iv) Another category encompasses statistical and machine learning-based methods, which estimate the segmentation of a novel test image based on the learned statistics of labeled training data [15, 16]. (v) The final category includes joint segmentation methods that in addition to PET, they take advantages of other modalities like CT and MRI. It has been shown that the fused images reduce uncertainty and are more suitable for visual perception [17, 18]. Among the mentioned approaches, machine learning based methods achieved high performance in terms of accuracy and time. Although the long training time is typically a drawback here, but the more critical inference time, e.g., segmenting a novel image, is generally much smaller than the run-time of energy-minimization methods. One of the key challenges, however, with supervised machine learning techniques (especially deep learning) is their reliance on the availability of a large set of training data.

1.2 Thesis Contributions

The purpose of this thesis is to investigate computerized PET image methods that address some of the aforementioned limitations of current quantitative analysis and automatic segmentation of cancerous lesion in the Head and Neck area. Such limitations cause some variation in overall tumor volume estimation and, subsequently, the radiotracer uptake measurement. Hence, we identified two major problems to tackle in order to boost the performance of PET quantitative analysis.

First, automatic lesion detection from PET scans is a challenging task due to the fact that both normal and abnormal regions in PET images have relatively high uptake of the radiotracer (commonly 18F-FDG) injection and result in voxels with similar appearance characteristics [19]. So, the presence of normal activity from healthy organs can alter the automatic localization of lesions resulting in less accurate segmentation or detection results or the need for additional forms of information (e.g., access to CT or manual post-processing). Therefore, for proper interpretation of PET images, it is important to detect the non-pathological active organs and exclude them from subsequent calculations when assessing cancer progression or treatment efficacy [20]. The first contribution of this thesis is on developing an efficient deep learning based method to automatically localize and classify normal active organs in 3D PET scans. By localization, similar to the state of the art methods introduced for organ localization in structural image modalities like CT, we particularly mean detecting organs in images and predicting a bounding box enclosing each organ. Our method encompasses training an object detection deep network with 2D PET image slices and aggregating result to estimate organ-labelled bounding boxes labels around active organs. We argue that the ability to automatically detect normal active regions in PET scans can facilitate lesion detection from PET images without the need for CT scans or post-processing steps.

Second, most of the recent successes of applying machine learning and deep learning for medical image analysis tasks (e.g., classification, detection or segmentation) can be attributed to supervised learning, for which a relatively large annotated set of trainig images must be made available. However, it is well-known that existing annotated medical imaging data sets are scarce and the effort involved in collecting them (especially ones with high quality annotated dense label-fields, i.e., segmentation masks) can be insurmountable [12]. It is also known that some annotations are easier and faster to collect than others. For instance, localizing lesions with bounding boxes or seeds (via one or a few mouse clicks) is less laborious than delineations. In this work, we contribute to developing a method to leverage such weak annotations in the context of PET lesion segmentation, where a deep learning model is trained to delineate lesions in 3D PET images from only bounding box annotations around the lesions of interest.

The second and third chapters of this thesis describe the details of the first and second contribution, respectively, which are based on the following published works:

- Saeedeh Afshari, Aïcha BenTaieb and Ghassan Hamarneh. Automatic localization of normal active organs in 3D PET scans, Computerized Medical Imaging and Graphics, Volume 70, Pages 111-118, 2018. doi:10.1016/j.compmedimag.2018.09.008.

- Saeedeh Afshari, Aïcha BenTaieb, Zahra Mirikharaji and Ghassan Hamarneh. Weakly Supervised Fully Convolutional Network for PET Lesion Segmentation, In SPIE Medical Imaging, 2019.

Chapter 2

Automatic Localization of Normal Active Organs in 3D PET Scan

2.1 Overview

Localization of anatomical structures in medical images is important for various clinical image interpretation applications. It can improve other medical image processing steps like segmentation. In case of PET image modality, to quantify metabolic activity and assessing tumor progression or treatment efficacy, detecting and parsing normal organ activities as a pre-processing step contributes efficiently to boost the accuracy of tumor detection/segmentation tools. Also, it helps to monitor organ activity concentration in dosimetry and therapeutic applications [21]. While CT contains valuable information useful for the detection and segmentation of organs, it often comes with additional costly steps such as the registration of CT and PET volumes. In this work, we adapt and extend an object detection deep CNN architecture (YOLO [22]) to locate, via 3D bounding boxes, and recognize all active organs in PET images by assigning a semantic (or anatomical) label to each organ (i.e., brain, heart, bladder, left and right kidneys). Our method is the first end to end trainable system that addresses the problem of organ localization in large 3D PET volumes and can handle a variable number of organs, i.e., it can gracefully ignore organs not visible within the field-of-view of a given PET image because we rely on the YOLO ability of detecting multiple organs visible in 2D slices of 3D PET volume. Next, we review some of the most related works to our method.

2.2 Related Work

Existing research on single or multi-organ localization approaches can be broadly categorized into two groups: (i) Atlas-based methods, which rely on a computationally intensive voxel-based image registration technique. (ii) machine learning-based (ML) methods, which rely on training an algorithm to learn the discriminatory image characteristics of different organs. Machine learning methods themselves may be divided into (ii-a) classical learning methods

that require extracting hand-crafted features and using them as input to train a subsequent classification or regression model; and (ii-b) deep learning methods that do not require hand-crafted features but typically require choosing or designing deep architectures and providing relatively larger training datasets.

The majority of previous works on organ localization in 3D volumes like Zhou et al. [23], Criminisi et al. [24], Gauriau et al. [25], and Wang et al. [26], focused on structural imaging modalities, such as CT scans and leverage ML models involving hand-crafted features and machine learning classifiers. For example, Zhou et al. [23] is based on Haar-like features and training ensembles of 2D location detectors. Similarly, Criminisi et al. [24] used visual features to train a multivariate regression forest. Gauriau et al. [25] encoded global spatial relationships between organs and their shape representations as prior knowledge to train random forest regressors. Wang et al. [26] proposed a fuzzy hierarchy model to encode anatomic information to predict the optimal pose of organs. These methods may not be applicable to functional nuclear medicine imaging, such as PET scans. For instance, in case of non-active organs, they should be capable to handle missing organs. Also, lack of sufficient differentiable visual and appearance features in PET images may result in weak performance of the proposed method for CT scans.

Recently, Convolutional Neural Networks (CNN) were also used to obtain 3D bounding boxes around organs in CT scans. Lu et al. [27] used CNN in a dual learning architecture to model local and global context for organ localization, trained on 2D slices from 3 orthogonal views (axial, coronal, sagittal) to localize organs in 3D CT volumes. de Vos et al. [1], proposed to feed each 2D orthogonal view image to a CNN which returns the binary label for the presence or absence of each target organ in the input image. Then, the binary predictions were aggregated to obtain 3D bounding boxes. de Vos et al. [1] was extended to use a single CNN able to handle all the orthogonal views using spatial pyramid pooling [28].

Few works targeted organ localization from PET images only (without requiring registered CT scans). Guan et al. [29] developed an automatic system to detect and segment hot spots (metabolically active regions in the body) in PET scans. However, they do not discriminate normal from abnormal activity values which limits the applicability of their work in the context of lesion detection. Bi et al. [30] attempted to specifically classify normal active organs from PET scans by applying an adaptive patch-based region classification to classify uptake regions in PET images obtained by automatic thresholding method (i.e., PERCIST). Their method [30], rely on hand-crafted texture features extracted from PET and registered CT patches with different scales. Then, Bi et al. [31] used CNN-based encoded features of multi-scale super-pixel regions to train binary classifiers for labeling normal uptake regions (organs) and abnormalities. Their result illustrates the benefits of using learned deep CNN features for classification. Fischer et al. [32] also tackled the problem of organ localization in PET images. They selected hundreds of visual features with an exhaustive search method

and used them to train regression forests. While these works tackle a similar problem, they rely on hand-crafted features and complex pipelines with extensive pre-processing steps (e.g., pre-processing using CT scans [20]).

In contrast to previous works on multi-organ localization in 3D volumes, our proposed method does not rely on 3 orthogonal views of 3D volume, and extensive pre-processing steps and feature extraction strategies. We report various comparisons of our approach to related state of the art methods and demonstrate improvement in evaluated measurements. The details of our proposed method are presented in the following section.

2.3 Method

Our approach localizes existing organs in each coronal slice of a PET scan and then aggregates the localized objects across all coronal slices (Figure 2.1). We chose the coronal slice orientation as multiple coronal slices expose various major organs and capture the spatial relationships between them (Figure 2.2). To perform a per-slice multiple organ localization, we adapted the YOLO deep CNN architecture (tiny version) [22], which has proven highly successful for object recognition tasks in non-medical computer vision. The YOLO network is a fully convolutional architecture (i.e., no fully connected layers) which consists of 3x3 convolution kernels (except for a 1x1 kernel at the last convolutional layer), batch normalization to avoid overfitting after each convolutional block, leaky ReLU as the activation function for the middle layers and max-pooling layers. As shown in Figure 2.1, the network has four types of convolutional blocks with different layers (i.e., convolution layer with the kernel size of 3x3 or 1x1, and max-pooling layer with the kernel size of 2x2 and stride of 2 or 1). We modified the last layer of the original architecture to handle 5 classes of organs. [22].

In YOLO, all features from the entire image are used to predict each object class with the bounding box around it. The input image is divided into a $S \times S$ grid. The cell that includes the predicted object's center is representing that detected object. Each cell also predicts B number of bounding boxes and confidence scores for those boxes which shows how much confident the model is about the box that enclose an object. The confidence score is defined as:

$$C = Pr(Object) \times IOU_{truth}^{pred} \tag{2.1}$$

$Pr(object)$ is the probability of an object existence in the box, and IOU is the intersection between the predicted box and the ground truth box.

