PET-Disentangler: PET Lesion Segmentation via Disentangled Healthy and Disease Feature Representations

by

Tanya Gatsak

B.Sc., University of Waterloo, 2019

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
Fall 2024

Declaration of Committee

Name:	**Tanya Gatsak**
Degree:	**Master of Science**
Thesis title:	**PET-Disentangler: PET Lesion Segmentation via Disentangled Healthy and Disease Feature Representations**
Committee:	**Chair:** Parmit Chilana Associate Professor, Computing Science
	Ghassan Hamarneh Supervisor Professor, Computing Science
	Saeid Asgari Committee Member Adjunct Professor, Computing Science
	Arman Rahmim External Examiner Professor Radiology, Physics and Biomedical Engineering University of British Columbia

Abstract

Positron emission tomography (PET) imaging is an invaluable tool in clinical settings as it captures the functional activity of both healthy anatomy and cancerous lesions. Developing automatic lesion detection methods for PET images is crucial since manual lesion segmentation is laborious and prone to inter- and intra-observer variability. We propose a 3D disentanglement method that learns robust disease features and predicts lesion segmentations by disentangling PET images into disease and normal healthy anatomical features. The proposed method, PET-Disentangler, uses a 3D UNet-like encoder-decoder architecture for feature disentanglement followed by simultaneous segmentation and image reconstruction. A critic network encourages the healthy latent features, which are disentangled from disease samples, to match the distribution of healthy samples and thus do not contain any lesion-related features. We train and evaluate PET-Disentangler on 3D PET images from the Cancer Imaging Archive (TCIA) whole-body FDG-PET/CT Dataset consisting of 1014 PET/CT scans, leveraging TotalSegmentator to obtain two anatomically aligned field-of-views of the whole-body scans referred to as the upper and lower torso regions. Compared to non-disentanglement segmentation methods, our quantitative results on the upper torso region show PET-Disentangler has similar performance while having the added advantage of visualizing, via the pseudo-healthy image, how a healthy (lesion-free) image might look like. Our quantitative and qualitative results on the lower torso show enhanced performance from our method as PET-Disentangler reduces the chances of incorrectly declaring high tracer uptake regions as cancerous lesions, since such uptake pattern would be assigned to the disentangled normal component.

Keywords: Positron Emission Tomography (PET); Image Segmentation; Disentangled Representations; Deep Learning; Computer Vision

Dedication

To all the physicians, researchers, and countless others involved in the fight against cancer: your dedication and collective efforts bring hope and progress to countless lives.

Acknowledgements

I would first like to thank my supervisor Prof. Ghassan Hamarneh for his invaluable guidance, support and patience during my graduate studies. His mentorship and passion for medical image analysis has allowed me to further develop my research skills while strengthening my appreciation for the field. I would also like to express my gratitude to the members of my examining committee, Prof. Saeid Asgari, Prof. Arman Rahmim, Prof. Parmit Chilana, for their time and feedback on this thesis.

I would also like to thank my labmates for their feedback on my research and many discussions over the years that have introduced me to new research and technical ideas. I would especially like to thank Kumar Abhishek and Kathleen Moriarty for their friendship, mentorship, and endless support during this time.

Finally, I'd like to thank my family and friends for their support and encouragement throughout my studies, as this source of strength has brought me here today.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CAM	Class Activation Mapping
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DOPA	Dihydroxyphenylalanine
FN	False Negative
FP	False Positive
FDG	Fluorodeoxyglucose
FCM	Fuzzy C-Means
GAN	Generative Adversarial Network
GP	Gradient Penalty
GT	Ground Truth
MRI	Magnetic Resonance Imaging
MRF	Markov Random Field
MIP	Maximum Intensity Projection
MONAI	Medical Open Network for Artificial Intelligence
PET	Positron Emission Tomography
ReLU	Rectified Linear Unit
SPADE	SPatially-Adaptive (DE)normalization
SUV	Standardized Uptake Value
TSNE	t-Distributed Stochastic Neighbor Embedding
TCIA	The Cancer Imaging Archive
TLE	Total Lesion Evaluation
TMTV	Total Metabolic Tumor Volume
TN	True Negative
TP	True Positive

TBR	Tumor-to-Background Ratio
WGAN	Wasserstein Generative Adversarial Network

List of Notations

X	Input image
X^-	Input image without tumour lesions (i.e. negative findings)
X^+	Input image with tumour lesions (i.e. positive findings)
E	Encoder
D_S	Segmentation Decoder
D_I	Image Decoder
C	Critic network
z_h	Healthy anatomy latent vector
z_h^-	Healthy anatomy latent vector obtained from X^-
z_h^+	Healthy anatomy latent vector obtained from X^+
z_d	Disease latent vector
c^-	Critic score from critic network for z_h^-
c^+	Critic score from critic network for z_h^+
M_{GT}	Ground truth segmentation mask
M	Predicted segmentation mask from D_S
M_0	Mask of zeros
R	Reconstructed X from $D_I(z_h, M)$
P	Pseudo-healthy image from $D_I(z_h, M_0)$

Chapter 1

Introduction

1.1 Background and motivation

Positron emission tomography (PET) is a medical imaging modality that measures the functional activity within the human body. Imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) capture mainly structural information whereas PET captures molecular-level biological changes throughout the body. PET imaging is widely used in oncology to detect, assess, and plan treatment for cancerous lesions, and it is also used in other areas of medicine including cardiology, neurology and psychiatry [60].

To acquire a PET image, a patient is injected with a radioactive tracer (i.e., radiotracer) and the patient will have the PET scan performed after a given time delay period. The radiotracer has radionuclides (i.e., radioactive isotopes) attached to a molecular compound, in which positrons will be emitted as the radionuclide decays. As soon as these positrons encounter electrons within the body, they will annihilate one another and create gamma rays travelling in opposite directions. The ring detector of the PET scanner detects these gamma rays and generates a sinogram that can then be reconstructed to produce a PET image [58]. The radiotracer will be distributed throughout the body where the molecular compound is used in physiological processes such that regions with higher radiotracer localization will correspond to the higher intensity values in the captured images. An example of a PET scanner is shown in Figure 1.1 where a patient lying down will be moved throughout the ring-shaped scanner.

Clinicians will qualitatively analyse these images as their understanding of the radiotracer distribution will allow them to delineate healthy radiotracer uptake from disease activity [44]. The most common radiotracer used in clinical settings for cancer management is fluoride-18 fluorodeoxyglucose (^{18}F–FDG) [49] as areas in the body that consume significant amounts of glucose will correspond to the bright areas in the images. Cells that require high levels of glucose include those in healthy organs and especially those in growing cancer tumours, making it a favourable choice in cancer detection.

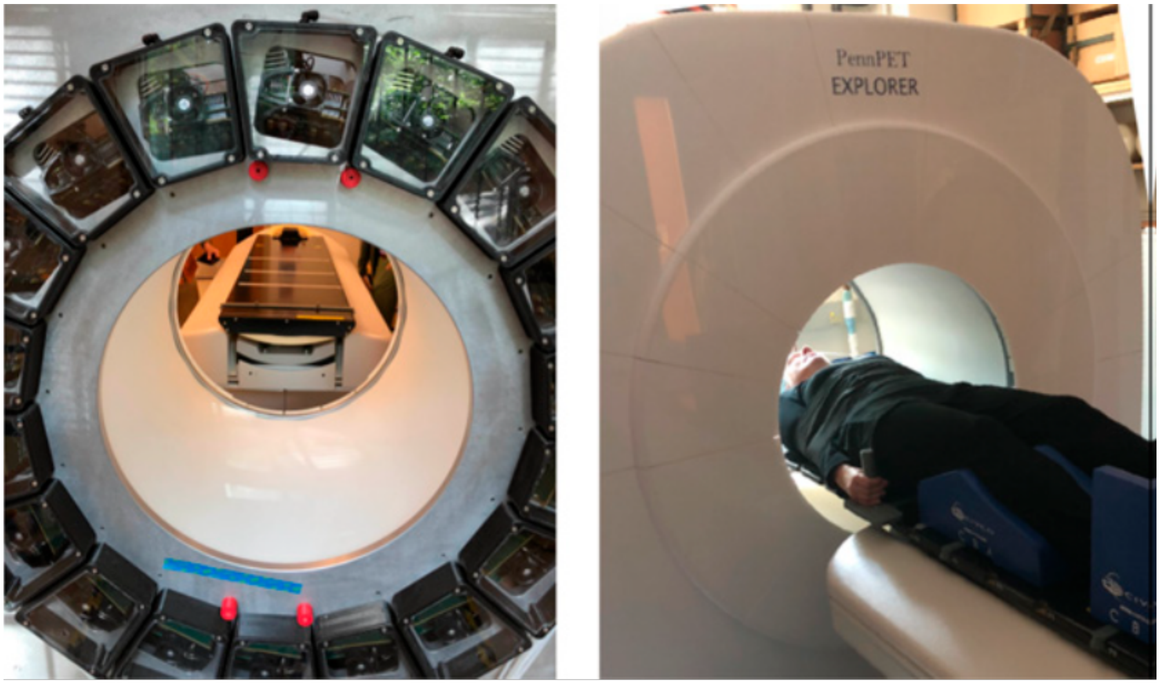

Figure 1.1: PennPET Explorer Whole-body PET scanner [36].

Other radiotracers include positron-labelled choline, such as ^{18}F-fluorocholine, as choline is metabolized as a precursor for cell membrane synthesis and cancer cells that are growing quickly have increased cell membrane requirements which correspond to higher uptake regions in obtained PET scans [45]. 18F-DOPA is another radiotracer that is used to study the dopaminergic pathway as 18F-DOPA is an analog of L-DOPA, an immediate precursor of dopamine [47]. Figure 1.2 shows coronal views of PET scans obtained using the three mentioned radiotracers, where each radiotracers distribution throughout the body creates different intensity uptake patterns.

Qualitative analysis of radiotracer distribution may be sufficient for lesion detection although quantitative measures are required for diagnosis and to measure treatment response over time. There are a number of methods to quantify the radiotracter distribution including tumor-to-background ratio (TBR) and total lesion evaluation (TLE), while the most widely used measure is the standardized uptake value (SUV) that is used to standardize PET image intensities. SUV represents the concentration of radiotracer throughout the body as follows:

$$SUV = \frac{C_{image}}{C_{whole\text{-}body}} = \frac{C^{*}_{image}}{\frac{C_{radiotracer}}{S}}. \tag{1.1}$$

SUV is calculated using the image-derived concentration of the radiotracer at acquisition time C_{image} normalized by the whole-body concentration of the radiotracer at time of injection $C_{whole\text{-}body}$. $C_{whole\text{-}body}$ is calculated using the concentration of the injected radiotracer at time of injection $C_{radiotracer}$ normalized by the patient's size S which can be accounted

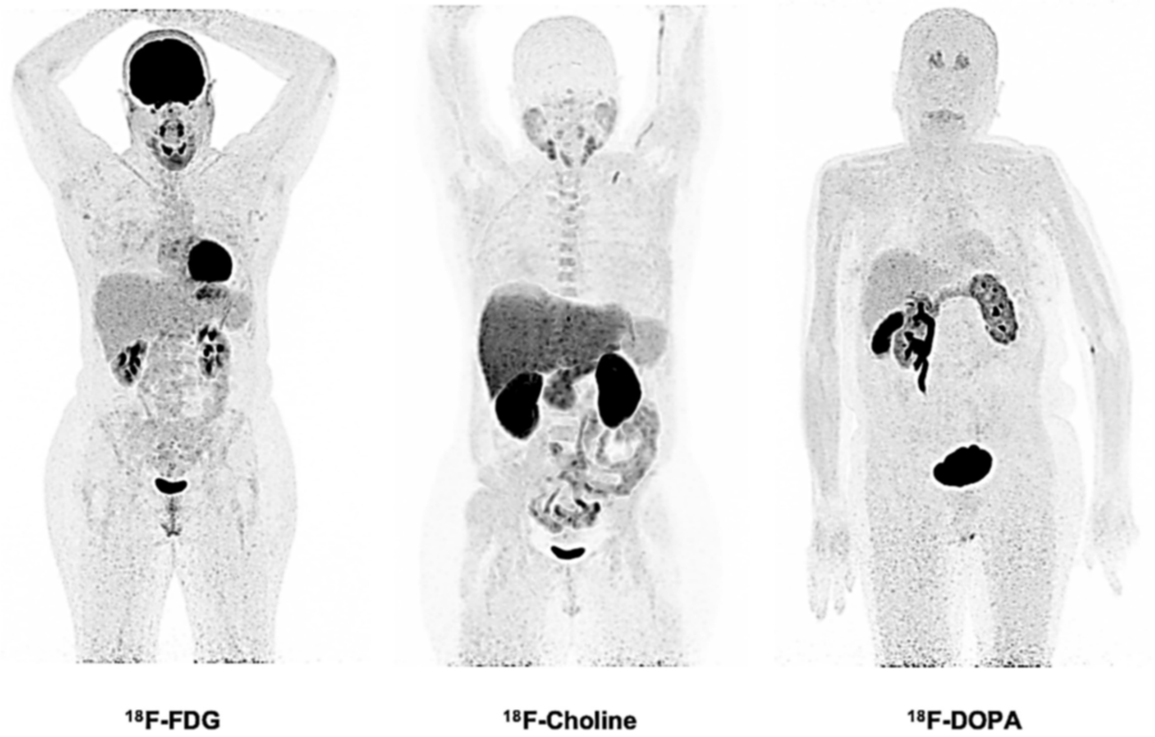

Figure 1.2: Coronal view examples of PET images from different radiotracers [15].

for using the patient's body weight, body surface area or lean body mass. C_{image} is further corrected by a decay factor that accounts for the radiotracer decay between time of injection to time of PET scan acquisition, producing C^{*}_{image}. There are a number of physiological, physical, and procedural factors that can influence the calculation of SUV including blood glucose concentration, partial volume effect, respiratory motion artifacts, image reconstruction algorithm, and resolution of the scanner used [17].

Nonetheless, SUV provides insight into metabolic activity and can help distinguish between healthy and abnormal levels of radiotracer uptake, in addition to providing measures of lesion aggressiveness and response to treatment. Segmentation of cancerous lesions facilitates diagnosis, treatment planning, and the monitoring of disease progression as segmentations can provide measures of total metabolic burden that can indicate disease prognostics. The development of automatic segmentation methods can further alleviate the error introduced and labour required for manual annotation.

1.2 The origins of automatic PET segmentation development

The origins of automatic PET lesion segmentation start with thresholding based techniques where the PET image is thresholded based on SUV values using either fixed threshold values, adaptive thresholding [16] or iterative thresholding methods [34]. Hofheinz et al. [29] provided an algorithm that can be considered an extension of adaptive thresholding by determining thresholds on a per-voxel level in which reference values for the calculation are obtained in each voxel's neighborhood of voxels. This algorithm improved upon previous thresholding techniques by being able to delineate strongly heterogeneous lesions while maintaining performance on more homogeneous lesions.

One of the first approaches moving away from thresholding-based techniques was to use segmentation algorithms based on hidden Markov models. In the work by Hatt et al. [26], a modified version of Hidden Markov chains that also introduced the fuzzy model was proposed with findings that the fuzzy nature can more realistically handle lesion borders. Another direction for segmentation was introduced at this time in which PET images were represented as a mixture of Gaussians. Aristophanous et al. [7] introduced a Gaussian mixture model for PET segmentation, where they modeled the PET distribution with three regions: background, target (for tumor regions), and uncertain. This method used one Gaussian density to represent the uncertain voxels, whereas the number of densities to represent the background and target were found to be 6 and 3, respectively. User input is also required to delineate a target region for modeling. More recently, Soffientini et al. [53] segmented lesions in PET images using an 8 class Gaussian mixture model clustering algorithm. This method required initial manual guidance of a rough lesion and background area delineation, and regularized the clustering by using Markov random field (MRF) for spatial priors.

In another direction, Yu et al. [68] approached PET/CT segmentation by using a decision-tree based k-nearest neighbour algorithm to classify each voxel as normal or abnormal based on PET and CT texture features for each voxel. Around the same time, Hatt et al. [27] introduced a 3D Bayesian statistical methodology that incorporated a fuzzy model by allowing a voxel to belong to one of a finite number of fuzzy levels in addition to one of the two hard classes of lesion or background. This method required an initialization step of delineating a region of interest. This methodology was extended in Hatt et al. [28] to include 3 hard classes that better handles heterogeneous uptake and has enhanced accuracy and robustness compared to the original methodology. Dewalle-Vignion et al. [14] aimed to delineate PET lesions by taking into account the uncertainty in labelling voxels as binary values, either belonging to the lesion volume or otherwise, by using possibility theory with maximum intensity projections and an initial selection of a 2D region of interest.

In the following years more methods were developing segmentation algorithms based on Markon random field models. Han et al. [23] produced a lesion segmentation by co-segmenting PET and CT using an MRF graph formulation. This method has an energy term that penalized the difference between PET and CT segmentations and produced the overall segmentation by computing a single maximum flow. Similarly, Song et al. [54] segmented lesions in both PET and CT scans by approaching co-segmentation as an energy minimization problem of an MRF model, in which the energy can be minimized using graph cuts and solved using a single maximum flow. This method encouraged consistency between segmentations produced for CT and PET images by using a consistency constraint that penalized differences between corresponding segmentation labels. This method also required initial user guidance of three seed points for each tumour volume, where one seed corresponds to the center point and two seeds are selected along the radius of the tumor volume.

Furthermore, Bagci et al. [9] approached co-segmentation of PET/CT and PET/MRI using a fully automated random-walk method that first finds foreground and background seeds using an automated method. To obtain these seeds, the image is first thresholded and foreground seeds are assigned to the voxel with maximum SUV in each foreground region, and the background seeds are assigned to the closest nearby voxels that have SUV below the threshold value. More recently, Ju et al. [35] used random walk and graph cut to approach PET/CT lesion segmentation, in which random walk is used to obtain initial seeds that are then used in the graph cut method. This method also used a context term to penalize differences between PET and CT segmentation maps, and furthermore introduced novel energy terms such as the downhill cost and 3D derivative cost for PET and a shape penalty cost for CT.

Another direction for segmentation has been based on clustering algorithms, particularly fuzzy c-means clustering. Belhassen and Zaidi [10] proposed a fuzzy c-means clustering algorithm (FCM-SW) that introduced spatial information and considered lesion heterogeneity by utilizing the results from the non-linear anisotropic diffusion filter and à trous wavelet transform in addition to the PET image in the FCM algorithm. Lapuyade-Lahorgue et al. [38] proposed a fuzzy c-means approach that provided a generalization of the algorithm by utilizing the Hilbertian norm where the norm parameter is estimated per image. This generalization was to improve accuracy and more accurately represent the non-Gaussian distributions within PET.

PET segmentation approaches have also been developed using the active contour model. Abdoli et al. [1] used an active contour model that replaced the input PET image with two transformed images, one being the anisotropic diffusion filtered image and the second being the contourlet transformed image, and included a curvature regularizing term to ensure the segmented lesions have a smooth surface. Zhuang et al. [73] used an active contour model that used transformed variations of the input PET image such as the histogram fuzzy c-means clustering, bilateral filter, and Gabor transformed image. The level set equation is also solved using the lattice Boltzmann method.

In more recent years, various other techniques have introduced their own novelties in PET segmentation. Hanzouli-Ben Salah et al. [25] approached co-segmentation of PET/CT images by using hidden Markov trees and applying the model to the original images in addition to the wavelet and contourlet transformed images. Tan et al. [56] proposed an adaptive region-growing algorithm (ARG_MC) that automatically determines the optimal relaxing factor for the algorithm using a maximum curvature strategy. Grossiord et al. [21] used machine learning techniques to first represent PET image as a component tree and obtain PET/CT descriptive features for each node in the tree, in which a random forest classifier used this representation to then predict the segmentation map.

We refer to reader to [17] for an in-depth review on pre-deep learning PET segmentation methods.

1.3 Deep learning techniques

1.3.1 PET lesion segmentation

Many studies have focused on segmenting lesions in PET images utilizing deep learning methodologies [39] both in standalone PET images and in hybrid scans (e.g. PET/CT, PET/MR). These studies have explored datasets encompassing a wide array of lesion locations.

Wang et al. [61] introduced a cascaded detection segmentation network for 18F-fluciclovine PET/CT prostate lesion segmentation, in which a fully convolutional one-stage object detection network is used to localize the lesions and provide a volume of interest to the segmentation network. Their method improved upon UNet and cascaded UNet comparisons.

For head and neck tumor segmentation, Afshari et al. [4] segmented lesions in PET using a fully convolutional network. A weakly supervised annotation approach that used ground truth bounding boxes and a novel Mumford-Shah piecewise constant loss term are leveraged to delineate lesions. Andrearczyk et al. [6] segmented head and neck lesions using VNets in a multi-modal setting with PET, CT, and PET/CT images, where they found the best performance using PET alone.

To segment lung lesions, Li et al. [41] used a fully convolutional network to create rough segmentations from CT scans that are then used as a prior to a PET fuzzy variational model to produce high quality segmentations. Zhong et al. [72] aimed to segment lung lesions in PET/CT by first using two separate 3D UNets to produce PET and CT based segmentation maps and then employed a graph co-segmentation model to refine the segmentation boundaries within the produced maps. Zhao et al. [71] introduced a 3D fully convolutional neural network for lung cancer PET/CT segmentation that used two separate VNet style networks for feature extraction with a subsequent feature fusion module that produced the final segmentation. This method used cropped regions of interest that contained the entire tumor area. Wu et al. [64] segmented lung lesions in PET by using an adversarial auto-encoder that took in 2D PET slices as input and produced the corresponding tumor free PET images, where lesions were found by identifying large differences between original input and model output.

To segment lymphoma lesions throughout a full-body PET/CT scan, Li et al. [40] trained a model that performed both segmentation and reconstruction tasks using DenseUNet. Jemaa et al. [33] also performed segmentation of lymphoma lesions throughout whole-body scans by first using 2D UNet to perform rough segmentation, then 3 separate 3D UNets for segmentation of the top, middle, and bottom regions, where the final segmentations are the average of 2D and 3D results. Sibille et al. [52] segmented lung and lymphoma lesions throughout the body by modifying AlexNet to take in multiple inputs (slices of PET/CT images around candidate center of mass, MIP, atlas position) and produced classification and segmentation of these regions. Blanc-Durand et al. [11] segmented lymphoma lesions throughout the body using nnUNet and randomly selected patches throughout the PET/CT scans. Hu et al. [30] segmented lymphoma lesions throughout the body using 4 networks trained separately, where the 4 outputs and original input are fused and passed to a single 3D convolutional layer to produce the final segmentation. Weisman et al. [63] segmented lymphoma lesions throughout the body using a 3D model of three DeepMedic models using different resolutions of the input, where the final segmentation is the intersection from the individual model's output. Liu et al. [42] approached PET lymphoma lesion segmentation using a multi-task framework in which lesion segmentation is jointly trained with prognosis prediction. Their method used a 3D UNet backbone that leveraged multi-scale features from the decoding stages for classification prediction, and used deep supervision for both segmentation and prediction tasks. Liu et al. [42] hypothesize that they outperformed competing methods because of the multi-task and deep supervision nature of their proposed method. Früh et al. [18] proposed a weakly supervised PET segmentation approach for the segmentation of lymphoma, melanoma, and lung cancer that first applied a VGG-based classification network to obtain slice level classification of tumor finding or no finding, in which CAMs of this network are used to obtain the segmentation mask. Yousefirizi et al. [67] introduced TMTV-Net for lymphoma, lung cancer, and melanoma lesion segmentation in PET/CT images for total metabolic tumor volume (TMTV) quantification. TMTV-Net is a cascaded segmentation network in which 5 3D UNets predict segmentations at different resolutions that are combined using a voting scheme to produce an intermediate prediction. The intermediate prediction is used with the original PET/CT in an additional 3D UNet to produce the final segmentation prediction. Yousefirizi et al. [67] found their cascaded approach improved upon state-of-the-art nnUNet and SWIN UNETR networks. A review of deep learning PET lesion segmentation methods can be found in [66].

1.3.2 Disentanglement

Disentangled representation learning aims to discern the underlying sources of variation within a dataset by isolating and encoding these distinct features of variation into separate latent vectors. Adopting a disentangled representation framework also provides explainability for neural network learning and allows the user the ability to manipulate the generated data. Disentangled representations have been seen in applications such as image-to-image translation, where domain specific features are separated from those that are domain invariant to map images from a source image domain to a target image domain. Disentanglement is often performed using generative models such as generative adversarial networks (GANs) [20] through regularization, where the disentangled components are learned implicitly.

The use of disentangled representations is growing in medical image analysis. For example, Han et al. [24] used disentanglement to separate ribs from chest x-ray images where they have seen an improvement in downstream task performance in disease classification and in detection on rib-suppressed chest x-rays. Another application is to disentangle healthy from disease as seen by Xia et al. [65], where they produced pseudo-healthy images and disease segmentation maps, while also producing reconstructions of the input given these two disentangled components. Their results showed they outperformed baselines for generating pseudo-healthy images and a conducted human study validated their improved performance.

Zhang et al. [69] also disentangled images into pseudo-healthy and disease components in an adversarial setting where the generator is competing against a segmentor rather than an image level classifier. This work also proposed a metric for healthiness and showed an enhancement technique to help downstream segmentation. Tang et al. [57] disentangled chest x-rays into a normal, healthy chest x-ray image and a disease saliency map, and produced a reconstruction image of the input, improving binary disease classification and detection performance. Kobayashi et al. [37] disentangled healthy anatomy from disease in the latent vector space and provided a framework for content-based image retrieval. Content-based image retrieval is to support comparative diagnostic reading, where physicians compare a given example to similar examples without findings or compare examples with similar abnormal findings.

1.4 Thesis Contributions

Automatic lesion segmentation remains a crucial area of development to alleviate the challenges associated with annotating medical images manually. The development of lesion segmentation methods continues to grow rapidly where most methods aim to refine the features of disease learned through various architectures, loss functions and data augmentation. Concurrently, novel computer vision and deep learning ideas are being introduced that may benefit when used in segmentation frameworks. In this thesis, we investigate the utility of integrating image disentanglement in the task of lesion segmentation, as better understanding of the components of the image may enhance the disease features learned. We propose PET-Disentangler to disentangle PET images into healthy and disease features in the latent space, and re-entangle these features during image reconstruction, to discover the utility of learning the components of PET for lesion segmentation. PET-Disentangler is the first 3D disentanglement approach for separating healthy and disease features. PET-Disentangler is also the first disentanglement of healthy and disease feature approach performed on PET data. Throughout our experiments, we show that the modelling of healthy anatomy via the healthy features reduces the false positive segmentation of healthy uptake patterns in which other methods fail.

This thesis extends upon our work that has been accepted and presented at the Annual Meeting of Society of Nuclear Medicine and Molecular Imaging. These extensions include training on more data, performing experiments on another anatomical region of interest, and additional evaluation methods.

> T. Gatsak, K. Abhishek, H. B. Yedder, S. A. Taghanaki, and G. Hamarneh, "PET-Disentangler: PET lesion segmentation via disentangled healthy and disease feature representations," *2024 SNMMI Annual Meeting Abstracts*, Journal of Nuclear Medicine, vol. 65 (supplement 2) 242461, 2024.

Chapter 2

Method

2.1 Overview

The current state of the art in automatic lesion detection is based on deep learning methods that aim at optimizing models to learn disease features. At the same time, the disentanglement of images in the latent space to identify sources of variation is proving beneficial in many settings [43], including image translation in both medical and non-medical applications. In this work, we approach lesion segmentation with a disentanglement framework to disentangle PET images into disease features and normal healthy anatomical features for a robust and explainable segmentation method.

In this work, we introduce a novel PET segmentation method, PET-Disentangler, that disentangles 3D PET images in the latent space into healthy and disease components. PET-Disentangler uses the disease features for lesion segmentation prediction, the healthy features to estimate pseudo-healthy images per input, and re-entangles both healthy and disease features for full reconstruction. PET-Disentangler is comparable to segmentation-only baselines, although it enhances the lesion segmentation task by providing explainability in the form of a pseudo-healthy image as to what the model expects the lesion-free image to look like per given input.

Furthermore, PET-Disentangler shows that learning the healthy component provides a solution to a critical challenge in PET lesion segmentation [5] where segmentation models can incorrectly segment healthy, high-intensity areas as disease. This challenge has seen solutions that focus on localizing healthy, high intensity regions [3] whereas the proposed PET-Disentangler can both capture the healthy anatomy features and delineate lesions through the learned disentangled representations.