For each box, 5 numbers are predicted by the network, x and y, the coordinates of the center of the box, and W and H, the relative width and height of the box to the input image, and finally confidence score C. The class probability of each detected object is predicted conditionally by each grid cell. It gives us class-specific confidence scores for each box representing the probability of that class appearing in the box and how well the

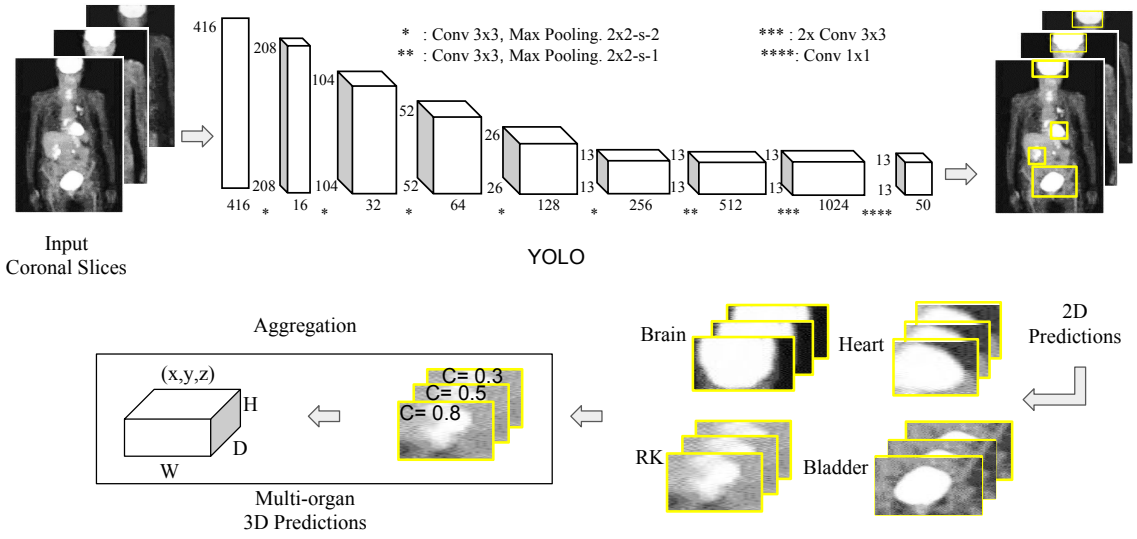

Figure 2.1: Proposed method overview. 2D coronal PET slices (left) are input to the YOLO model. * indicate convolutional blocks with different kernel sizes. Organs detected in successive 2D slices with confidence score C>0.5 are aggregated into an organ-specific 3D bounding box with dimensions W×H×D and centroid coordinates (x,y,z).

predicted box fits the object.

$$Pr(Class_i|Object) \times Pr(Object) \times IOU_{truth}^{pred} = Pr(Class_i) \times IOU_{truth}^{pred} \tag{2.2}$$

We modified the architecture to take as input resized coronal 2D PET slices and to detect up to 5 different organ classes: brain, heart, bladder, and left and right kidneys. We initialized the YOLO model using weights tuned for the PASCAL VOC dataset and then fine-tuned the model on coronal slices manually-labelled with bounding boxes. At test time, each coronal slice of a novel PET scan is fed into the trained YOLO model which returns a vector containing the coordinates of the 2D bounding boxes of multiple organs visible in the slice. Without any specific aggregation strategy, all the predicted 2D bounding boxes can be used to define a 3D bounding box that we refer to that as original YOLO.

For aggregation, we (i) collected all the bounding boxes that share the same semantic label (i.e., localizing the same organ) and their associated localization confidence scores; (ii) excluded any 2D bounding box whose confidence score is below 50%; (iii) computed a weighted, by confidence score, average 2D bounding box (X center, Y center, Width, and Height) for each organ over the predicted bounding boxes in successive slices; and, finally, to obtain 3D localizations, we (iv) set the Z center of the 3D bounding box to the mean z value of the successive slices containing the organ and set the 3D bounding box Depth to be equal to the z span of the same slices.

It has been shown that YOLO learns general representations of objects and its detection process is extremely fast . Hence, when generalizing from natural images to other domains, it outperforms other object detection methods, including R-CNN. It also is proven to predict less false positives on an image background [33]. Due to these reasons, we choose to adapt this network. However, there are other versions and newer networks for object detection task in natural images that are worth trying them in our data in the future work.

2.4 Material

We applied our method on 479 18F-FDG PET scans of 156 patients (QIN-HEADNECK) obtained from the public collection of head and neck cancer from the Quantitative Imaging Network of the US National Cancer Institutes [34] in compliance with all ethical clearance policies. The number of scans per patients varied from 1 to 7. The intra-patient scans are visually quite different in terms of organ activities, scan resolution and the field of view. The scan size varied between 128×128×74 and 168×168×545 voxels, with voxel sizes ranging between 3.39 and 4.68 mm along x and y, and 2 to 5 mm along z. The 479 scans were split as follows: 79 patients were used for training; and 77 different patients were used for testing. To feed to YOLO, all slices were resized to YOLO default input size 416×416.

To create ground truth bounding boxes, we used tight 3D bounding boxes drawn by an expert around the 5 normal organs of interest (or a subset thereof when some organs were out-of-view, i.e., not all organs are present in every PET scan). The organs of interest are the brain, heart, bladder, and left and right kidneys. The sizes of bounding boxes ranged from 32×45×61 mm^3 to 160×150×120 mm^3 (Figure 2.2). After training, our model detects normal organ activity in a novel test PET scan and gives the location of each organ as a 3D bounding box.

2.5 Experiments

We trained the network for 60 epochs and adopted the hyper-parameters used in the original YOLO: initial learning rate = 0.0001, decay = 0.0005 and momentum = 0.9. The size of the minibatch was set to 50 images. We implemented our method using Keras. We evaluated the performance of our approach for the object recognition (i.e., organ label) and 3D object localization (bounding boxes) tasks. Sample qualitative results are shown in Figure 2.3. We calculated the following performance metrics commonly used in other works for assessing the quality of anatomical structure detection and localization.

Direct and fair comparison to other works is difficult because, first, there are limited works that address the same task (i.e., detect and classify normal organs in PET). The closest works to ours are the works from Bi et al. [30, 31], which do not detect but only classify normal active organs after a thresholding step (Table 2.1). To the best of our knowledge,

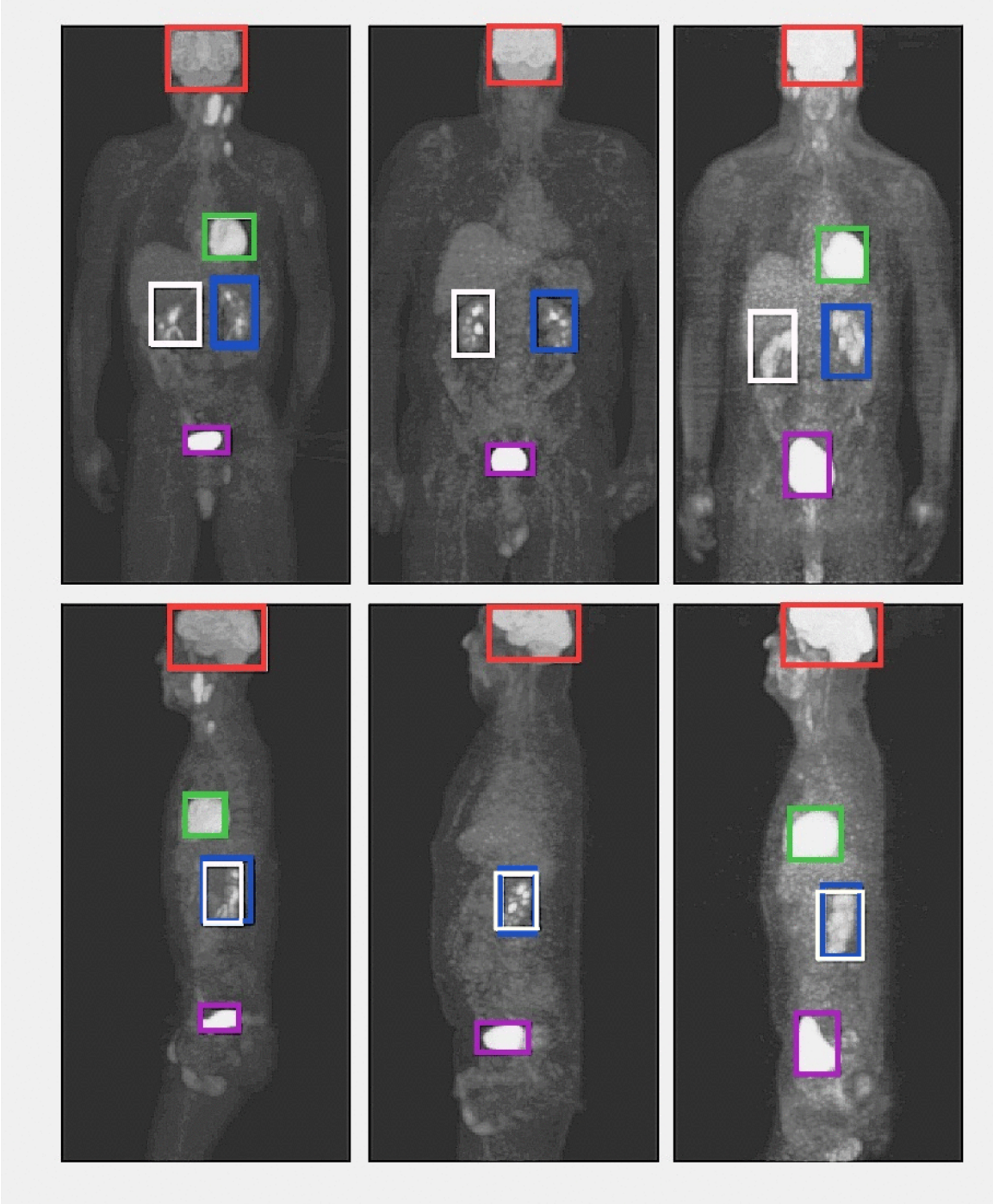

Figure 2.2: Reference bounding boxes drawn overlaid on coronal (top) and sagittal (bottom) maximum intensity projection (MIP) views of 3 PET scans of 3 patients (3 columns). The boxes surround the brain (red), heart (green), left and right kidneys (blue and white) and bladder (purple).