2.2 Proposed network architecture

In this work, we propose approaching lesion segmentation using a deep learning disentanglement framework that learns to separate an input PET image into healthy and disease

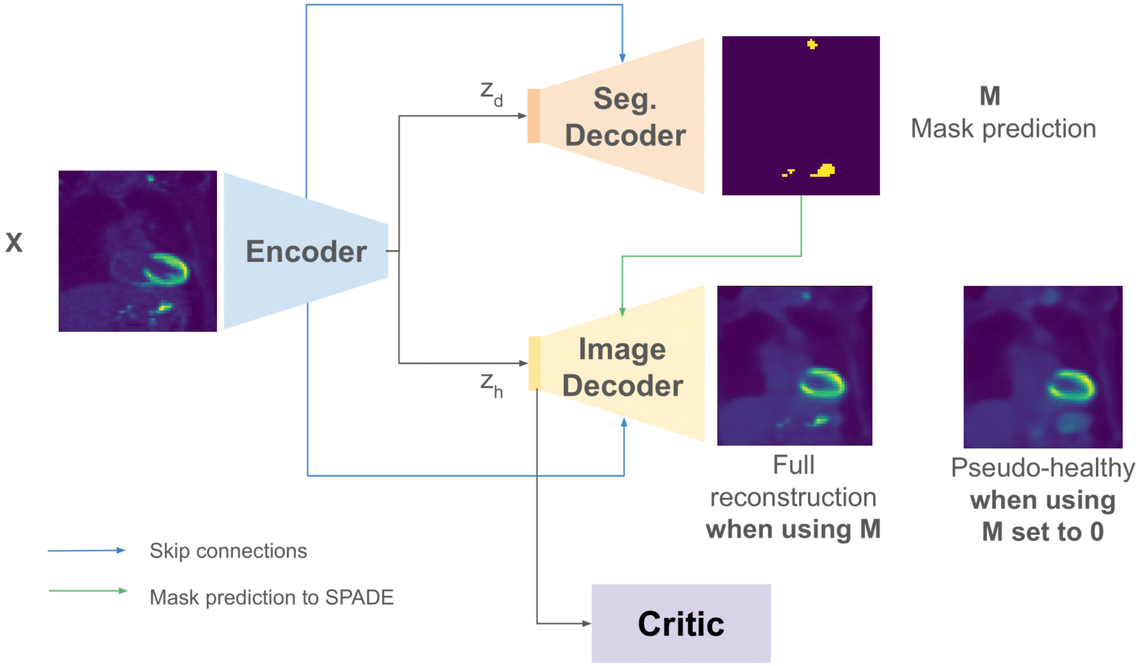

Figure 2.1: Proposed architecture of PET-Disentangler that consists of an encoder, segmentation decoder, reconstruction decoder and a critic network to enforce feature disentanglement. The encoder produces two latent vectors z_h and z_d that represent the healthy and disease features, respectively. The black arrows represent the utilization of features throughout the network. The blue arrows represent the skip connections between the encoder and two decoders. The green arrow represents the use of the mask prediction in the image decoder to combine healthy and disease features, via SPADE blocks, for image reconstruction.

features in the latent space. The disentanglement framework facilitates the learning of disease features in a segmentation prediction path and via re-entangling healthy and disease features in an image reconstruction path. This framework includes an additional critic network that is used to ensure healthy features do not contain lesion features. An overview of the proposed architecture is shown in Figure 2.1.

The proposed method, PET-Disentangler, adopts a UNet architecture that is further extended for the task of disentanglement. UNet is a fully convolutional network initially introduced for and widely utilized in semantic segmentation [51]. Additionally, UNet can be adapted for various other tasks, such as image generation. The name UNet arises from the u-shaped architecture of the network, consisting of symmetric contracting and expanding paths, where the input and output from the model have the same spatial resolution.

The contracting path, also known as the encoder, extracts features from the input via a series of convolution, batch normalization, ReLU and maxpooling operations to obtain a compressed representation of the features, referred to as the bottleneck latent vector. The expanding path, also known as the decoder, uses the bottleneck latent vector to generate the output for the given task through a series of convolution, transpose convolution, batch normalization and ReLU operations. Between encoding and decoding blocks of the same spatial resolution, skip connections are used to pass output from an encoding block to the corresponding decoding block where it will be concatenated with the input features for decoding. In deep convolutional networks, important features can be lost and gradients can diminish with the depth of the network. Skip connections are used to provide features that otherwise can be lost and provide a better flow of gradients during backpropagation, which facilitates the overall model learning better representations and generating high quality output.

In PET-Disentangler, the UNet architecture is extended from 2D to 3D and modified such that there is one encoder and two decoders for segmentation prediction and image reconstruction. The encoder takes as input a PET image, X, and outputs two bottleneck latent vectors, z_h and z_d, which encode the healthy and disease features of the input image, respectively. The disease features are passed to the segmentation decoder that predicts a segmentation mask, M. The healthy features are passed along with the segmentation mask to the image decoder to re-entangle the healthy and disease features and produce a reconstruction of the input, R. Skip connections are used between each encoder and segmentation decoder block. In contrast, the skip connections between the encoder and image decoder are modified to prevent introducing disease features through these connections as the disease features, if any exist, should be introduced via the segmentation mask. The ground truth segmentation masks, GT, are binary such that voxels labelled as 0 belong to either background or healthy anatomy whereas voxels labelled as 1 belong to cancerous lesions.

2.2.1 Modified encoder architecture

The encoder, E takes in a 3D PET image of size $64 \times 64 \times 64$ voxels and applies a series of encoding blocks, consisting of a 3D convolution, batch normalization and ReLU, repeated once, to produce the latent vectors z_h and z_d:

$$z_h, z_d = E(X). \tag{2.1}$$

The first encoding block takes X as input and produces an output vector that is then split into healthy and disease features via the application of two separate encoding blocks. For the rest of the encoding process, separate encoding blocks, followed by max-pooling to reduce the dimensions, are applied to healthy and disease features with identical operations. This design is to ensure the spatial dimensions are the same between vectors at each stage of encoding without sharing weights in the encoding process. All convolutions in the encoder have a kernel size of $3 \times 3 \times 3$ with a stride of 1.

2.2.2 Segmentation decoder and prediction

The disease features z_d are passed to the segmentation decoder, D_S, with the corresponding skip connections to generate a segmentation mask prediction, M, via a series of decoding blocks, as described by:

$$M = D_S(z_d). \tag{2.2}$$

Each decoding block consists of a transposed 3D convolution to upsample the spatial dimensions of the feature vector, concatenation between the feature vector and skip connection from the encoder, followed by a sequence of 3D convolution, batch normalization and ReLU, repeated once as seen in the encoding blocks. The transpose convolution layers have a kernel size $2 \times 2 \times 2$ and stride of 2. The final layer of the decoder consists of a single convolution followed by sigmoid activation. Given the task at hand is binary segmentation, we can use the decoder to produce output with two channels where each class is represented by its own channel. We use the argmax operation to obtain which channel, and subsequently which class label, has higher activation at each voxel to generate the segmentation prediction.

To optimize the segmentation mask prediction, ComboLoss [55] is used between the predicted mask and the ground truth mask, M_{GT}:

$$L_{seg} = L_{ComboLoss}(M, M_{GT}) = L_{Dice} + L_{cross-entropy}. \tag{2.3}$$

ComboLoss takes the weighted sum of a Dice loss term and cross-entropy loss term. The Dice loss is commonly used for semantic segmentation and measures the overlap between ground truth and predicted segmentations whereas the cross-entropy loss penalizes false positive and false negative results. In our experiments, the Dice and cross-entropy terms are weighted equally. The Dice loss is written as follows:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} m_i \cdot \hat{m}_i}{\sum_{i=1}^{N} m_i + \sum_{i=1}^{N} \hat{m}_i} \tag{2.4}$$

where m_i and $\hat{m}_i$ represent the ground truth and predicted label at voxel i, respectively, over the total number of voxels N. The cross-entropy loss is written as follows:

$$L_{cross-entropy} = -\frac{1}{N}\sum_{i=1}^{N} w_0\, m_i\, log(\hat{m}_i) + w_1(1 - m_i)\, log(1 - \hat{m}_i) \tag{2.5}$$

where, in our experiments, we use equal weights for w_0 and w_1.

2.2.3 Image decoder and reconstruction

The image decoder, D_I, uses the healthy features z_h and corresponding skip connections along with the segmentation prediction M during decoding to generate image reconstructions. When the mask prediction contains lesion features, the output of the image decoder will be the full reconstruction of the input, R:

$$R = D_I(z_h, M). \tag{2.6}$$

When the mask prediction is set to empty such that it contains all zeros, M_0, the image decoder will only be using healthy features to generate the output and will therefore predict a healthy estimate of the input image, P:

$$P = D_I(z_h, M_0). \tag{2.7}$$

For input images that have no tumours, the reconstruction and the healthy estimate should be identical.

The image decoder re-entangles the healthy and disease features using spatially-adaptive normalization (SPADE) blocks to combine the healthy features at each resolution during image decoding with a downsampled version of the segmentation mask of the same spatial dimensions. SPADE blocks were initially introduced for semantic image synthesis where semantic masks are used to produce photorealistic images [50]. This work was motivated by the limitation of using semantic masks directly as input to an image synthesis model as normalization layers tend to lose the information contained in the masks. SPADE alternatively combines features from semantic masks with generated features in a conditional normalization layer, where activations are adjusted by an affine transformation based on external data. In SPADE, the activations are modulated by learned scale and bias tensors that are obtained from applying a two-layer convolution block to the segmentation mask, in which the scale and bias are then element-wise multiplied and added to the previous features. Figure 2.2 describes the two-layer convolution structure of the SPADE block as seen in the original paper. To obtain downsampled masks, a series of convolutions with kernel size $3 \times 3 \times 3$ and stride of 2, followed by batch normalization and LeakyReLU are repeatedly applied to the segmentation prediction.

The decoding blocks in the image decoder are similar to those in the segmentation decoder, with the key difference being the introduction of SPADE blocks following the transpose convolutions to combine healthy and disease features. Another difference is that the skip connections are removed from the last three decoding blocks to prevent disease features from potentially being introduced from outside of the mask prediction. The final layer of the image decoder consists of the following series of operations: convolution, batch normalization, ReLU, convolution, sigmoid activation. All convolutions in the segmentation and image decoder, including the SPADE blocks, have a kernel size of $3 \times 3 \times 3$ with a stride of 1.

To optimize the model to learn the reconstruction R, L_1 and L_2 reconstruction losses are used:

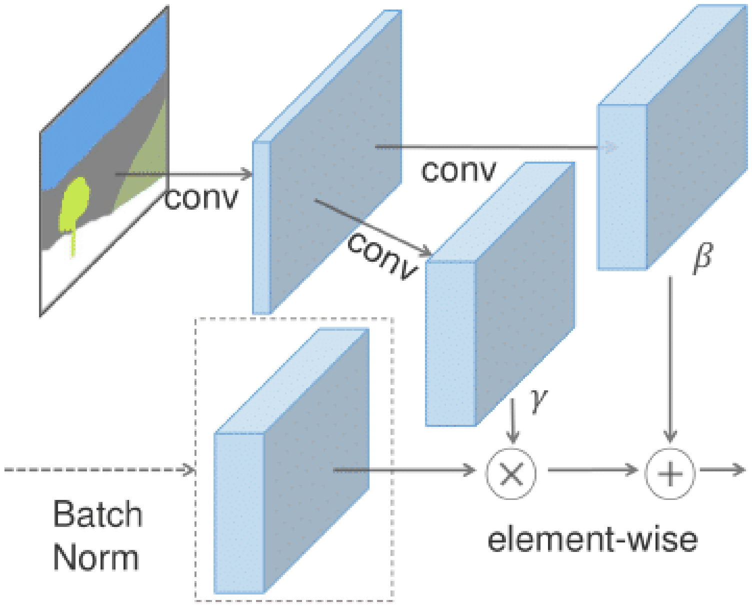

Figure 2.2: SPADE block structure [50].

$$L_{recon} = ||X - R||_1 + ||X - R||_2. \tag{2.8}$$

The L1 loss minimizes the mean absolute error between X and R whereas the L2 loss minimizes the mean squared error between the two. Both loss functions are used in image reconstruction tasks, where L2 is commonly seen in image restoration and L1 has seen more advantageous in super resolution applications. L2 is beneficial as it penalizes large errors and facilitates faster convergence but it also tends to produce over-smoothed output images. In contrast, L1 can produce images with sharper details and has also been shown to reach better minima during optimization than L2 as the L2 tends to get stuck in local minima [70]. Overall, the L1 and L2 loss functions are used in combination to obtain the benefits of both independent functions.

2.2.4 Critic network for healthy distribution matching

To ensure that z_h only contains features relating to healthy anatomy regardless of whether the input image has lesions, a critic network C is used in a similar manner to Kobayashi et al. [37] in which the Wasserstein GAN (WGAN) with gradient penalty is used to align healthy features to a healthy distribution. A set of images X can be partitioned into those without any tumour lesions (i.e., negative findings), X^-, and those with tumor lesions (i.e., positive findings), X^+, where the corresponding healthy feature vectors are z_h^- and z_h^+, respectively. Ideally, the healthy feature vectors should only contain features for healthy anatomy regardless of whether the input image has disease, and the distributions corresponding to the sets of healthy feature vectors should match. As such, WGAN with gradient penalty is used to ensure the distribution of healthy feature vectors obtained from X^+ matches the distribution obtained from X^-. As the only features that could cause a discrepancy between distributions correspond to those of disease, the process of distribution matching should eliminate the disease features ensuring z_h from all examples only contain healthy features.

The loss to optimize the critic network is described as follows:

$$L_{critic} = \Big(-\Big(C(z_h^-) - C(z_h^+)\Big) + \lambda_{GP}\left(||\nabla_{z_m} C(z_m)||_2 - 1\right)^2\Big) \cdot w_c. \tag{2.9}$$

The first term in the critic loss L_{critic} corresponds to the Wasserstein distance between the z_h^- and z_h^+ distributions. The critic is learning to identify healthy feature vectors originating from the two different distributions such that this distance is maximized. To ensure the distance is maximized while aiming to minimize the overall loss, the negative of the distance is used.

To use Wasserstein distance in the provided form, the critic network must be Lipschitz continuous meaning that the derivative of the network is less than or equal to a constant K everywhere, where $K \geq 0$. The original WGAN paper proposed to enforce this constraint by clipping the weights to a certain value range $[-c, c]$, although the most significant limitation in this approach is the careful selection of a clipping value parameter [8]. Selecting a parameter too large results in slow training and potentially exploding gradients whereas a parameter too small results in vanishing gradients. Gradient penalty was introduced to replace the usage of weight clipping by enforcing 1-Lipschitz continuity in the critic by adding a term to the critic loss that ensures the gradients have L2 norm values close to 1 everywhere [22].

The gradient penalty corresponds to the second term in L_{critic}, where λ_{GP} scales the magnitude of contribution of the gradient penalty to the critic loss and w_c weighs the overall loss. The gradient penalty is calculated by first interpolating an image z_m between z_h^- and z_h^+ using a random variable α to weigh the contribution of each vector, as follows:

$$z_m = \alpha z_h^- + (1 - \alpha) z_h^+. \tag{2.10}$$

Using the critics score on this interpolated image, the gradient and gradient norm are calculated and the final value is the squared difference between the norm and 1.

As the critic network is learning to identify the difference between sets of healthy feature vectors, the encoder is trying to produce healthy feature vectors that appear from the same distribution. As such, the corresponding WGAN term used in the overall model loss is referred to as the pseudo-healthy loss $L_{pseudo\text{-}healthy}$, described by:

$$L_{pseudo\text{-}healthy} = -C(z_h^+). \tag{2.11}$$

The overall model is trying to maximize the critic's score on the "fake" healthy vectors z_h^+ such that they are recognized as "real" healthy vectors, where the negative of this value is used as the overall loss to be minimized.

The architecture of the critic network consists of three 3D convolutions with LeakyReLU activations with slope of 2 in between. The first convolution has a kernel size of $4 \times 4 \times 4$ with stride of 2, whereas the second and third convolutions have a kernel size of $3 \times 3 \times 3$ with stride of 1.

2.2.5 Overall objective function

To train the proposed architecture, the critic network is optimized separately from the rest of the network components. The critic loss is used to optimize the critic whereas the overall objective function to optimize the encoder, segmentation decoder, and image decoder is:

$$L_{overall} = w_s\, L_{seg} + w_r\, L_{recon} + w_{ph}\, L_{pseudo\text{-}healthy}. \tag{2.12}$$

The parameters w_s, w_r, and w_{ph} refer to the weights of the contribution for the segmentation, reconstruction, and pseudo-healthy losses to the overall loss, $L_{overall}$.

2.3 Experimental design and implementation details

2.3.1 Dataset

We use the Whole-body FDG-PET/CT dataset [19] obtained from The Cancer Imaging Archive (TCIA) that is publicly available and has been used in both autoPET and autoPET2 challenges. This dataset consists of 900 patients and a total of 1014 scans, where 513 scans have no cancerous lesions and 501 scans have lesions from either lymphoma, melanoma, or lung cancer. Although additional information is provided to identify cancer type per scan, we do not consider cancer type when predicting lesion segmentation masks as the goal of this work is to investigate the impact of disentanglement on lesion segmentation.

In the dataset, 819 patients have a single scan, 59 patients have two scans, 14 patients have 3 scans, 5 patients have 4 scans and 3 patients have 5 scans. For every scan, the axial slices have dimensions of 400×400 pixels where the number of axial slices ranges from 200 to 661 slices throughout the dataset. The voxel size for every scan is $2.04 \times 2.04 \times 3.0$mm.

These scans were obtained at the University Hospital Tübingen using a single Siemens Biograph mCT PET/CT scanner, and the data is provided in anonymised DICOM files. A radiologist and nuclear medicine physician analysed the dataset to identify and manually segment all lesions in agreement. The authors of the dataset have also made available preprocessing scripts to register and resample the available PET and CT images, in addition to transforming the PET images from raw PET intensities to SUV intensities.

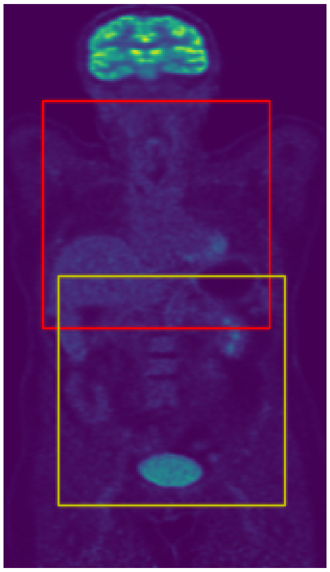

Figure 2.3: Upper (red) and lower (yellow) torso regions selected based on aorta and bladder locations.

2.3.2 Data preprocessing

A core contribution of our method is the ability to model and learn the healthy anatomy component of given PET images to ultimately distinguish disease features when they are present. This requires volumes of the same relative area between all healthy and disease examples to learn from. TotalSegmentator [62] is leveraged to obtain anatomical segmentations for each whole-body scan, where these segmentations can act as landmarks to center spatial croppings around to obtain a set of aligned subvolumes for the dataset. TotalSegmentator is an automatic segmentation tool built off of a nnUNet backbone [31] that can produce segmentations for 117 classes of anatomical structures in CT images. We can leverage the aligned CT scans corresponding to each PET scan in the dataset to obtain anatomical structure segmentations based on the CT images and use the corresponding locations in the PET images. As this dataset consists of scans that are mainly covering from eyes to thighs, two subregions are selected to cover the span of the full-body scans.

The first selected region covers the anatomy of the upper half of the torso, capturing the lung and heart field-of-view, and is obtained using the aorta segmentation mask as the aorta lays roughly in the center of this region. To produce the set of upper-region cropped volumes, the middle voxel location of the aorta from the aorta mask is used to center a $128 \times 128 \times 128$ crop in the PET scans.

The second region covers the lower half of the torso, with a field-of-view containing the kidneys and bladder. This region is obtained using the location of the bladder, where similar to the aorta-based cropping, a $128 \times 128 \times 128$ volume is cropped from the PET scan using the center pixel coordinate of the bladder segmentation mask. An example of both regions is shown on a coronal slice in Figure 2.3, where the upper-torso region is outlined by the red box and the lower-torso region is outlined by the yellow box, with a slight overlap between the two regions.

As larger volumes require more computational resources, the cropped volumes are resized to smaller dimensions to maximize the amount of volumes that can be used in a single batch during model training. The volumes are resized to $64 \times 64 \times 64$ voxels using bilinear interpolation. The PET scans have SUV values clipped from [0, 15] and further normalized between [0, 1].

The dataset is split into approximately 80:10:10 splits for training, validation, and testing sets where the ratio of healthy and disease examples is approximately the same between each split. For the upper-torso region, 460 healthy and 352 disease examples are used for training, and 58 healthy and 43 disease examples are used for both validation and testing. For the lower-torso region, 566 healthy and 244 disease examples are used for training, and 71 healthy and 31 disease examples are used for validation and testing.

2.3.3 Experiments

As the goal of this work is to investigate whether disentanglement of healthy and disease features can improve lesion segmentation in PET images, PET-Disentangler will be compared to a baseline segmentation-only method in addition to a few variations to study the importance of each modification introduced in PET-Disentangler.

We compare PET-Disentangler to the following methods:

- **SegOnly**: a method that performs segmentation only, with a 3D UNet architecture of one encoder that produces disease features that are passed to a segmentation decoder that produces the final segmentation prediction. This method is the baseline segmentation method for which we introduce or modify components in the following experiments to investigate their impact on segmentation performance.

- **SegRecon**: a method for simultaneous segmentation and reconstruction without feature disentanglement, that uses a modified 3D UNet architecture in which the encoder produces two sets of features: disease features that are passed to a segmentation decoder and image features that are passed to an image reconstruction decoder. In this method, we are observing the impact of learning more than disease features has on lesion segmentation, as PET-Disentangler extends upon this method by learning semantically disentangled features in addition to re-entangling these features.

- **SegReconHealthy**: a method that learns disease and healthy features via a modified 3D UNet architecture, in which the encoder produces a set of healthy and disease features, and the disease features are passed to a segmentation decoder and the healthy features are passed to an image reconstruction decoder. In SegReconHealthy, only examples that are healthy are passed to the reconstruction path of the model, such

Table 2.1: Summary of deep learning components used in each experiment

Method	Segmentation	Reconstruction	Disentanglement	SPADE	Critic
SegOnly	✓	✗	✗	N/A	N/A
SegRecon	✓	✓	✗	N/A	N/A
SegReconHealthy	✓	Healthy examples	✗	N/A	N/A
PET-Disentangler-withoutCritic	✓	✓	✓	✓	✗
PET-Disentangler-withoutSpade	✓	✓	✓	✗	✓
PET-Disentangler	✓	✓	✓	✓	✓

that the healthy features encode only healthy anatomy and the decoder produces only healthy reconstructions. This method is used to investigate the impact of simply learning healthy and disease features has on the task of lesion segmentation. However unlike PET-Disentangler, SegReconHealthy does not see examples with disease in the image reconstruction path, and furthermore it does not re-entangle both healthy and disease features for reconstruction.

- **PET-Disentangler-withoutCritic**: a method of the optimal PET-Disentangler configuration without the critic network to enforce that the healthy features only contain healthy anatomy. This method assumes that the features in the healthy vector may contain disease features as it is not regularized via the critic network. Therefore, it investigates how incorporating disease features during image reconstruction affects the overall lesion segmentation task.
- **PET-Disentangler-withoutSpade**: a method of the optimal PET-Disentangler configuration without SPADE blocks in the image decoder such that only healthy features are used during image reconstruction. This method is used to investigate the impact of disentangling healthy and disease features alone has on lesion segmentation, compared to PET-Disentangler where additionally re-entanglement of healthy and disease features is learned in the SPADE blocks.
- **PET-Disentangler-n**: the proposed method for PET-Disentangler with modified skip connections, where n represents the skip connections dropped between the first n encoding blocks and the last n decoding blocks. These methods are used to investigate the impact of skip connections of the quantitative results of lesion segmentation and the qualitative impact on healthy and disease feature disentanglement. The optimal skip connection configuration based on the qualitative and quantitative results will correspond to the final PET-Disentangler architecture.

Table 2.1 summarizes the deep learning components used in each experiment. As the SPADE and critic components are only used during disentanglement, the non-disentanglement experiments have these fields set to N/A to specify they are not applicable components.

Table 2.2: Losses used during each experiment

Method	L_{seg}	L_{recon}	$L_{pseudo\text{-}healthy}$	L_{critic}
SegOnly	All examples	—	—	—
SegRecon	All examples	All examples	—	—
SegReconHealthy	All examples	Healthy examples	—	—
PET-Disentangler-withoutCritic	All examples	All examples	—	—
PET-Disentangler-withoutSpade	All examples	All examples	Disease examples	All examples
PET-Disentangler	All examples	All examples	Disease examples	All examples

Table 2.2 further describes each method with respect to the losses and data seen during training. For learning disease features, each experiment uses all the examples during training and uses L_{seg} between the predicted and ground truth segmentation masks. SegRecon and SegReconHealthy both perform image reconstruction using L_{recon} differing in that SegRecon uses all examples and learns image features whereas SegReconHleathy only uses healthy examples to learn healthy features. PET-Disentangler extends upon these experiments to use all examples for image reconstruction, and additionally uses L_{critic} and $L_{pseudo\text{-}healthy}$ to enforce the learning of healthy features where L_{critic} uses data from all examples and $L_{pseudo\text{-}healthy}$ only uses data from disease examples.

2.3.4 Evaluation methods

To evaluate the proposed method, we will first evaluate the observed disentanglement of PET-Disentangler-n variants to determine an optimal configuration. Subsequently, we will evaluate the lesion segmentations produced by each method.

Determining optimal PET-Disentangler configuration based on feature disentanglement: To determine the optimal configuration of PET-Disentangler, we will investigate which configuration is successfully disentangling healthy and disease features. Disease features may be incorrectly introduced in the method, referred to as disease feature leakage, in two ways. The first being in the healthy vectors if the input image has disease and the distributing matching is not correctly removing these disease features. This prevents the model from learning to disentangle the healthy and disease components. The second possibility is via the image decoder through the skip connections as these skip connections are not regularized to ensure they contain only healthy information as done with the healthy feature vector. This disease leakage can prevent the image decoder from learning how to re-entangle the healthy features with the disease features for image reconstruction. The optimal configuration of PET-Disentangler will be tuned based on the number of skip connections used between the encoder and image decoder. There are 5 possible skip connections that can be used and dropping the last 1 to 5 skip connections will be analysed.

To identify whether the healthy and disease features are disentangled, TSNE plots capturing the 2D embeddings of z_h and z_d are analysed to determine whether these features belong to distinct clusters capturing different semantic information or whether there is overlap between the information stored. Furthermore, TSNE plots of the 2D embeddings belonging to z_h^- and z_h^+ will be analysed to determine if the healthy vectors obtained from disease input, z_h^+, appear the same as z_h^-. This comparison is to determine if the healthy features belong to the same distribution or whether there is disease leakage in z_h^+. The TSNE plots will be generated for each PET-Disentangler-n variant to determine if the results are consistent regardless of the skip connections used.

To identify whether skip connections introduce disease features, the healthy reconstructions from input examples with disease, generated by variants of PET-Disentangler-n, will be visually inspected for the appearance of lesions.

Evaluating segmentation results: To evaluate the quality of the segmentations produced, we use the Dice-Sørensen coefficient (also referred to as Dice similarity coefficient) to measure the overlap between ground truth and predicted segmentations, normalized by the size of both segmentations. Given a mask prediction M and ground truth mask, M_{GT}, the Dice can be computed as:

$$Dice(M, M_{GT}) = \frac{2|M \cap M_{GT}|}{|M| + |M_{GT}|}. \tag{2.13}$$

We can further rewrite the Dice coefficient using confusion matrix variables, true positive (TP), true negative (TN), false positive (FP), false negative (FN), as follows:

$$Dice(M, M_{GT}) = \frac{2TP}{2TP + FP + FN}. \tag{2.14}$$

Referring to Equation 2.13, if a given example has lesions and the model correctly predicts the exact lesion segmentation, then the size of the ground truth segmentation, predicted segmentation, and overlap between the segmentations will all be equal. In this scenario, the size of the overlap doubled divided by the sum of both segmentation sizes will be equal and the perfect Dice value will be 1. In the case where lesions exist but the predicted segmentation has no overlap with the ground truth, the Dice value will go to 0 corresponding to the worst Dice possible.