Fischer et al. [32] is the only work that tackles similar task to ours, i.e., organ localization in PET images by generating bounding box around detected organs but the technical details

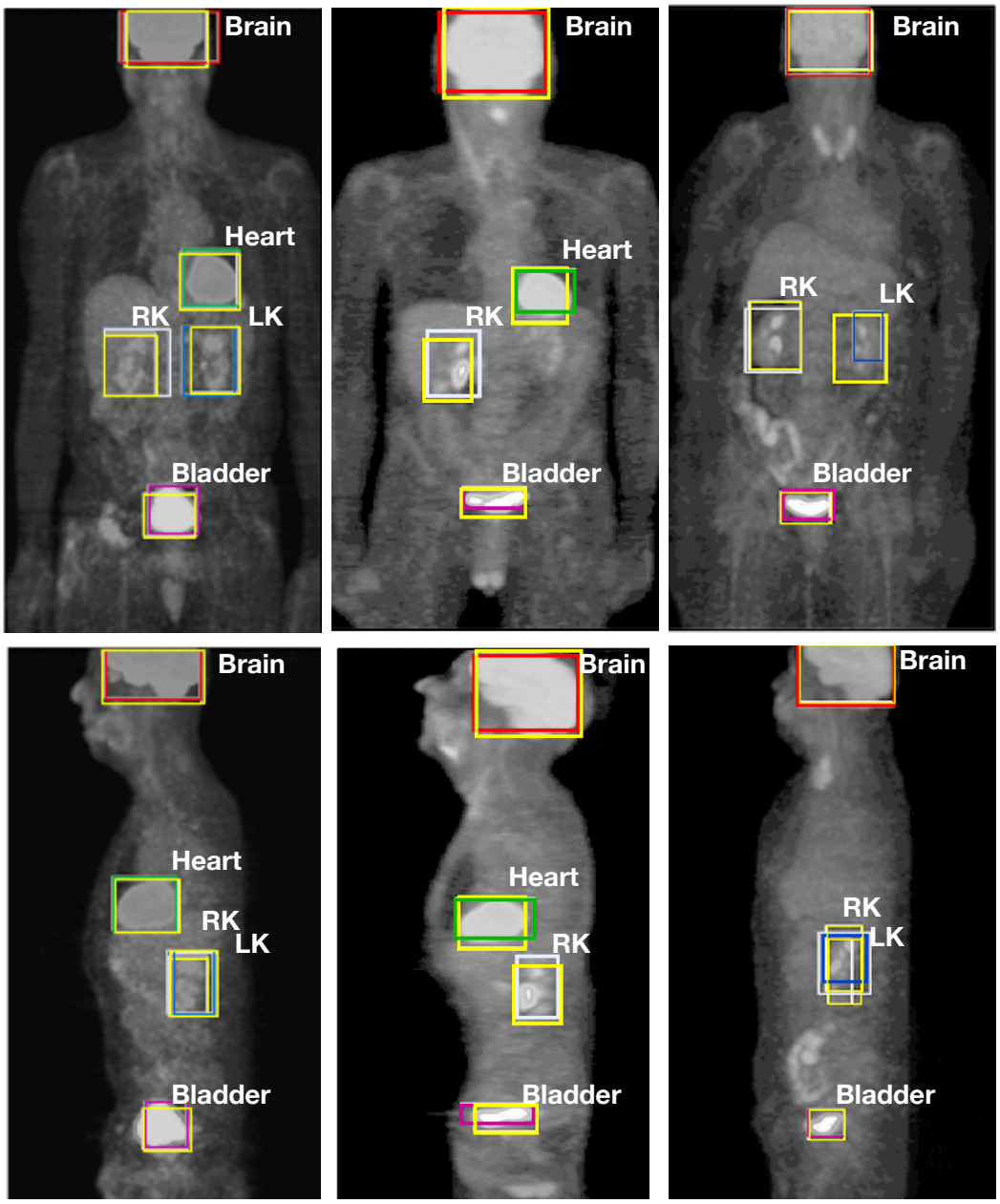

Figure 2.3: Examples of detected (yellow) and reference 3D bounding boxes (other colors) for 3 PET scans from the coronal (top) and sagittal (bottom) plane in MIP views. The predicted class label is shown at the top right corner of the box. RK and LK correspond to the right and left kidneys.

of the method are not clear in cited reference [32]. Second, other works like the ones listed in Table 2.4 that use a related learning-based approach for organ localization are mainly applied to CT images (not PET). Finally, another challenge is in finding the right evaluation metric to compare different works. Most works report localization accuracy in mm that can not be compared to directly since it is dependent on the dataset voxel size. Nonetheless, we perform the following comparison:
1. We identified two baseline works that our method can be compared to Bi et al. [30] and Bi et al. [31]. While we were not able to apply the method directly to our dataset (as the code was not released and the different steps involved in the algorithm were not easily generalizable to new datasets), we report the accuracy of these works on similar organs and indirectly compare our classification performance to them in Table 2.2.
2. We also identified the work of de Vos et al. [1], as the closest work to ours when it comes to organ localization. Although the method was applied to CT, we implemented it and evaluated it on our PET dataset. In a nutshell, the work of de Vos et al. [1] involves three convolutional networks trained on each orthogonal view (axial, sagittal, coronal) independently, to predict the presence of each organ in 2D slices of a 3D CT volume and then combines the prediction results from successive slices to create 3D bounding boxes around the organs (Table 2.1 and 2.3).
3. Along with these prior works, we also compared our method to the original YOLO to evaluate the benefits of our proposed aggregation strategy (Table 2.1 and 2.3).
4. To compare different works, we reported the original results from several previous works [23, 24, 25, 26, 27, 1, 32] on the same organs of interest (Table 2.4).

2.6 Results

The localization and classification performance of all models are reported in terms of wall distance, centroid distance error in millimeter (mm), precision, recall and F1-score in Table 1- 4. Note that for a fair comparison in Table 2.4, the centroid distance error is also reported in terms of minimum and maximum number of voxels. For several works [26, 1], we calculated this metric approximately based on the reported voxel size in the dataset used. Similarly, for the wall distance error, we reported the mean number of voxels over 6 walls of a 3D bounding box. We calculated the approximate mean number of voxels error in wall distance for some compared works [24, 25, 1, 32]. We discussed our results in the discussion section.

Organ Detection Performance. To evaluate the object-level detection of the proposed method and compared methods, we calculated precision, recall and F1 scores over our test set (Table 2.1 and 2.2). On average, we observed high precision and recall scores (75-100%) when detecting organs using our proposed method. In fact, while using YOLO allows us to

efficiently capture the context within each coronal slice, it results in noisy organ predictions (YOLO in Table 2.1).

Organ	YOLO			2D ConvNets			PROPOSED		
	P	R	F1	P	R	F1	P	R	F1
Brain	0.78	0.98	0.86	**0.97**	0.96	**0.96**	0.82	**1.00**	0.90
Heart	0.84	0.84	0.75	0.79	0.89	0.84	**0.86**	**0.96**	**0.90**
Bladder	0.86	0.93	0.86	0.89	0.94	0.91	**0.98**	**1.00**	**0.98**
RK	0.75	0.78	0.66	0.80	0.81	0.80	**0.83**	**0.95**	**0.88**
LK	0.69	0.74	0.57	0.73	0.84	0.78	**0.75**	**0.94**	**0.83**

Table 2.1: Per-organ detection accuracy for the original YOLO model, 2D ConvNets [1], and our proposed extension to YOLO. P, R, and F1 stand for Precision, Recall and F1-score.

Methods	train-test images	Brain	Heart	Bladder	RK	LK
Bi et al. (2015) [30]	40 (5-fold)	**0.93**	0.87	0.81	**0.97**	**0.93**
Bi et al. (2017)[31]	40 (LOO)	0.85	**0.92**	0.93	0.95	0.89
Proposed	254 - 225	0.90	0.90	**0.98**	0.88	0. 83

Table 2.2: Classification performance (F1-score) for different organs reported in previous works. LOO stands for Leave One Out cross validation approach.

Centroid and Wall Distance. For each correctly detected organ in all test scans, we calculated the distance between the center of the predicted 3D bounding box and the center of the reference bounding box. In addition, we calculated the average wall distance over the 6 walls of the bounding boxes for each organ. On average across all patients and organs, our method achieves 10.3 mm error in terms of centroid distance and 17.9 mm error in terms of wall distance (Table 2.3). Note that these results are obtained from low-resolution PET images (see Figure 2.2) for which localization is more challenging than regular CT images.

Organ	2D ConvNets		Proposed	
	Centroid (mm)	Wall (mm)	Centroid (mm)	Wall (mm)
Brain	7.23 ± 11.56	**11.64 ± 12.88**	**6.1 ± 5.0**	12.6 ± 7.7
Heart	13.60 ± 18.72	**18.92 ± 19.01**	**10.4 ± 9.1**	19.7 ± 12.4
Bladder	9.49 ± 15.63	**14.73 ± 47.97**	**8.5 ± 6.3**	15.1 ± 9.1
RK	18.36 ± 25.98	25.87 ± 24.61	**13.1 ± 10.6**	**19.2 ± 13.0**
LK	19.5 ± 28.52	27.38 ± 25.92	**13.4 ± 11.7**	**23.48 ± 14.3**

Table 2.3: Average and standard deviation of 3D Euclidean distances between the centroid and walls of predicted and reference 3D bounding boxes for each organ.