Furthermore, if an example is healthy and the model correctly predicts no lesions, then both the numerator and denomator would go to zero. As this is the ideal scenario, an offset variable is added to both the numerator and denominator such that when there are no lesion pixels in the ground truth or predicted segmentations, the Dice value will go to 1. Dice values below 1 for healthy examples indicate the presence of false positives in the lesion prediction.

We also measure the sensitivity of the lesion segmentation which is the proportion of lesion voxels that are accurately detected as lesion voxels. The sensitivity is calculated based on confusion matrix variables as follows:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{2.15}$$

The sensitivity can be calculated per volume and lesion level, where the volume level sensitivity would simply use the confusion matrix variables directly obtained from the predicted and ground truth segmentations. For lesion level segmentation, the ground truth segmentation is first processed by a connected components algorithm to identify each individual lesion and its location. Once these ground truth lesions are identified, the sensitivity per lesion can be calculated by locating the corresponding lesions in the predicted segmentations.

Ideally, we would report more lesion level metrics in addition to sensitivity although reporting metrics such as specificity and precision would require calculation from non-trivial graph matching algorithms. This is because these metrics require the false positive values associated with each ground truth lesion although to obtain false positives would require graph-matching algorithms to match pixels from the ground truth to predicted segmentation masks. Thus for the current investigation we will only report lesion sensitivity whereas future work can explore additional lesion-level metrics.

2.4 Training and implementation details

All development of this work was done in the Python programming language and PyTorch was used to develop the neural network architectures, in addition to the use of predefined loss functions, optimizers, and dataset handling operations. The Medical Open Network for Artificial Intelligence (MONAI) library [12] is an open-source library built off of PyTorch to support medical imaging workflows and it was used in this work for their predefined loss functions and data transformation operations such as intensity scaling, resizing, and transformation of data into tensor objects.

To train the proposed method, two separate optimizers are required to optimize the critic and overall model architecture. We use the Adam optimizer to optimize both the critic and overall model architecture, in addition to optimizing each of the models in the proposed experiments, with a learning rate of $1e$-3. In each experiment, the models are trained using 300 epochs with a batch size of 4 consisting of 2 healthy and 2 disease examples per batch. Although the model is trained for 300 epochs, the model may not be at it's optimal performance at the 300th epoch, such that a better performing model was seen earlier during training. To obtain the best performing model for each experiment, the version of the model with the lowest ComboLoss on the validation set is saved as the best model. To balance the model's objective function during training, the weights on the individual loss terms were tuned on the validation set and found empirically to be 100, 10, 0.001, and 0.01 for w_s, w_r, w_{ph}, and w_c, respectively.

To train the critic and overall model to optimal solutions, the critic is updated multiple times per training iteration while the model is updated only once. As introduced in the original WGAN paper, this is to train the critic well such that the generator model has high quality gradients to learn from [8]. We refer to the number of critic updates per iteration as n_{critic} and set it to 5 as proposed in the original work.

To ensure reproducibility of the experiments performed, we use random seeding at the beginning of each experiment to initialize the random seed of Python, NumPy, and PyTorch variables. To execute the experiments, the hardware used consists of an AMD EPYC 7543 32-core CPU with 64GB of RAM and a single NVIDIA RTX A5000 24GB GPU. As the purpose of this work is to investigate the utility of image disentanglement in lesion segmentation, we do not pretrain any PET-Disentangler path or module such that we can have a direct performance comparison to non-disentanglement methods.

Chapter 3

Results and Discussion

In this chapter, we present a discussion on determining the optimal PET-Disentangler configuration based on analysis of TSNE plots of healthy and disease features, visual inspection of pseudo-healthy images, and Dice values of the PET-Disentangler-n variants (Section 3.1). Next, we present and discuss the qualitative and quantitative results of PET-Disentangler on the upper torso region (Section 3.2) and on the lower torso region (Section 3.3). We also present an analysis of the observed denoising effect of PET-Disentangler in image reconstruction (Section 3.4). Finally, we present a summary of our key findings and observations (Section 3.5).

3.1 Determining the optimal PET-Disentangler configuration

Figure 3.1 shows the TSNE plots of z_h and z_d embeddings from the upper torso region across PET-Disentangler-n variants, where there is consistently a large separation between healthy and disease embedding clusters in each variant. This shows the semantic information captured between healthy and disease vectors is significantly different, and as there is no overlap between the clusters, this shows there is no leakage of disease information in the healthy vectors.

Figure 3.2 shows the TSNE plots obtained from the healthy feature vectors z_h^- and z_h^+ across PET-Disentangler-n variants. In each plot, the healthy feature embeddings form a distribution in which z_h^- and z_h^+ blend together and cannot not be separated. These plots show that the healthy features belong to the same distribution across all PET-Disentangler-n variants, ensuring the critic network is working on the task of distribution matching. As the TSNE findings are consistent between PET-Disentangler-n variant, these plots show that the healthy and disease features are disentangled and that healthy features match the same distribution regardless of the number of skip connections used.

Figure 3.3 shows the reconstructions of the disentangled healthy component from PET-Disentangler-n with n ranging from 1 to 5. For the model variations with 1 and 2 skip

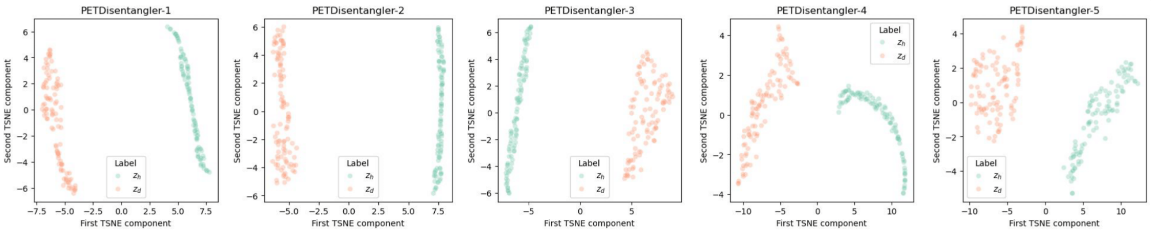

Figure 3.1: TSNE plots capturing the 2D embeddings of the healthy and disease latent vectors z_h and z_d from the upper torso region. These plots show the clear distinction between healthy and disease features by the large distance between the healthy and disease embedding clusters.

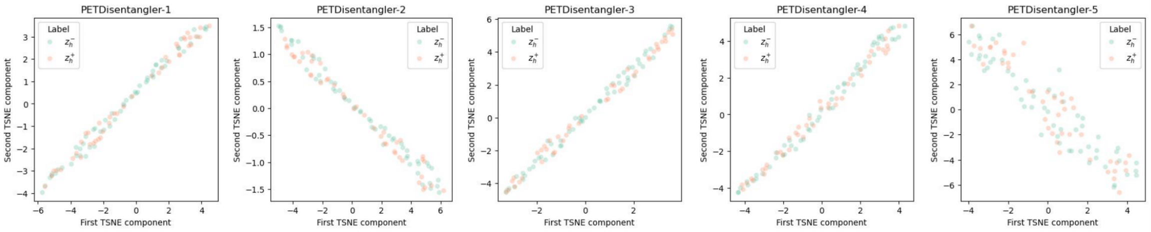

Figure 3.2: TSNE plots capturing the 2D embeddings of the healthy latent vectors z_h^- and z_h^+ across the variants of PET-Disentangler-n for the upper torso region. These plots show the healthy features from both sets are captured in the same embedding space and indistinguishable.

connections removed, the healthy reconstructions appear to have lesion information as the healthy reconstructions have similar appearance to the input images in the location of the lesions. A disentanglement of disease and healthy features seems to begin when the model has 3 skip connections removed as lesions appear to be removed or blurred out when comparing to the input image. Furthermore, PET-Disentangler with 4 or 5 skip connections removed appears to enhance the disentangled component as they do not reconstruct lesion features although at the expense of producing less sharp, blurrier images. In particular, PET-Disentangler-5 produces output images that lose a lot of visual features and structure of the anatomy while appearing quite faint opposed to the other generated images.

From the analysis of the TSNE plots in Figures 3.1 and 3.2 in addition to Figure 3.3, it appears that the healthy and disease feature vectors are disentangled although it appears the skip connections can introduce a leakage of disease features. As the image decoder should only introduce disease features via the SPADE blocks, removing skip connections can ensure the model is producing disentangled healthy and disease representations.

Given the visual inspection, it appears that disentanglement begins when 3 skip connections are removed thus the proposed architecture will be evaluated on the test set with 3 or more skip connections removed before a single configuration is chosen.

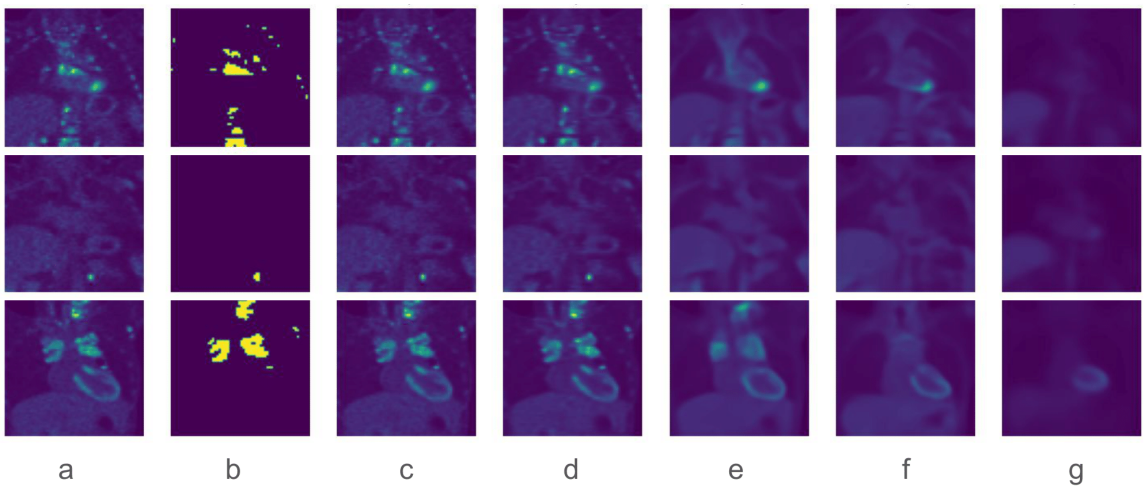

Figure 3.3: Qualitative analysis of disentanglement by viewing the reconstructed healthy component from the upper torso region: (a) Coronal slices of sample input PET. (b) The ground truth segmentation mask. (c, d, e, f, g) The healthy component reconstructed from the disentangled healthy representation from the model with the last 1, 2, 3, 4, and 5 skip connections removed, respectively.

Table 3.1: Dice metrics on the test set for the PET-Disentangler-n variants

Method	Healthy (57)	Disease (43)	Overall (100)
PET-Disentangler-3	**0.7946 ± 0.3877**	0.7084 ±0.2387	**0.7575 ± 0.3332**
PET-Disentangler-4	0.7716 ±0.3899	0.6998 ±0.2336	0.7434 ±0.3326
PET-Disentangler-5	0.7075 ±0.4126	**0.7178 ± 0.2179**	0.7119 ±0.3413

Table 3.1 summarizes the Dice metric of the PET-Disentangler with 3, 4, and 5 skip connections removed on the test set. PET-Disentangler-3 produces the highest Dice on healthy examples and overall set of examples, while PET-Disentangler-5 outperforms slightly on the disease examples. As PET-Disentangler-5 produces poor quality reconstructions, PET-Disentangler-3 demonstrates a more optimal configuration as it produces a high-quality image reconstruction at the expense of having slight disease leakage via the skip connections. Given PET-Disentangler-3 has the highest healthy and overall example Dice values while performing similarly to PET-Disentangler-5 on the disease examples, in addition to the high quality image reconstructions, PET-Disentangler-3 will be used as the optimal architecture.

Figure 3.4 shows the TSNE plots of healthy and disease feature embeddings for the lower torso region. Across all PET-Disentangler-n variants, there is a clear separation between sets of features although PET-Disentangler-3 shows a smaller distance separating the two clusters. Figure 3.5 shows the TSNE plots of healthy feature embeddings for the lower torso region. All healthy features are indistinguishable across all the PET-Disentangler-n

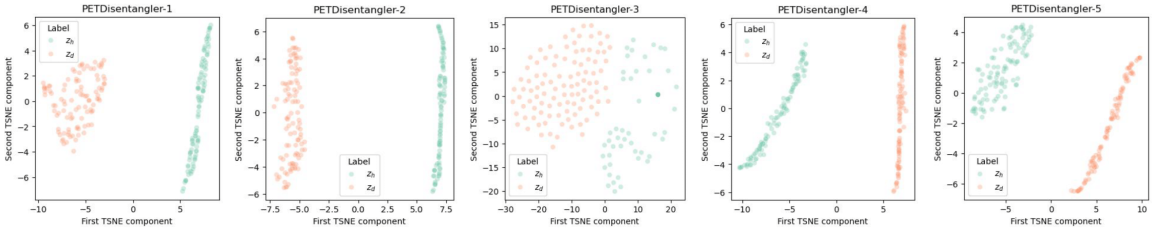

Figure 3.4: TSNE plots capturing the 2D embeddings of the healthy and disease latent vectors z_h and z_d from the lower torso region. These plots show there is consistent distinction between healthy and disease features.

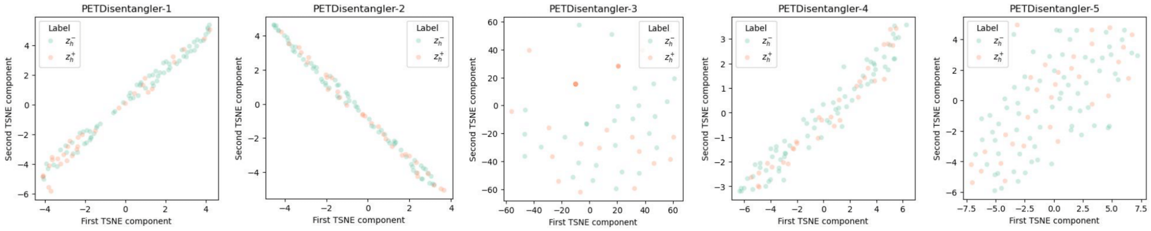

Figure 3.5: TSNE plots capturing the 2D embeddings of the healthy latent vectors z_h^- and z_h^+ across the variants of PET-Disentangler-n for the lower torso region. These plots show the healthy features from both sets are indistinguishable among all PET-Disentangler-n variants.

variants, although PET-Disentangler-3 and PET-Disentangler-5 have larger, more spread out clusters.