Intersection Over Union (IOU) in 3D space We show in Figure 2.4 examples of the predicted bounding boxes for different organs with their corresponding 3D IOU scores which

Method	Modality	Voxel size (mm× mm× mm)	Organ	Centroid (mm)	Centroid (min_{voxel}-max_{voxel})	Wall (mm)	Wall ($mean_{voxel}$)
Zhou et al. (2012) [23]	CT	0.6×0.6×2-5	Heart	-	1 - 19	-	-
			RK	-	1 - 19	-	-
			LK	-	1 - 34	-	-
Criminisi et al. (2013) [24]	CT	0.5-1×0.5-1×1-5	Heart	-	-	13.4	10.1
			RK	-	-	13.6	10.2
			LK	-	-	16.1	12.1
Gauriau et al. (2014) [25]	CT	0.5-1×0.5-1×0.5-3	RK	-	-	7.3	6
			LK	-	-	6.8	5.3
Wang et al. (2015) [26]	CT	1.2×1.2×4	Bladder	9.9	2.4 - 8.25	-	-
			RK	8.3	2.1 - 6.9	-	-
			LK	12.9	3.2 - 10.7	-	-
de Vos et al. (2016) [1]	CT	0.55-0.87×0.55-0.87×1	Heart	7	7 - 12.7	4	5.1
Lu et al. (2016) [27]	CT	2×2×2	RK	-	3.9 (mean)	-	-
Fischer et al. (2014) [32]	PET	5×5×5	Brain	-	-	15.4	3.1
			Liver	-	-	13.0	2.6
			Bladder	-	-	15.9	3.2
			RK	-	-	11.1	2.2
			LK	-	-	13.4	2.7
Proposed	PET	3.3-4.7×3.3-4.7×2-5	Brain	6.1	0.5 - 1.4	12.6	2.8
			Heart	10.4	1.4 - 2.4	19.7	4.5
			Bladder	8.5	0.7 - 2.1	15.1	3.5
			RK	13.1	1.2 - 3.5	19.2	4.4
			LK	13.4	1.5 - 3.6	23.48	5.4

Table 2.4: Results reported in previous works for organ localization accuracy on different datasets and various metrics.

is the 3D extension of 2D IOU. We obtained an average of 72% IOU for the brain, 55% for the bladder, 52% for the heart, 48% for the right kidney and 43% for the left kidney. On average, we observed that the detection of some organs is sensitive to the quality of the annotations when the context (anatomy and shape information) is not sufficient. This can explain the lower localization results obtained for kidneys.

Run-time Using non-optimized Python code, training our model took around 5 hrs on a Titan X GPU. At test time, locating and labeling all organs took only 19 seconds per 168×168×487 PET scan. Comparing to the 2D ConvNet method [1] that the test step running time for the same volume size and localizing 5 organs is ∼14 seconds, we reduced the training time and the space required for storing the network parameters by a factor of ∼15 (training 3 independent CNNs for localizing 5 organs of interest, separately for each one) . Note that our runtime is significantly faster than the previous work of Fisher et al. [32] on PET images who reported a 2-minute runtime per scan. We chose to rely on a per-frame organ detection using an efficient 2D detection model and to combine this model with a simple yet accurate 3D aggregation strategy. This allowed us to process large 3D volumes in a reasonable amount of time but also to use larger spatial resolutions when processing the 2D input images. Processing large 3D volumes with a 3D convolutional network architecture would require significantly larger models, datasets and longer training time.

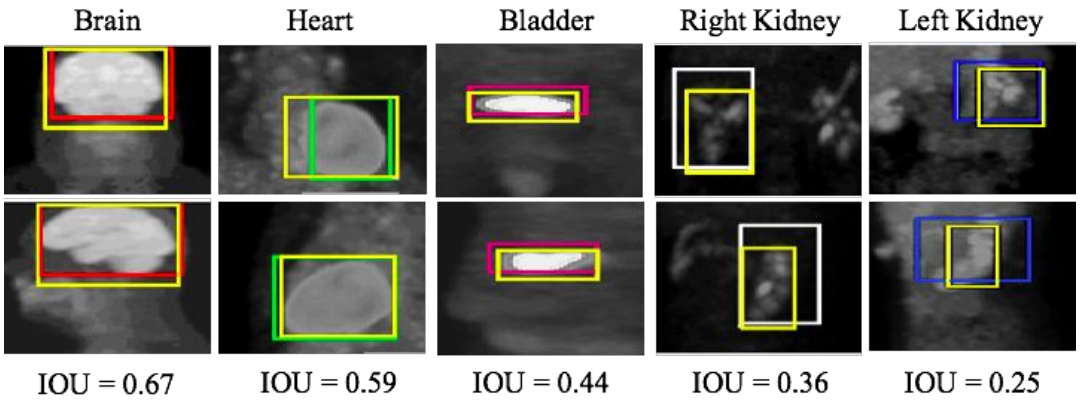

Figure 2.4: Examples of detected 3D bounding boxes for different organs with the corresponding 3D IOU scores. A coronal (top row) and (bottom row) sagittal view is shown.

2.7 Discussion

The results in Table 2.1 show that on average our method outperforms other works [1] and original YOLO. In fact, we show in Table 2.1 that the method of de Vos et al. [1], 2D ConvNets, shows better performance in classifying only the brain in 3D PET images (i.e., 15% higher precision than our proposed method), whereas our method improves the results on the other 4 organs, i.e. heart, bladder, right and left kidneys with 7%, 9%, 3%, 2% increase in precision accuracy for each organ respectively. Note also that our proposed method only uses coronal slices which requires less computational effort than feeding multiple slices as proposed in de Vos et al. [1].

The improvement observed in terms of classification performance of our method as shown in Table 2.1, could be due to filtering the noisy per-slice detections with the proposed aggregation strategy (described in section 2). We note that the brain and the bladder were consistently detected by our method likely due to their high FDG uptake values. Finally, we also indirectly compare our work to Bi et al. [30, 31] in Table 2.2 in terms of classification F1-score. Note that the comparison is indirect as it is on different datasets (the dataset used in Bi et al. [30, 31] is private and not public as ours so we were not able to apply our method to this data or use the authorsâĂŹ method on our dataset either as this previous work involves multiple steps that are not trivial to replicate without available code). Our method shows comparable performance to both methods with noticeable improvements in classification accuracy of the bladder in the relatively larger dataset.

We also compared our method to the original YOLO network without using our proposed aggregation strategy to eliminate low confidence predictions and observed lower classification accuracy for such setting. In fact, our aggregation approach shows that it can complement and improve the results of YOLO by eliminating some of the false positive

predictions and we were able to integrate the 3D context to refine the detection results (+26% F1 score).

In Table 2.3, the mean and standard deviation of absolute centroid and wall distances between the automatically obtained and reference bounding boxes are reported for the proposed method and 2D ConvNets method [1]. The proposed method obtained lower error (3.33 mm on average) in terms of the distance between the center of the predicted and the ground truth bounding boxes for all the target organs.

We observed that the detection and localization of left and right kidneys (75-83 % precision detection, and 13 mm centroid and 19.2 - 23.5 mm wall distance error) is the lowest result compared to all other organs of interest which we hypothesize is due to the noisy annotation in kidneys and heterogeneous FDG uptake in different regions of kidneys causing fragmented kidney masks. Such fragmentations increase the complexity of detection of these organs.

Although our method had relatively greater mean wall distance than 2D ConvNets, it is worth mentioning that this slightly higher error (less than 1mm) in predicting six walls of the 3D bounding boxes is due to the lack of context and difficulties in identifying anatomical boundaries from a single view compared to the results of having information from multiple views of an organ in 3D volumes. Empirically, we observed that using multiple views as opposed to only coronal views improved the wall distance results for 3 organs only (i.e., brain, heart, and bladder). Also, using multiple views did not improve the detection performance in terms of centroid distance, precision and recall on our dataset. From these observations, we concluded that using only coronal views allowed us to achieve competing results while reducing the total training computation time by a factor of 3 (slice orientations) $\times$ 5 (organs).

Clinical Application. Automatic localization of organs in PET images contributes to a diverse set of clinical applications. It provides context information for clinical tasks like identifying sites of abnormal metabolism, disease monitoring, and tumor tracking during therapy for an early assessment of treatment response and plans for future therapy as it is suggested in PET Response Criteria in Solid Tumors (PERCIST) guideline [35]. Among standard methods to differentiate involved regions in FDG PET/CT studies, thresholding uptake values is more common and vastly used in qualitative and quantitative assessments of metabolic activities in PET. One of the limitations of such standard methods is background normal FDG excretions which make image interpretation problematic. Therefore, detecting and discriminating normal uptakes facilitates lesion detection models and improves their results by describing statistics and spatial features of normal active regions from abnormalities. Our method potentially can be used as a preprocessing step for highlighting normal and active organs in 3D PET images that contributes to tumor response assessment approaches qualitatively and quantitatively. Furthermore, our method is applicable where registering multi-modality images (i.e., PET-CT, PET-MRI) is required. It can provide complemen-

tary information regarding organs and their anatomical location and size. Also, in retrieving images or part of images (cropping) in PET data collection, our method can be applied to obtain information associating to different target body parts automatically and save manual efforts significantly. For example, in head and neck cancer dataset, the focus is on head and neck area which can be retrieved from 3D scans of different scales automatically instead of identifying and cropping target areas by an expert.

2.8 Conclusion

In this work, we propose an automatic system that can directly detect and localize organs with normal activity. Our system consists of a deep neural network architecture specifically designed to detect and localize anatomical structures with normal activity directly from 3D PET images. We process 2D slices and generate 3D bounding boxes using an efficient aggregation strategy, which makes our method fast, reliable and efficient enough to be easily utilized in more complex abnormality detection/segmentation pipelines. Although we obtained promising results, we noticed in cases where very small lesions are close to normal organs, our method might assign normal activity label to those abnormal regions. We believe this limitation is due to the network not being trained to distinguish lesions from normal activities. We plan to extend our work by combining our organ detection method with previous PET lesion detection methods [21], i.e., to train the network to discriminate between normal activities and pathologies, and not just segment each independently, making our method more useful in clinical practice. Future works will also involve testing our method on quantifying the size and activity of lesions proximate to these detected organs. Furthermore, it would be interesting to explore applying our approach to localization of other anatomical structures in other imaging modalities.

Chapter 3

Weakly Supervised Segmentation of Tumor Lesions in 3D PET Scan

3.1 Overview

Most of the success of machine learning and deep learning based medical image analysis tasks (e.g., classification, detection or segmentation) so far can be attributed to supervised learning, for which a relatively large annotated set of images must be made available. However, it is well-known that existing annotated medical imaging datasets are scarce and the effort involved in collecting them (especially ones with high quality annotated dense label-fields, i.e., segmentation masks) can be insurmountable [12]. It is also known that some annotations are easier and faster to collect than others. For instance, localizing lesions with bounding boxes or seeds is less laborious than delineations. In this work, we develop a method to leverage such weak annotations in the context of PET lesion segmentation, where a deep learning model is trained to delineate lesions in 3D PET images from only bounding box annotations around the lesions of interest.

FDG-PET images are used for diagnosis, staging and monitoring cancerous lesions in radiation therapy [3]. Medical image processing and analysis of PET images is used to detect and localize cancerous lesions in order to quantify tumor metabolic activity from segmented tumor volume and to design radiation therapy treatment plans. This, in turn, requires more reliable, accurate and reproducible PET image tumor delineation [7]. An important challenge in designing automatic lesion segmentation techniques in PET images is handling the low signal to noise ratio and low resolution of these images.

Some segmentation techniques leverage anatomical information from CT scans registered to PET, especially for tumors located near the lungs, where the air creates high contrast in the tumor areas. Otherwise, using CT may cause over-estimations of tumor lesions [13]. Assuming CT scans are available (which cannot be guaranteed), PET-CT registration is prone to error and, even with hybrid PET/CT scanners, the two modalities are not perfectly aligned because of patient movement or respiratory motion [8]. Further, certain tumor

boundaries are difficult to see in CT due to similarities in Hounsfield units (HU) of tumor lesions and surrounding tissues, so delineating tumors based on metabolic activity presented in PET may be more reliable. In this work, we focus on head and neck lesion segmentation only from PET.

3.2 Related Work

To address the need of annotation for training machine learning models, different weakly supervised segmentation models that leverage bounding box annotations around objects have been proposed for medical and non-medical segmentation tasks. For natural images, deep convolutional neural network (CNN) models that exploit bounding box annotations give excellent results in semantic segmentation tasks [36, 37, 38]. For medical image segmentation, few works use CNN models to leverage bounding box annotation. One of the closest works to ours is Rajch et al.'s [39], where the authors built a model similar to the popular GrabCut algorithm by replacing the Guassian mixture model (GMM) with a CNN whose network parameters are optimized by minimizing a conditional random field (CRF) energy function. They applied their method to segment the brain and the lungs from MR images using bounding box annotations. While our proposed approach aims at leveraging a similar type of weakly supervised training data for medical image segmentation, our work has important differences: We are segmenting lesions in PET (not organs from MR); our method is truly 3D whereas other works process 3D volumes as 2D slices, which results in loss of context when segmenting 3D lesions; and our approach does not require user interaction at inference time whereas other methods do.

Another closely related work to ours is Deep Level Set by Hu et al. [40]. At a high level, both our method and their method combine a convolutional network with energy-minimization based segmentation. In contrast, while our model is based on the 3D network architecture, we formulate a different level-set energy as a regularization term in a new loss function to segment target lesions in PET images in an unsupervised manner. This regularization term allows us to train the model in a weakly supervised context (i.e., using only bounding box annotations for training).

3.3 Method

To leverage weakly supervised data yet achieve 3D lesion delineation at test time without any user-interaction, we propose a novel loss function that dynamically combines two loss terms; one that is guided by the training bounding boxes, and the other designed to realize Mumford-Shah-inspired piecewise constant segmentation within the box. We show competitive results with other works on a challenging public dataset of PET lesions and compare

our model to different baseline methods, experts' delineations, and other semi-automatic approaches.

Given a set of 3D PET training images and their corresponding bounding box annotations, our goal is to optimize the parameters of an FCN in an end-to-end manner to identify and segment all lesions of interest. Our model is trained with a new loss function:

$$L = \beta L_{MDice} + \lambda L_{MS} \tag{3.1}$$

where L_{MDice} is the modified version of the Dice loss function, L_{MS} is the Mumford-Shah inspired term and β and λ are hyper-parameters balancing the contribution of each term. The interaction between both terms of the loss function is controlled via these hyper parameters and allows the network to first predict segmentations that are relatively close to the real segmentation masks, and then refine them using the regularization term.

In segmentation tasks where the majority of voxels are background voxels, the Dice similarity coefficient is often deployed as a loss function, L_{Dice}, to address this class imbalance [41]:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} p_i \cdot \hat{p}_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} \hat{p}_i} \tag{3.2}$$

where p_i and $\hat{p}_i$, in order, are the ground truth label and prediction label (class probability) for a voxel i, and N is the total number of voxels in the volume.

To encode the Mumford-Shah piecewise constant image model, commonly used with energy minimizing level-set segmentation, and train a deep network under weak supervision, we modify L_{Dice} as follows.

Modified Dice loss term. Optimizing L_{Dice} when annotations are bounding boxes forces the network to generate box-like masks, which is undesirable for a lesion segmentation task. To generate segmentation masks instead of bounding boxes, we modify the denominator of L_{Dice} by proposing a new term L_{MDice} that penalizes the misclassified voxels inside and outside of the box annotations in different ways. Outside the box, the classical Dice loss is always applied (to discourage any false positive voxels). Inside the box, however, the loss is dynamically modified such that the Mumford-Shah piecewise constant image model takes effect, instead of the classical Dice, whenever sufficient voxels are labeled as foreground. This dynamic loss behavior trains the network to use the weakly supervised labeling to effectively localize lesions and the Mumford-Shah model to delineate them. In particular, we set:

$$L_{MDice} = 1 - \frac{2\sum_i p_i \cdot \hat{p}_i}{\sum_i p_i + \sum_i \hat{p}_i - M \cdot H(\frac{\sum_i p_i \cdot \hat{p}_i}{\sum_i p_i})} \tag{3.3}$$

$$M = \sum_{i=1} p_i - \sum_{i=1} p_i \cdot \hat{p}_i \quad \text{and} \quad H(x) = \begin{cases} 1, & x > \alpha \\ 0, & x \leq \alpha \end{cases} \tag{3.4}$$

where $H(x)$ is a Heaviside function shifted by hyper-parameter α. Note that L_{MDice} is equal to L_{Dice} as long as the fraction of voxels, inside the ground truth bounding box and classified as foreground, is less than or equal to α. Otherwise, the denominator of L_{MDice} becomes $\sum_i \hat{p}_i + \sum_i p_i \cdot \hat{p}_i$, which means that voxels inside the box yet predicted as background are not penalized; this is when L_{MS}, described next, is most critical.

Mumford-Shah inspired loss term. In order to train our network to produce segmentation masks, inside the bounding box, that mimic results obtained using energy-minimizing level-set methods that encode the Mumford-Shah piecewise constant image model (e.g., [42]), we define L_{MS} as follows:

$$\begin{aligned} L_{MS} = \frac{1}{\sum_i p_i} \Bigg(& \omega_1 \sum_i \left(|I_i \cdot H_\epsilon(\phi_i) - C_1| - |I_i \cdot H_\epsilon(\phi_i) - C_2| \right) \cdot p_i + \\ & \omega_2 \sum_i \left(|I_i \cdot (1 - H_\epsilon(\phi_i)) - C_2| - |I_i \cdot (1 - H_\epsilon(\phi_i)) - C_1| \right) \cdot p_i \Bigg) \end{aligned} \tag{3.5}$$

where I_i is the input image intensity of the i^{th} voxel. $H_\epsilon(x)$ is the regularized Heaviside function proposed in [42]:

$$H_\epsilon(x) = \frac{1}{2} \left(1 + \frac{2}{\pi} arctan(\frac{x}{\epsilon}) \right). \tag{3.6}$$

We set ϵ to 0.03 as suggested in [40]. ϕ is the level-set function, which we obtain by linearly shifting the prediction map values to be in the range [-0.5, 0.5] so as to encode the lesion boundary as its zero level-set. Multiplying by p_i (defined in (3.2)) confines L_{MS} calculations to the interior of the bounding box. Normalization by $\sum_i p_i$ is used to obtain similar contribution to the loss from bounding boxes of different sizes. ω_i are scalar weights, and C_1 and C_2, in order, are the average intensities of voxels inside and outside of the prediction boundary, which are calculated by:

$$C_1 = \frac{\sum_i I_i \cdot H(\phi_i) \cdot p_i}{\sum_i H(\phi_i) \cdot p_i} \quad \text{and} \quad C_2 = \frac{\sum_i I_i \cdot (1 - H(\phi_i)) \cdot p_i}{\sum_i (1 - H(\phi_i)) \cdot p_i} \tag{3.7}$$

Neural Network Architecture. We adopt the U-Net architecture with contracting path to efficiently exploit contextual information. U-Net demonstrated good performance when the number of training images is limited [43]. We changed the 2D architecture of U-Net to 3D FCN and reduced the total number of feature maps by a factor of 4 (Figure 3.1). The size of the input image, the output segmentation mask, and the ground truth mask is

set to $160 \times 160 \times 128$. Scaled exponential linear units (Selu) are selected as the activation functions in the middle layers and sigmoid function for the output layer. The model is trained end-to-end with the proposed loss function in 3.1.