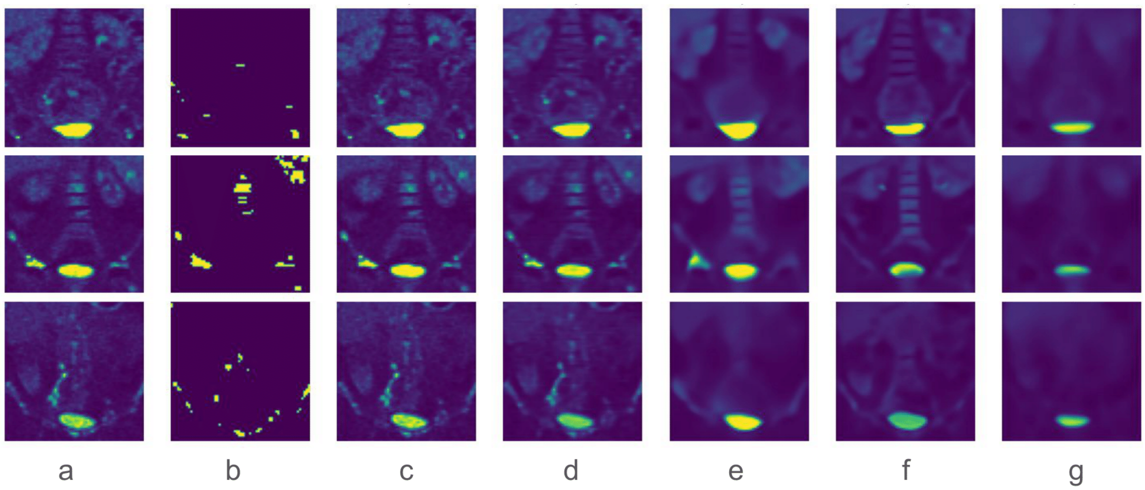

Figure 3.6: Qualitative analysis of disentanglement by viewing the reconstructed healthy component from the lower torso region: (a) Coronal slices of sample input PET. (b) The ground truth segmentation mask. (c, d, e, f, g) The healthy component reconstructed from the disentangled healthy representation from the model with the last 1, 2, 3, 4, and 5 skip connections removed, respectively.

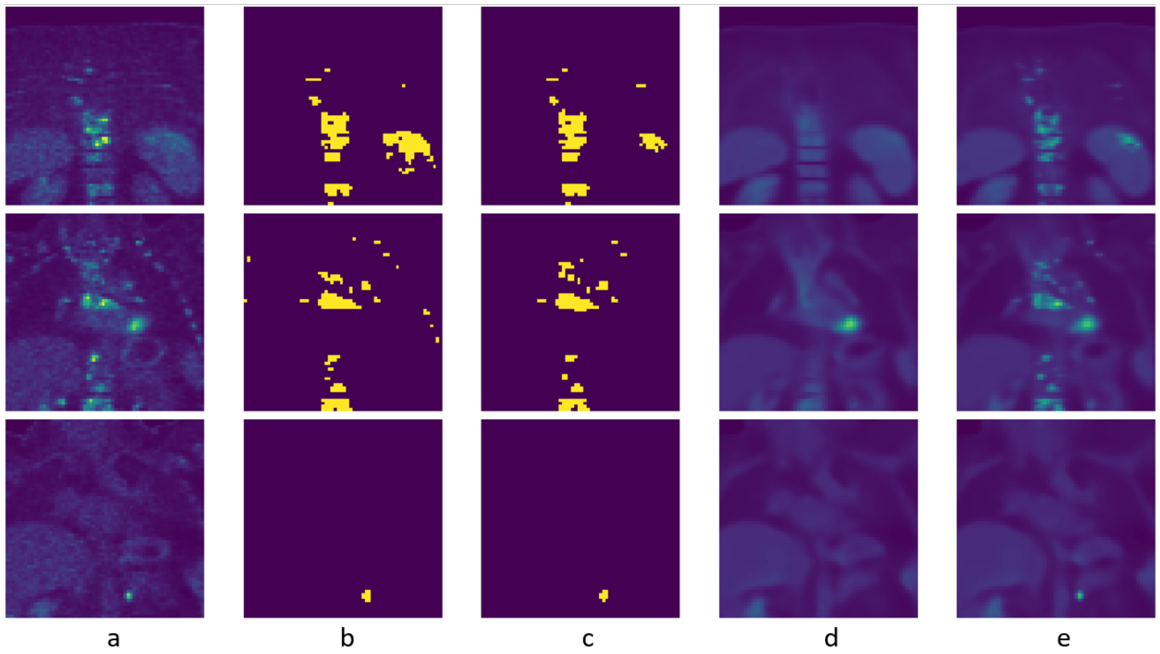

Figure 3.7: Qualitative performance of PET-Disentangler for the upper torso region: (a) Coronal slices of sample input PET volume. (b, c) The ground truth and the predicted segmentation masks, respectively. (d) The healthy component that is reconstructed from the disentangled healthy representation. (e) The reconstructed output using healthy and disease components.

Figure 3.6 shows the same patterns as Figure 3.3 where disease leakage is present in PET-Disentangler-1 and PET-Disentangler-2, in addition to PET-Disentangler-5 producing significantly blurry reconstructions. Furthermore, PET-Disentangler-4 appears to introduce structures that are not present in the input image, such as bones along the spine in the first two rows.

As PET-Disentangler-3 has disentangled healthy and disease features, healthy features that match the same distribution, and produces pseudo-healthy images with minimal issues, PET-Disentangler-3 will also be used as the optimal configuration for the lower torso region. In addition, this will ensure the method used for the lower torso is consistent with the upper region.

3.2 Upper torso experiments

Figure 3.7 shows the qualitative results from PET-Disentangler. By comparing the segmented lesions and pseudo-healthy images from columns (c) and (d), the pseudo-healthy images show an absence of lesion features that correspond to high intensity and abnormal shape. The pseudo-healthy images fill in these lesion regions with the expected "healthy" appearance. These lesion features re-appear in the full reconstructions of the images in col-

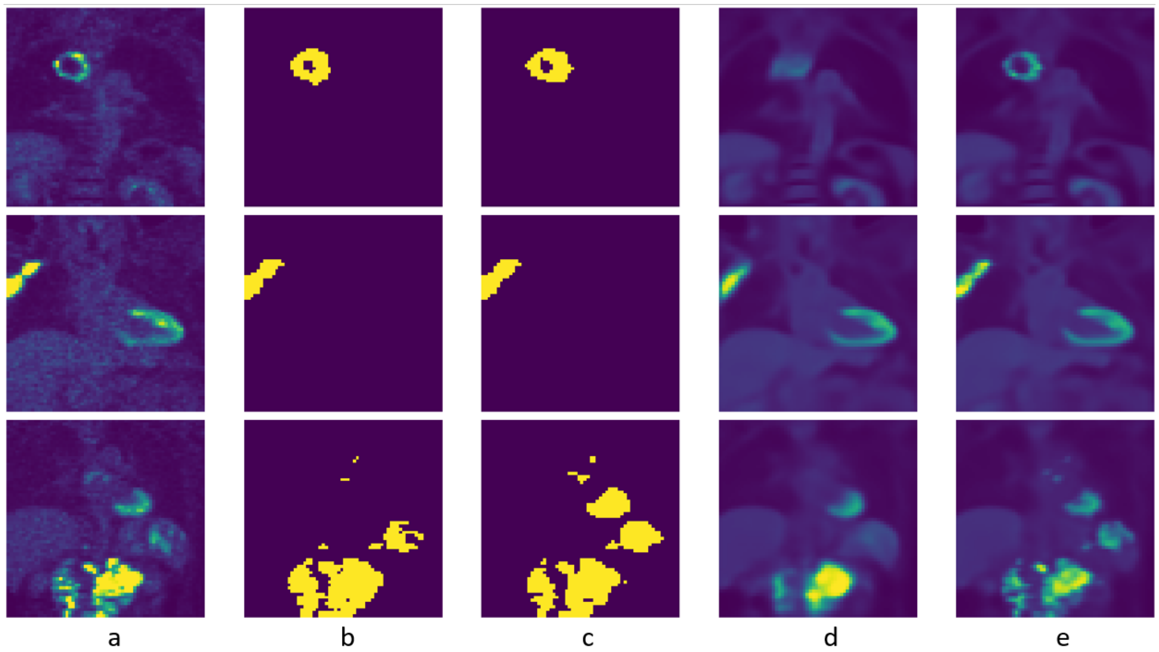

Figure 3.8: Qualitative performance of PET-Disentangler in which disentanglement of disease from healthy appears to fail: (a) Coronal slices of sample input PET volume. (b, c) The ground truth and the predicted segmentation masks, respectively. (d) The healthy component that is reconstructed from the disentangled healthy representation. (e) The reconstructed output using healthy and disease components.

umn (e) which are obtained by re-entangling the healthy and disease components via the SPADE blocks in the image decoder.

Figure 3.8 shows examples where the disentanglement of healthy and disease features are not as strong as in Figure 3.7. By examining the pseudo-healthy images, we can see that in areas that correspond to lesions in the ground truth segmentation, there are high intensity patterns in the pseudo-healthy images. These regions look similar to the corresponding appearance in the input PET images indicating that the disease features seem to be leaked into the pseudo-healthy images. From the previous discussion, this is likely due to the use of skip connections. These lesions have high intensities and likely have strong features that are present even in the later blocks of encoding, which are then used as skip connections in the image decoder. This is a tradeoff of using skip connections to stabilize training as it will produce pseudo-healthy images with disease leakage for examples with high intensity lesions.

Table 3.2 presents the average Dice coefficients obtained on the test set from the various experiments on the upper torso region. The results are shown for the healthy examples, disease examples, and the overall test set of all healthy and disease examples combined. SegOnly achieves the best performance on the healthy examples and overall test set, while the PET-Disentangler-withoutSpade achieves the best performance on the disease exam-

Table 3.2: Dice metric for lesion segmentation from the upper torso region

Method	Healthy (57)	Disease (43)	Overall (100)
SegOnly	**0.8085 ± 0.3770**	0.7084 ±0.2341	**0.7654 ± 0.3267**
SegRecon	0.6742 ±0.4523	0.6966 ±0.2411	0.6838 ±0.3749
SegReconHealthy	0.5579 ±0.4530	0.7026 ±0.2369	0.6201 ±0.3809
PET-Disentangler-withoutCritic	0.7848 ±0.3766	0.7048 ±0.2451	0.7504 ±0.3276
PET-Disentangler-withoutSpade	0.7794 ±0.3950	**0.7091 ± 0.2256**	0.7492 ±0.3333
PET-Disentangler	0.7946 ±0.3877	0.7084 ±0.2387	0.7575 ±0.3332

Table 3.3: Sensitivity of lesion segmentation from the upper torso region

Method	≤ 10 voxels (165)	10 ≤ 50 voxels (157)	50+ voxels (122)	Overall (444)
SegOnly	0.3692 ± 0.4382	0.6243± 0.3839	0.7237± 0.3457	0.5242±0.4309
SegRecon	0.3972 ±0.4607	**0.6633 ± 0.4010**	0.7610 ±0.3239	**0.5577 ± 0.4448**
SegReconHealthy	**0.4108±0.4551**	0.6084 ±0.4065	**0.7749± 0.3063**	0.5501 ±0.4370
PET-Disentangler-withoutCritic	0.3491 ±0.4363	0.6012 ±0.4056	0.6821±0.3480	0.4986 ±0.4343
PET-Disentangler-withoutSpade	0.3686 ±0.4332	0.5884 ±0.3837	0.7172±0.3404	0.5115 ±0.4247
PET-Disentangler	0.2864 ±0.4082	0.5805 ±0.3982	0.7062±0.3511	0.4676 ±0.4320

ples. However, it is worth noting that all methods have similar performance on the disease examples. The PET-Disentangler-withoutSpade performance on the disease examples highlights that learning to disentangle slightly improves on lesion segmentation compared to the baseline segmentation method, SegOnly. As this method is PET-Disentangler without SPADE blocks, this indicates that the architecture of the image decoder with SPADE blocks can be further fine-tuned and improved.

The baseline segmentation method and PET-Disentangler achieve similar performance on the disease examples. Although the baseline segmentation method slightly outperforms PET-Disentangler on the healthy examples, PET-Disentangler has comparable performance while also providing explainability using the disentangled components. SegOnly model learns disease features without any reasoning to how it determines lesions from background, whereas PET-Disentangler uses the segmentation mask in the image decoder for image reconstruction where the re-entanglement of disease and healthy features offers explainaibility in the shapes and sizes of the segmented lesions as they correspond to the features that are absent from the pseudo-healthy image.

Referring to Table 3.2 and the discussion of Dice scores on healthy examples in Section 2.3.4, each method has more false positives than the baseline segmentation model on healthy examples evident by the lower Dice values. SegRecon and SegReconHealthy in particular must have significantly more false positives than the other methods.

Table 3.3 presents the lesion sensitivities obtained on the test set, where the results are summarized for lesions with less than or equal to 10 voxels, lesions within 10 to 50 voxels, lesions greater than 50 voxels, and for all lesions combined to obtain an overall sensitivity measure. Before comparing methods, one thing to note is that the sensitivity of lesion

Table 3.4: Dice metric for lesion segmentation from the lower torso region

Method	Healthy (71)	Disease (31)	Overall (102)
SegOnly	0.0007 ±0.0026	0.1864 ±0.2474	0.0572 ±0.1598
SegRecon	0.0013 ±0.0048	0.1847 ±0.2474	0.0570 ±0.1593
SegReconHealthy	0.0008 ±0.0012	0.1791 ±0.2403	0.0550 ±0.1547
PET-Disentangler-withoutCritic	**0.7271 ±0.4115**	**0.5540 ± 0.2910**	**0.6745 ±0.3937**
PET-Disentangler-withoutSpade	0.6196 ±0.4306	0.5378 ±0.2790	0.5947 ±0.3912
PET-Disentangler	0.7174 ±0.4200	0.5153 ±0.2843	0.6560 ±0.3937

segmentation increases as lesion sizes increase. This can be seen as the lesion sensitivities for lesions with less than 10 voxels are significantly lower than those corresponding to lesions with more than 50 voxels.