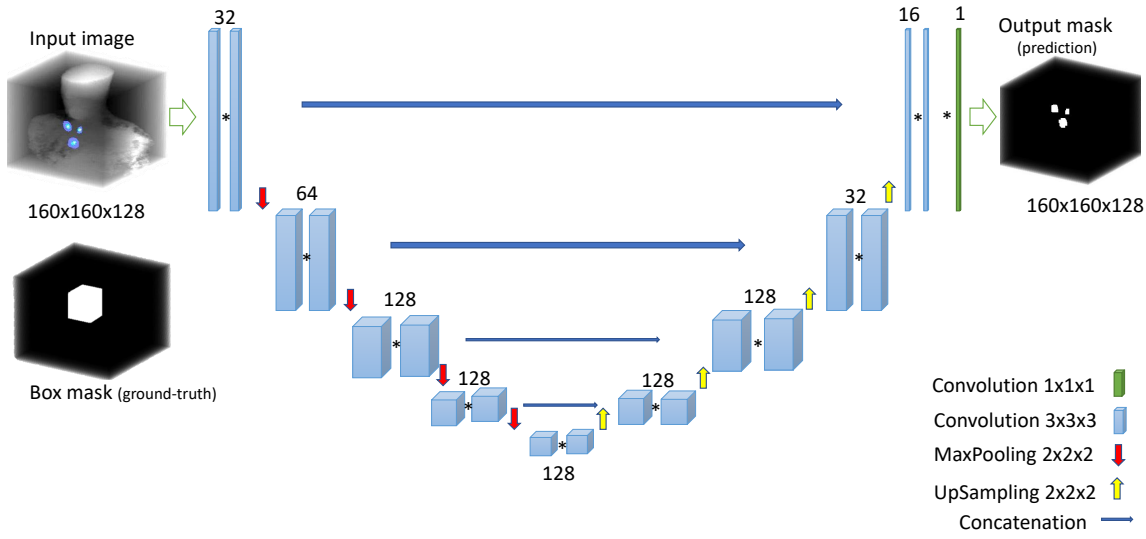

Figure 3.1: U-Net architecture. The inputs are 3D PET volume (Head and Neck area) and the single bounding box ground truth mask with the size of $160 \times 160 \times 128$ voxels. The output is the segmented lesions mask in the same size of the inputs.

Training procedure illustration. The proposed multi-loss handles error computation different within and outside the bounding box annotation. In Figure 3.2, the interplay between the proposed loss function terms is illustrated for different situations. When the predicted segmentation mask (the green object) and the annotation bounding box drown in blue do not overlap (a), the modified Dice loss calculates the segmentation error similar to the original Dice loss function. As the segmentation mask moves closer to the bounding box and overlap happens (b-e), the modified Dice term value decreases and becomes zero when segmentation map locates in the box (d). At the same time, the Mumford-Shah loss term starts calculating the segmentation error of miss-classified voxels inside the bounding box till the network learns to segment the foreground lesions inside the bounding box annotation (e).

3.4 Material

We applied our method to 57 FDG-PET scans of unique patients from the public collection of head and neck cancer provided by the Quantitative Imaging Network of the US National Cancer Institute [44]. From this dataset, a subset of 10 PET scans was utilized for the QIN PET segmentation challenge [45]. On average, the dataset includes 3.8 ± 2.5 lesions per PET volume (ranging from 1 to 12) with the volume ranges from 1 to 13+ ml. The selected test

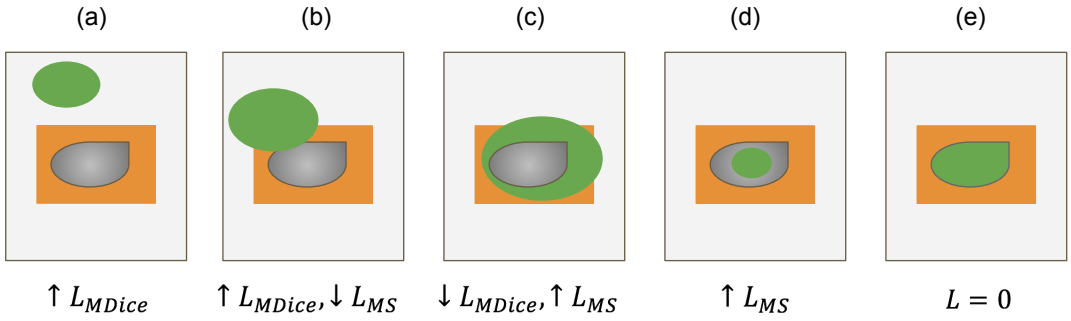

Figure 3.2: An illustration of the proposed multi-loss function value during the training in different situations. The upward arrows and downward arrows demonstrate high and low values of the specified loss terms, respectively.

cases included different complexity levels (ranging from low to high) [46]. The PET images vary in resolution from 128×128 to 168×168 and the number of axial slices ranged from 191 to 545. The average voxel size in this dataset is $3.5 \times 3.5 \times 3.1$ mm.

We used the challenge test set and the provided delineation by experts (average over experts' delineations to reduce inter- and intra- variability) for evaluation of our model and all other baselines reported in the results. We train the network on the remaining 47 cases. The box annotations are coarse bounding boxes around all lesions for each image with an average Dice similarity, with the ground truth delineations, of $28\% \pm 11\%$.

Given the low resolution of PET images and the complexity of the task, we found the following pre-processing steps necessary to accurately train our model and all baselines. First, as this dataset only includes lesions in the head and neck area, we reduced the computational cost by cropping all volumes inferior to the neck area such that we could process the entire volumes (not patch-based, and not slices) with 3D networks. We normalized all images to zero mean and unit variance and performed different data augmentation strategies (i.e., random flipping left-right and superior-inferior orientation, and additional Gaussian noise). We assume that flipping the 3D images increase the diversity of tumor shapes in the training dataset and allows the model to be spatially invariant. After augmentation, the total number of images in the training and validation set were 430 and 40, respectively.

3.5 Experiments

We compared our method with a fixed architecture CNN trained with fully supervised ground truth delineations by experts (L_{FS}) and a CNN trained with bounding box annotation (L_{Dice}). The performance of these two networks can be considered as the upper bound and the worst-case accuracy, respectively, that our method can achieve. In addition, we in-

clude the result of training our model with only L_{MDice}. We also tested using morphological (M) operation (dilation and erosion) as a post-processing ($L_{MDice}+L_{MS}+M$) to show that our result can be improved by separating boundaries of touching lesions and removing small sized false positive predictions. Furthermore, we compare our results to those obtained when active contours without edges level-sets method [42] is applied, as a post-processing step, to the prediction output of L_{Dice} ($L_{Dice}+LS$). We also added the result of the semi-automatic approach (SA) [46], which is a graph based optimization approach guided by a user provided approximate lesion center point and enables finding an optimal segmentation surface of tumor lesions from PET images. To compare with experts performance, we include the experts agreement by averaging Dice coefficient over all experts within each trial ($Experts$).

For training the network, all hyper-parameters are tuned on the validation set. For the proposed loss function, ω_1 and ω_2 are selected empirically based on the strength of two terms in equation 3.5 for the validation dataset, and are set to 1 and 3, respectively. β and λ are set to 10 and 1, respectively, in order to emphasize the Dice loss. For the modified Dice loss term, we used $\alpha = 0.1$, which means at least 10% of the voxels inside the box are forced to be predicted as tumor. We used stochastic gradient descent as the model optimizer and set the momentum, weight decay, and learning rate to 0.9, $1e-6$ and $1e-5$, respectively. The batch size is 1 and the maximum number of epochs is 100, which is limited by an early stopping criteria when there is no improvement on validation loss after 20 epochs. We used the trained network on box mask annotations to initialize the weights.

3.6 Results

To evaluate the performance of the proposed framework for segmentation, we adopted the Dice similarity coefficient which is a volume-based measurement and unsigned distance error (UDE) metrics, a surface-based measurement used in [45]. Let P and $\hat{P}$ be binary volume describing the reference segmentation and the predicted segmentation, respectively.
The volume error of segmentation is determined by the Dice coefficient with the below equation.

$$Dice_{coef}(P,\hat{P}) = 2\left|P\bigcap\hat{P}\right| / \left[|P| + \left|\hat{P}\right|\right] \tag{3.8}$$

The distance between a point x and a surface A is defined as:

$$d(x,A) = min_{x\prime\in A}\,\|x - x\prime\| \tag{3.9}$$

The mean unsigned distance error between two surface S_P and $S_{\hat{P}}$ is evaluated as following:

$$UDE_{mean}(S_P, S_{\hat{P}}) = \left[\bar{d}(S_P, S_{\hat{P}}) + \bar{d}(S_{\hat{P}}, S_P)\right] / \left[|S_P| + |S_{\hat{P}}|\right] \tag{3.10}$$

where

$$\bar{d}(S_1, S_2) = \int_{x \in S_1} d(x, S_2) dx \tag{3.11}$$

To compare different methods more efficiently, results on the test data set are summarized in Figure 3.3 and two examples of successful and failed segmentation results for two different 3D PET images are shown in Figure 3.4.