SegReconHealthy achieves the highest sensitivity for lesions less than 10 voxels and lesions greater than 50 voxels, whereas SegRecon achieves the highest sensitivity for lesions between 10 and 50 voxels and for the overall set of lesions. These results are in direct contrast to the Dice values in Table 3.2 where SegRecon and SegReconHealthy have the lowest Dice among healthy, disease, and overall examples. Referring to Equation 2.15, a higher sensitivity would occur for a given lesion if there are more true positives and less false negatives. These results show that the SegRecon and SegReconHealthy methods have more true positive and false negative predictions, whereas the Dice performance is lower due to more false positive predictions. In comparison to SegOnly, the enhanced sensitivity from SegRecon and SegReconHealthy indicates that learning additional features, compared to only disease features, enhances lesion sensitivities.

PET-Disentangler has the lowest sensitivity for lesions less than 10 voxels, lesions between 10 and 50 voxels, and the overall set of lesions, where they values are significantly lower than the highest per category. As PET-Disentangler has Dice results on par with SegOnly, this highlights that PET-Disentangler has more false negatives while less false positives compared to the other methods. Although reducing false positives is valuable, decreasing false negatives is of equal or higher importance.

3.3 Lower torso experiments

Table 3.4 presents the average Dice coefficients obtained on the test set from the various experiments on the lower torso region. SegOnly, SegRecon, and SegReconHealthy have significantly lower values compared to PET-Disentangler variants for healthy, disease, and overall examples. In addition, SegOnly, SegRecon, and SegReconHealthy each have values close to 0 for healthy examples. Given the healthy example Dice values and the previous discussion, these methods must have significant false positives compared to the PET-Disentangler variants. The disease examples show that PET-Disentangler is still significantly

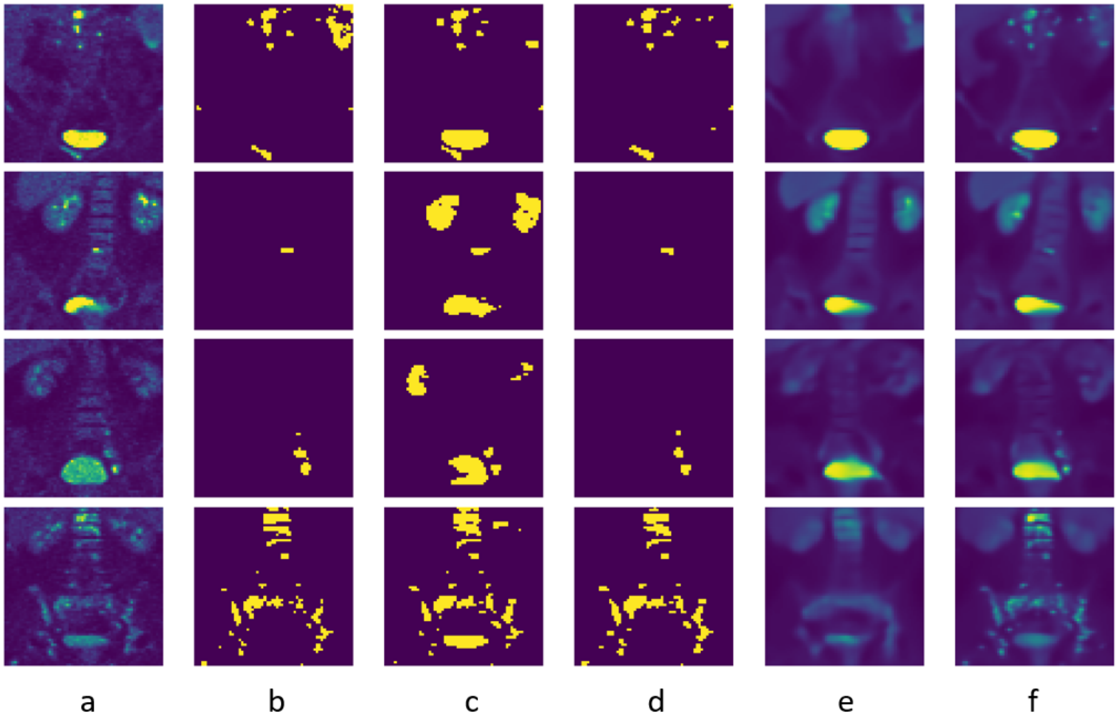

Figure 3.9: PET-Disentangler compared to SegOnly for the lower torso region, highlighting that SegOnly produces false positives for healthy uptake whereas PET-Disentangler does not: (a) Coronal slices of sample input PET volume. (b, c, d): The ground truth, baseline UNet predicted segmentation masks, and PET-Disentangler segmentation masks, respectively. (e) The healthy component from PET-Disentangler using the healthy components. (f) The reconstructed output from PET-Disentangler.

Table 3.5: Sensitivity of lesion segmentation from the lower torso region

Method	$\leq$ 10 voxels (165)	10 $\leq$ 50 voxels (134)	50+ voxels (74)	Overall (373)
SegOnly	0.4341± 0.4527	0.5762± 0.3943	0.7155± 0.3260	0.5410± 0.4233
SegRecon	0.4963 ±0.4599	**0.6052 ±0.3656**	0.6831 ±0.3561	**0.5725± 0.4149**
SegReconHealthy	0.4024 ±0.4573	0.5580 ±0.3777	0.6331 ±0.3465	0.5040 ±0.4201
PET-Disentangler-withoutCritic	0.4501±0.4461	0.5914 ±0.3863	0.7155 ±0.0.3101	0.5535 ±0.4137
PET-Disentangler-withoutSpade	**0.5061 ± 0.4614**	0.5956 ±0.3646	0.6868 ±0.3264	0.5741 ±0.4097
PET-Disentangler	0.3883±0.4598	0.5727 ±0.4178	**0.7175 ± 0.3029**	0.5199 ±0.4368

outperforming the competing methods and likely due to the same false positive source as on healthy examples. Furthermore, PET-Disentangler-withoutCritic appears to be the best performing model, in addition to PET-Disentangler-withoutSpade performing better than PET-Disentangler on disease examples. As both PET-Disentangler variants improve upon PET-Disentangler on disease examples, this shows that PET-Disentangler can be further fine-tuned to leverage the benefit of both the SPADE blocks and the critic network to become the best performing configuration.

When looking at examples from the lower torso region, Figure 3.9 shows qualitative results from SegOnly in column (c) and the PET-Disentangler segmentation in column (d) as well as the PET-Disentangler pseudo-healthy image and reconstruction in columns (e) and (f), respectively. SegOnly results show that the bladder is consistently segmented incorrectly as lesions. Additionally, the second and third rows indicate that high intensity activity that corresponds to healthy kidney uptake is also incorrectly segmented as lesions. In contrast, PET-Disentangler has learned the healthy anatomy component of the images and can identify these high uptake regions as healthy activity and focus on the remaining activity to delineate the lesions.

When reviewing the lesion sensitivities for the lower region in Table 3.5, PET-Disentangler-withoutSpade have the highest sensitivity for lesions with less than 10 voxels, SegRecon the highest sensitivity for lesions between 10 and 50 voxels, and the overall set of lesions, whereas PET-Disentangler has the highest sensitivity for lesions greater than 50 voxels. Similar to as seen in the upper torso results, SegOnly does not achieve the highest sensitivity in any category indicating that learning more than disease features enhances lesion sensitivity.

The lower torso results indicate that PET-Disentangler can accurately learn high intensity uptake although there are instances where PET-Disentangler can succeed or fail nearby these areas. For example, Figure 3.10 shows two instances of lesions nearby high intensity pixels of the bladder. The first row highlights a failure of the model to detect abnormal high intensity uptake where the bladder normally is located and PET-Disentangler wrongly assumes this uptake corresponds to bladder activity. To contrast, the second row shows that tumors that are outside of the bladder can be detected and delineated even with their close proximity to the bladder. These examples show that PET-Disentangler can delineate

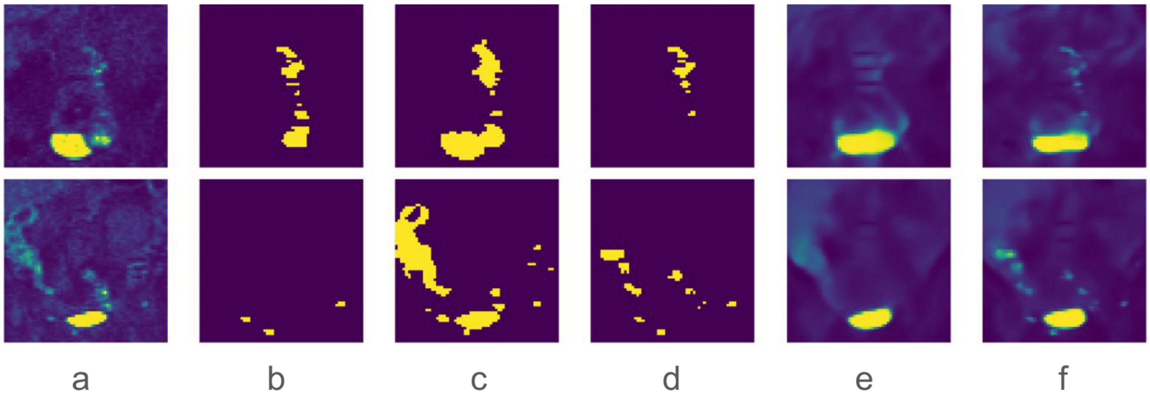

Figure 3.10: Analysis of PET-Disentangler performance nearby high intensity uptake of normal activity compared to Segonly for the lower torso region: (a) coronal slices of input PET, (b) ground truth segmentation, (c) SegOnly segmentation prediction, (d, e, f) PET-Disentangler segmentation prediction, pseudo-healthy image, and reconstruction, respectively.

lesions nearby healthy anatomy uptake although may not recognize abnormal uptake in regions where its expected that healthy uptake occurs.

3.4 The importance of learning the healthy anatomy component

As discussed in the previous sections, PET-Disentangler is able to reduce false positive segmentations due to healthy anatomy uptake by learning the healthy anatomy component of PET images. This finding also raises the question of whether disentanglement is necessary to model and subsequently omit healthy organ uptake. For example, we have healthy organ segmentations obtained from TotalSegmentator that could be leveraged to mask the false positive uptake patterns in the produced segmentations. This would eliminate the need for disentanglement as we would have already identified healthy uptake regions from these organ segmentations. Secondly, as we obtained aligned regions between upper and lower torso examples, we can model the healthy anatomy uptake by taking the average healthy PET image. This would alleviate the need for disentanglement as we can easily identify healthy uptake and furthermore lesions by subtracting this average healthy component from examples with disease.

To address the discussion on leveraging TotalSegmentators' organ segmentations, we first look at Figure 3.11 that shows examples with uptake from the heart, bladder, and kidneys outlined in blue and lesions outlined in red contours. This figure highlights that the bladder segmentations obtained from using the CT scan and TotalSegmentator do not fully correspond to the contours of the PET image uptake. This shows we cannot always

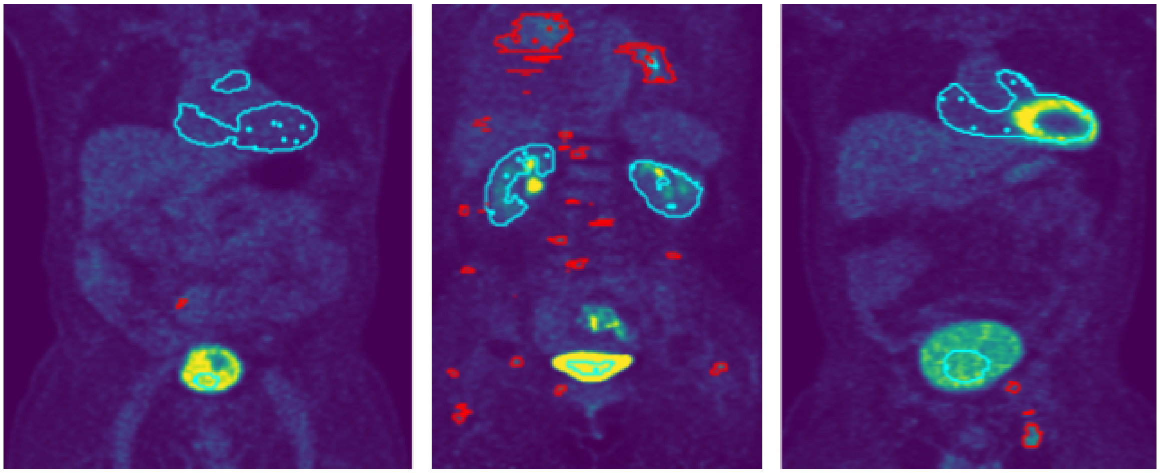

Figure 3.11: Bladder contours in blue highlight that the segmentations obtained from CT scan and TotalSegmentator may not fully align with the corresponding uptake patterns in PET images.

assume the organ segmentations obtained from the CT scan will have a 1:1 correspondence for PET uptake segmentation.

Furthermore, Figure 3.12 shows in (a) and (b) that lesions overlap the bladder segmentation contour, and that lesions overlap the heart uptake in (c). Figure 3.13 shows examples of lesion uptake overlapping healthy kidney uptake in (a), (b), and (c). Both Figures 3.12 and 3.13 show that we cannot use organ segmentations to omit healthy uptake regions as there are examples where lesions are in these healthy organs and have overlapping segmentations.

Secondly, as we obtained aligned regions between upper and lower torso examples, we can model the healthy anatomy uptake by taking the average healthy PET image. This would alleviate the need for disentanglement as we can easily identify healthy uptake and furthermore lesions by subtracting this average healthy component from examples with disease.

To discuss whether an average healthy image can be computed and then used to remove healthy uptake patterns, we first look at the histogram of SUV values within the heart, kidneys, and bladder in Figure 3.14. This histogram shows the number of occurrences for SUV intensities of each pixel from these 3 organs, where heart pixels correspond to the blue bars, right kidney pixels to the orange bars, left kidney pixels to the green bars, and bladder pixels to the red bars. The heart and kidney histograms show a spread of the pixel intensities between 0-6 SUV indicating heterogeneity, whereas the histogram for the bladder intensities show significant heterogeneity from the wide SUV distribution ranging from SUV 0 to 15.