3.7 Discussion

Figure 3.3 shows that in the case of using our proposed multi loss function, i.e., L_{MDice}+L_{MS} to train the model, the improvement of the Dice coefficient is approximately 30% greater than the performance of the same model trained using the original Dice L_{Dice}. This improvement is obtained because the network is not penalized to predict a box like segmentation guided by the bounding box annotation when our modified Dice loss term is used and moreover, our L_{MS} loss term encourages the network to learn how to segment the foreground voxels with high intensity values from the background voxels with low intensity values inside the bounding box annotation. We also found that by applying morphological (M) operations to our weakly supervised outputs, our performance of $(L_{MDice} + L_{MS} + M)$ differs from the results of the fully supervised (L_{FS}) approach trained by the expert's delineations by only 7% in Dice similarity. Furthermore, while our method does not need either user interaction or ground truth segmentation masks, its performance is not very far from the performance of the SA approach, in which a user selects a point in the center of each lesion of interest (Dice similarity 68% and 82%, respectively). While it seems desirable for physicians to apply more accurate semi-automated segmentation tool, for difficult circumstances (i.e., a large clinical trial analysis), it may be costly and time consuming.

The mean UDE for the proposed method is 3.1 mm, which is comparable to the L_{FS} model with an error of 2 mm.

In Figure 3.4, for the first row, the proposed method is successful in delineating and splitting the tumor lesions while in the second image, it fails to find and delineate all lesions. The reason for missing to detect the small lesions would be the variability seen in manually generated reference segmentations, and the inhomogeneity of tracer uptake in these lesions.

3.8 Conclusion

Given the clinical importance of segmenting lesions from PET images (e.g., for radiation therapy treatment planning or assessing treatment efficacy) and given the tedious task of manually delineating datasets for training machine learning algorithms, we proposed a deep learning lesions segmentation method that requires only weakly labeled data (in the form of bounding boxes). We developed a novel multi-loss function with an adaptive Dice coefficient

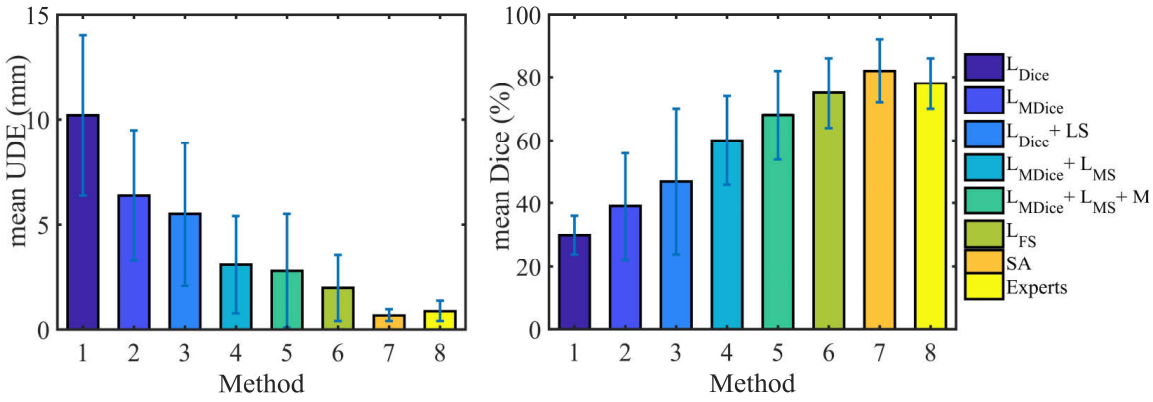

Figure 3.3: Mean Dice and Mean unsigned distance error (UDE) for different methods. The error bars represent the standard deviation of the metrics for each method.

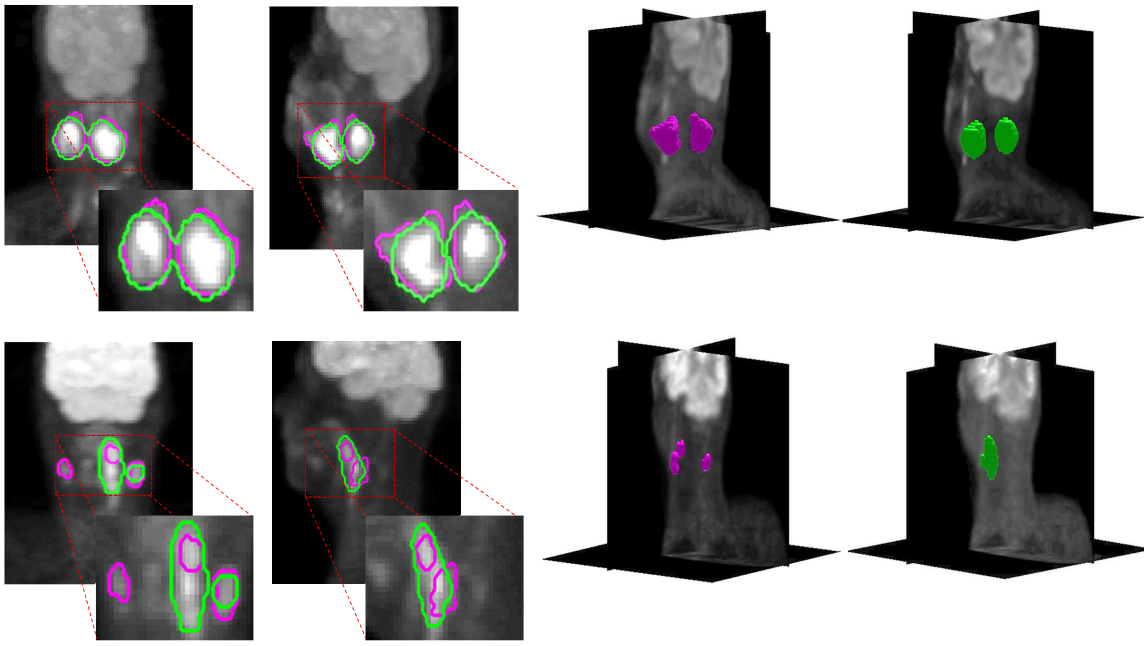

Figure 3.4: 2D (coronal and sagittal views) and 3D rendering of segmented lesions in two test cases. Green and purple colors, in order, show the lesions segmented by the proposed method and the average of ground truth lesions segmented by experts. In the first row, the proposed method successfully segments the lesions (Dice similarity of 83%). In the second row, our method fails to segment the lesions (Dice similarity of 31%)

term and a Mumford-Shah piecewise constant model term. Our results validated that the provision of training data in the form of bounding boxes is a viable option to careful expert delineation. Future work will focus on testing the method on other types of cancer lesions.

Chapter 4

Future Work

The first contribution of this thesis is on automating the detection and localization of physiological FDG uptake and normal FDG excretion, as in the kidneys, bladder, brain and heart, from 3D whole-body PET scans. The presence of these normal activities can obscure sites of abnormal uptake and render image interpretation problematic. The proposed method demonstrated promising results and improvement compared to the state of the art methods. The second contribution, automatic lesion segmentation from weak annotations, incorporated a joint, simultaneous detection and segmentation method of lesions in PET scans.

Future work could focus on some these following directions:

- Extending the first contribution by examining the more recent advancements in the rapidly developing field of deep learning based object detection, such as methods that demonstrated better performance than YOLO for non-medical image analysis tasks (e.g., object recognition in natural images).

- Extending the first conteribution by departing from aggregating slice-by-slice analysis results and instead adopt an inherently 3D analysis of the volumetric PET data. For example, comparing the accuracy and feasibility of the 3D extension of YOLO or any other object detection networks would be interesting in the task of normal activity detection and localization in 3D PET images.

- Investigating dedicated models that learn and perform lesion detection and segmentation tasks separately, in sequence, as opposed to the hybrid approach of the second contribution.

- Investigating the trade-off between fine-tuning (i.e., transfer learning) the prediction models that were trained on large numbers of natural images vs retraining the models from scratch on smaller sets of medical images.

- Examining various image augmentation techniques to artificially increase the number of training images [47, 48] or using physics based simulations to create reaslitic artificial data [49, 50].
- Considering adopting techniques that request from the user to label only images or slices that automated methods are uncertain about, e.g., active learning [51, 52].
- Extending the loss functions with other target object priors [53, 54, 55]
- Exploring training models even without bounding box or other annotations, i.e., leveraging unsupervised learning approaches similar to auto-encoders and clustering algorithms.
- Increasing the number of test images for a stronger evaluation of the generalizability of the models, and exploring possible extensions, especially for dealing with novel datasets acquired with, e.g., different scanners, reconstruction methods, and radiotracers.
- Utilizing the dynamic aspect of PET images, in which a time activity curve is collected at each pixel [56]
- Exploring extending the methods to analyze multi-modal (multi-channel) imaging datasets, such as fused PET-CT images, to leverage the complementary information provided by different modalities.

Bibliography

[1] B. de Vos, J. Wolterink, P. de Jong, M. Viergever, and I. Išgum, "2D image classification for 3D anatomy localization: employing deep convolutional neural networks," in *Medical Imaging 2016: Image Processing*, vol. 9784, p. 97841Y, 2016.

[2] S. R. Cherry, T. Jones, J. S. Karp, J. Qi, W. W. Moses, and R. D. Badawi, "Total-body pet: maximizing sensitivity to create new opportunities for clinical research and patient care," *Journal of Nuclear Medicine*, vol. 59, no. 1, pp. 3–12, 2018.

[3] H. Zaidi, H. Vees, and M. Wissmeyer, "Molecular PET/CT imaging-guided radiation therapy treatment planning," *Academic Radiology*, vol. 16, no. 9, pp. 1108–1133, 2009.

[4] C. C. Meltzer, J. D. Luketich, D. Friedman, M. Charron, D. Strollo, M. Meehan, G. K. Urso, M. A. Dachille, and D. W. Townsend, "Whole-body fdg positron emission tomographic imaging for staging esophageal cancer: comparison with computed tomography," *Clinical nuclear medicine*, vol. 25, no. 11, pp. 882–887, 2000.

[5] E. Miele, G. P. Spinelli, F. Tomao, A. Zullo, F. De Marinis, G. Pasciuti, L. Rossi, F. Zoratto, and S. Tomao, "Positron emission tomography (PET) radiotracers in oncology–utility of 18F-Fluoro-deoxy-glucose (FDG)-PET in the management of patients with non-small-cell lung cancer (NSCLC)," *Journal of Experimental and Clinical Cancer Research*, vol. 27, no. 1, p. 52, 2008.

[6] J. A. Taverna, H. M. Babiker, S. Yun, M. C. Bishop, S. Lau-Braunhut, P. N. Meyer, and T. Enzler, "The great masquerader of malignancy: chronic intestinal pseudo-obstruction," *Biomarker research*, vol. 2, no. 1, p. 23, 2014.

[7] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present, any future?," *European journal of nuclear medicine and molecular imaging*, vol. 44, no. 1, pp. 151–165, 2017.

[8] M. Hatt, J. A. Lee, C. R. Schmidtlein, I. E. Naqa, C. Caldwell, E. De Bernardi, W. Lu, S. Das, X. Geets, V. Gregoire, *et al.*, "Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211," *Medical physics*, vol. 44, no. 6, 2017.

[9] H. Vees, S. Senthamizhchelvan, R. Miralbell, D. C. Weber, O. Ratib, and H. Zaidi, "Assessment of various strategies for 18 f-fet pet-guided delineation of target volumes in high-grade glioma patients," *European journal of nuclear medicine and molecular imaging*, vol. 36, no. 2, pp. 182–193, 2009.

[10] E. Day, J. Betler, D. Parda, B. Reitz, A. Kirichenko, S. Mohammadi, and M. Miften, "A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients," *Medical physics*, vol. 36, no. 10, pp. 4349–4358, 2009.

[11] H. Li, W. L. Thorstad, K. J. Biehl, R. Laforest, Y. Su, K. I. Shoghi, E. D. Donnelly, D. A. Low, and W. Lu, "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours," *Medical physics*, vol. 35, no. 8, pp. 3711–3721, 2008.

[12] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.

[13] M. Wanet, J. A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Grégoire, and X. Geets, "Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens," *Radiotherapy and Oncology*, vol. 98, no. 1, pp. 117–125, 2011.

[14] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on pure and applied mathematics*, vol. 42, no. 5, pp. 577–685, 1989.

[15] M. Hatt, C. C. Le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive bayesian segmentation approach for volume determination in pet," *IEEE transactions on medical imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[16] M. S. Sharif, M. Abbod, A. Amira, and H. Zaidi, "Artificial neural network-based system for pet volume segmentation," *Journal of Biomedical Imaging*, vol. 2010, p. 4, 2010.

[17] U. Bagci, J. K. Udupa, J. Yao, and D. J. Mollura, "Co-segmentation of functional and anatomical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 459–467, Springer, 2012.

[18] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered fdg pet/ct-based textural characterization of head and neck cancer for radiation treatment planning," *IEEE transactions on medical imaging*, vol. 28, no. 3, pp. 374–383, 2009.

[19] S. A. Sarji, "Physiological uptake in FDG PET simulating disease," *Biomedical imaging and intervention journal*, vol. 2, no. 4, 2006.

[20] L. Bi, J. Kim, D. Feng, and M. Fulham, "Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 569–576, 2014.

[21] Y. Song, W. Cai, H. Huang, X. Wang, Y. Zhou, M. J. Fulham, and D. D. Feng, "Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies," *IEEE transactions on medical imaging*, vol. 33, no. 2, pp. 408–421, 2014.

[22] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *arXiv preprint arXiv:1612.08242*, 2016.

[23] X. Zhou, S. Wang, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, and H. Fujita, "Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning," *Computerized Medical Imaging and Graphics*, vol. 36, no. 4, pp. 304–313, 2012.

[24] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical Image Analysis*, vol. 17, no. 8, pp. 1293–1303, 2013.

[25] R. Gauriau, R. Cuingnet, D. Lesage, and I. Bloch, "Multi-organ localization combining global-to-local regression and confidence maps," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 337–344, Springer, 2014.

[26] H. Wang, J. K. Udupa, D. Odhner, Y. Tong, L. Zhao, and D. A. Torigian, "Body-wide anatomy recognition in PET/CT images," in *Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9415, p. 941518, 2015.

[27] X. Lu, D. Xu, and D. Liu, "Robust 3D organ localization with dual learning architectures and fusion," in *International Conference on Medical Image Computing and Computer-Assisted Intervention-LABELS*, pp. 12–20, 2016.

[28] B. D. de Vos, J. Wolterink, P. de Jong, T. Leiner, M. Viergever, and I. Isgum, "Convnet-based localization of anatomical structures in 3D medical images," *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1470–1481, 2017.

[29] H. Guan, T. Kubota, X. Huang, X. S. Zhou, and M. Turk, "Automatic hot spot detection and segmentation in whole body FDG-PET images," in *IEEE ICIP*, pp. 85–88, 2006.

[30] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, "Adaptive supervoxel patch-based region classification in whole-body PET-CT," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Computational Methods for Molecular Imaging*, 2015.

[31] L. Bi, J. Kim, A. Kumar, L. Wen, D. Feng, and M. Fulham, "Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies," *Computerized Medical Imaging and Graphics*, vol. 60, no. Supplement C, pp. 3 – 10, 2017.

[32] P. Fischer, V. Daum, D. A. Hahn, M. Prümmer, and J. Hornegger, "Regression forest-based organ detection in normalized PET images.," in *Bildverarbeitung für die Medizin*, pp. 384–389, 2014.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[34] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[35] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, "From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors," *Journal of nuclear medicine: official publication, Society of Nuclear Medicine*, vol. 50, no. Suppl 1, p. 122S, 2009.

[36] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proceedings of the IEEE international conference on computer vision*, pp. 1742–1750, 2015.

[37] P. O. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1713–1721, 2015.

[38] J. Dai, K. He, and J. Sun, "Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation," in *ICCV*, pp. 1635–1643, 2015.

[39] M. Rajchl, M. C. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Damodaram, M. A. Rutherford, J. V. Hajnal, B. Kainz, *et al.*, "Deepcut: Object segmentation from bounding box annotations using convolutional neural networks," *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 674–683, 2017.

[40] P. Hu, B. Shuai, J. Liu, and G. Wang, "Deep level sets for salient object detection," in *CVPR*, vol. 1, p. 2, 2017.

[41] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.

[42] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.

[44] A. Fedorov, D. Clunie, E. Ulrich, C. Bauer, A. Wahle, B. Brown, M. Onken, J. Riesmeier, S. Pieper, R. Kikinis, *et al.*, "Dicom for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research," *PeerJ*, vol. 4, p. e2057, 2016.

[45] R. Beichel, B. J. Smith, C. Bauer, E. J. Ulrich, P. Ahmadvand, M. M. Budzevich, R. J. Gillies, D. Goldgof, M. Grkovski, G. Hamarneh, *et al.*, "Multi-site quality and variability analysis of 3D FDG PET segmentations based on phantom and clinical image data," *Medical physics*, vol. 44, no. 2, pp. 479–496, 2017.

[46] R. Beichel, M. Van Tol, E. J. Ulrich, C. Bauer, T. Chang, K. A. Plichta, B. J. Smith, J. J. Sunderland, M. M. Graham, M. Sonka, *et al.*, "Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach," *Medical physics*, vol. 43, no. 6, pp. 2948–2964, 2016.

[47] B. G. Booth and G. Hamarneh, "Dti-deformit: Generating ground-truth validation data for diffusion tensor image analysis tasks," in *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*, pp. 730–733, 2014.

[48] G. Hamarneh, P. Jassi, and L. Y. W. Tang, "Simulation of ground-truth validation data via physically- and statistically-based warps," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 5241, pp. 459–467, 2008.

[49] G. Hamarneh and P. Jassi, "Vascusynth: Simulating vascular trees for generating volumetric image data with ground truth segmentation and tree analysis," *Computerized Medical Imaging and Graphics (CMIG)*, vol. 34, no. 8, pp. 605–616, 2010.

[50] H. B. Yedder, A. BenTaieb, M. Shokoufi, A. Z. F. Golnaraghi, and G. Hamarneh, "Deep learning based image reconstruction for diffuse optical tomography," in *Medical Image Computing and Computer-Assisted Intervention Workshop on Machine Learning for Medical Image Reconstruction (MICCAI MLMIR)*, vol. 11074, pp. 112–119, 2018.

[51] A. Top, G. Hamarneh, and R. Abugharbieh, "Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3d image segmentation," in *Medical Image Computing and Computer-Assisted Intervention Workshop on Medical Computer Vision (MICCAI MCV)*, vol. 6533, pp. 204–213, 2010.

[52] A. Top, G. Hamarneh, and R. Abugharbieh, "Active learning for interactive 3d image segmentation," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 6893, pp. 603–610, 2011.

[53] A. BenTaieb and G. Hamarneh, "Topology aware fully convolutional networks for histology gland segmentation," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9900, pp. 460–468, 2016.

[54] Z. Mirikharaji and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 11073, pp. 737–745, 2018.

[55] M. Nosrati and G. Hamarneh, "Incorporating prior knowledge in medical image segmentation: a survey," Tech. Rep. arXiv:1607.01092, 7 2016.

[56] A. Saad, B. Smith, G. Hamarneh, and T. Moeller, "Simultaneous segmentation, kinetic parameter estimation, and uncertainty visualization of dynamic pet images," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 4792, pp. 726–733, 2007.