Figure 3.15 shows examples of PET images with disease (a), alongside their corresponding ground truth segmentation (b), the corresponding average PET image (c), and the residual image between (c) and (a) thresholded by 0.3. Figure 3.15 (c) shows that the aver-

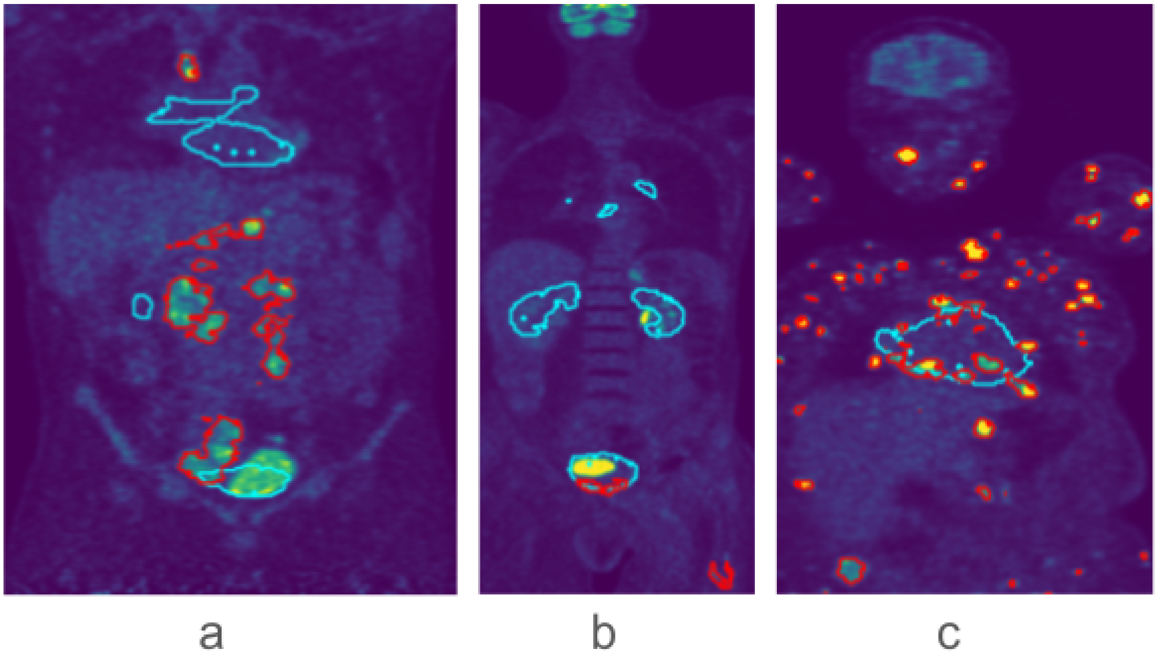

Figure 3.12: Red contours outline the lesion uptake that is overlapping the organ segmentations outlined in blue in the bladder (a), (b), and heart (c).

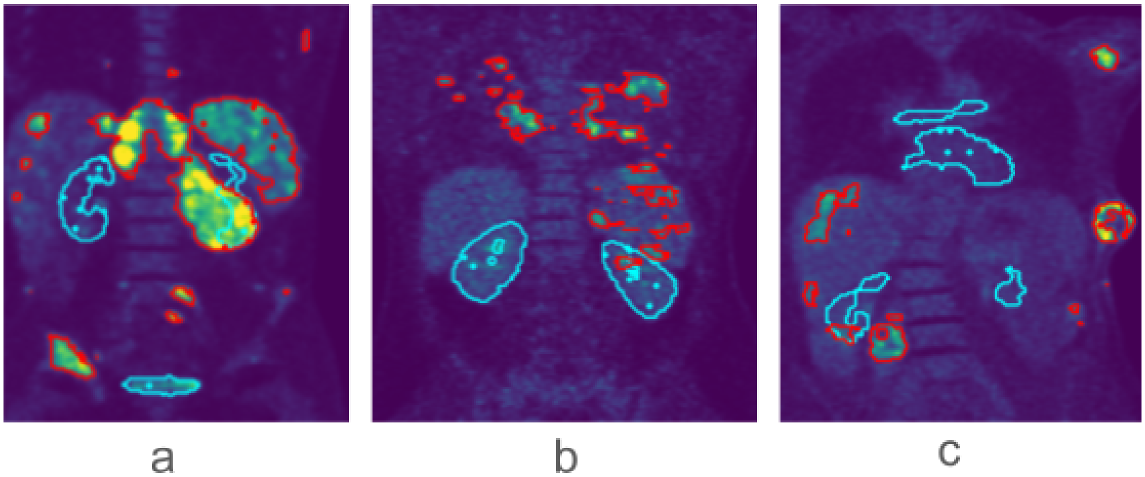

Figure 3.13: Red contours outline the lesion uptake overlapping the kidney segmentation contours outlined in blue in (a), (b), and (c).

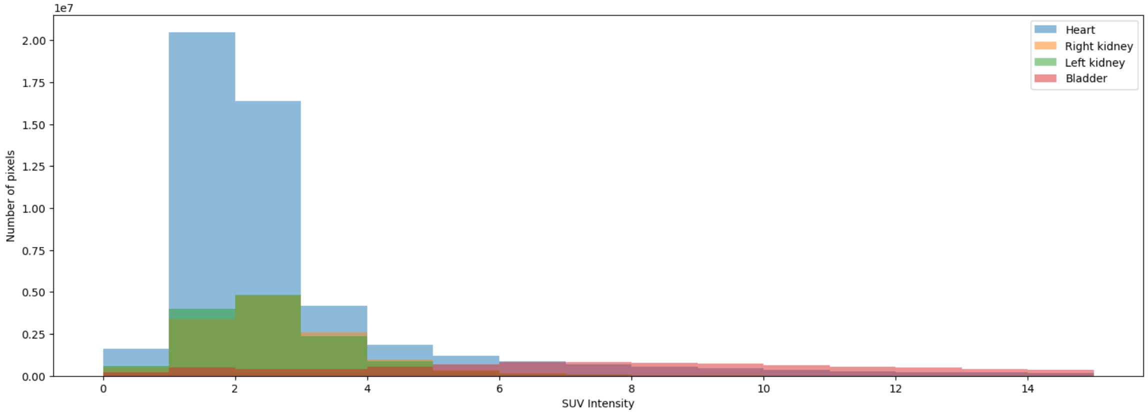

Figure 3.14: Histogram of pixel SUV intensities from heart, kidney, and bladder pixels.

age PET image signal is blurry and that it does not have the same uptake contours as seen in (a). The residual images also show that the differences between average healthy image and PET example with disease do not correspond to lesions rather they indicate healthy uptake patterns not seen in the average PET image. The first row also shows the lesions overlap with the average PET image such that no lesions are identified in the residual. The results in Figure 3.15, along with the heterogeneity observed in Figure 3.14, indicate that calculating a mean healthy PET image cannot model healthy uptake patterns and delineate lesions as can be done by PET-Disentangler.

3.5 Evaluating denoising effect on reconstruction

From Figures 3.7 and 3.9, the reconstructed images from PET-Disentangler show that the images appear to be denoised. To investigate this denoising effect, patches of $10\times10\times10$ were sampled from the liver between examples of the model input and the reconstructed output from the healthy examples. These patches were sampled from the liver due to its consistent appearance across patients, both in terms of size and intensity distribution. Additionally, the liver's size allows for the selection of relatively large foreground 3D patches.

These samples of patches were compared using the signal-to-noise ratio (SNR), described as:

$$SNR = \frac{\mu_{signal}}{\sigma_{noise}}, \tag{3.1}$$

where μ_{signal} and σ_{noise} represent the mean of the signal and standard deviation of the noise, respectively.

These patches before being passed to the model had an average SNR of 4.98 ± 1.15, whereas the average SNR of the reconstructed patches was 15.46 ± 10.29. This significant SNR increase shows that PET-Disentangler can significantly denoise and smooth the signal from the noisy PET input. This denoising effect is likely due to the modification of skip

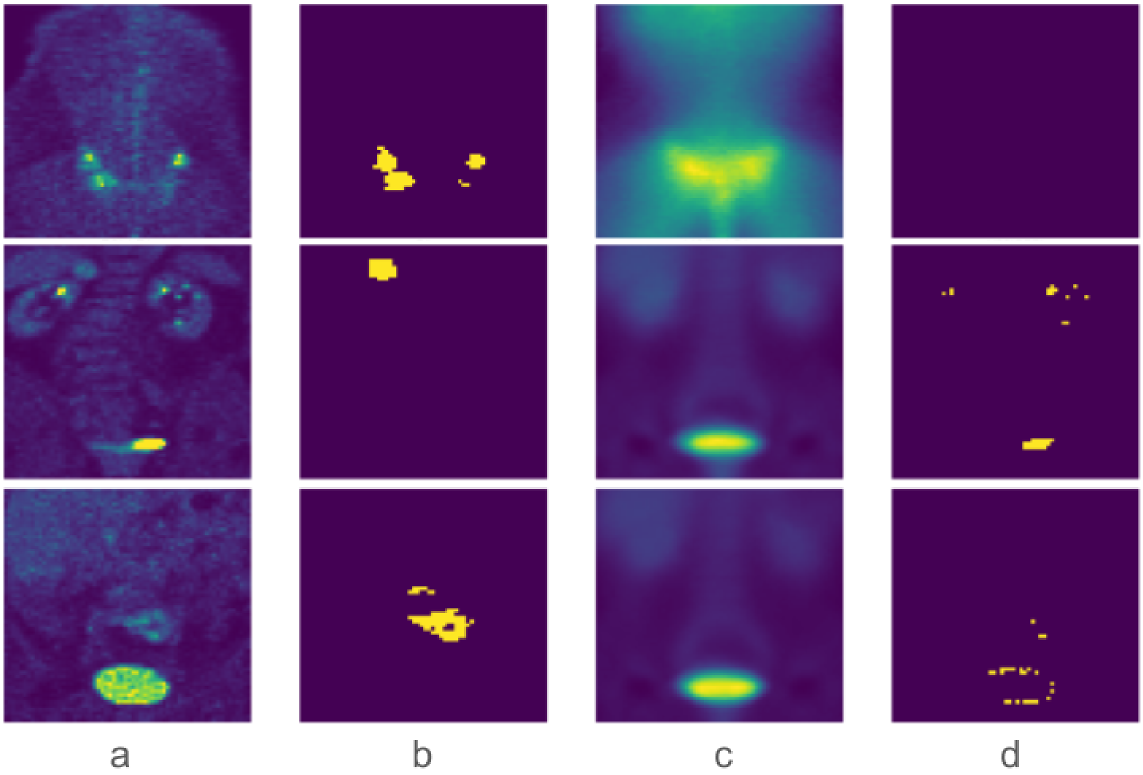

Figure 3.15: Examples of (a) coronal PET images, (b) corresponding ground truth segmentation, (c) corresponding average PET image, (d) residual image thresholded by 0.3.

connections between the encoder and image decoder. When all skip connections are used, the image decoder is able to restore high frequency information more easily such as noise. This is due to the skip connections being able to provide and reintroduce features that would otherwise be lost in the depths of the encoding and subsequent decoding blocks.

Ulyanov et al. [59] further discuss that although it is in theory possible for a neural network to learn parameters that fit both the noise and signal during image reconstruction, it is much more likely for a network to capture a smoother signal and filter out noise. Due to the high frequency, unstructured nature of noise, the network would need to make many fine adjustments to capture these fluctuations whereas the network's parameters easily adjust to the smooth signal within the image. This suggests that it would take significantly more parameter optimization to reach a point in which PET-Disentangler would lose its denoising effect and begin to recover high frequency noise features.

To evaluate whether the denoising effect improves the contast-to-noise ratio (CNR) of lesions after reconstruction, we report the CNR values of lesions for upper and lower torso disease examples in the test sets using the following formula:

$$CNR = \frac{\mu_{signal} - \mu_{background}}{\sigma_{background}}, \tag{3.2}$$

where μ_{signal} is the mean lesion intensity, and $\mu_{background}$ and $\sigma_{background}$ are the mean and standard deviation of the background area surrounding the lesion. The lesion CNRs between original PET images and PET-Disentangler reconstructions are 4.50 ± 2.81 and 2.73 ± 2.84 for the upper torso examples and 4.60 ± 2.79 and 2.51 ± 2.53 for the lower torso examples, respectively. Unfortunately in both upper and lower torso examples, the CNR of lesions decreases after reconstruction indicating that although PET-Disentangler appears to have a denoising effect, the resulting reconstructions are not enhancing the contrast of lesions to the surrounding areas.

3.6 Summary

Given the results and discussion presented in the previous sections, we identify the following key observations:

- Healthy and disease features are disentangled as seen in the TSNE plots in Figures 3.1 and 3.2, where it is also observed that healthy features match the same distribution with no disease leakage. In contrast, skip connections can introduce disease feature leakage seen in the analysis of pseudo-healthy images.
- PET-Disentangler performs similarly to the segmentation baseline on Dice metrics while providing additional explainability in the shapes and sizes of lesions segmented as they correspond to missing features in the pseudo-healthy images.

- PET-Disentangler significantly reduces false positives in segmentation of healthy uptake patterns compared to non-disentangling methods by modelling the healthy anatomy as discussed in Section 3.3. This is seen visually in Figure 3.9 and reflected quantitatively in the Dice values in Table 3.4.

- Network architectures that learn more than disease features have enhanced lesion sensitivity compared to a baseline segmentation only method as seen in Tables 3.3 and 3.5.

- PET-Disentangler has a denoising effect such that when comparing the SNR in liver patches between original and reconstructed images, the SNR improves from 4.98 ± 1.15 to 15.46 ± 10.29.

Chapter 4

Conclusion and Future Work

Development of automatic lesion segmentation tools to facilitate the annotation of clinical PET images remains an important task. Many automatic segmentation tools focus on optimizing the learning of disease features while concurrently there are increasing advancements in novel computer vision and deep learning techniques that could potentially provide utility in lesion segmentation.

This thesis investigated the use of image disentanglement for lesion segmentation by decomposing a PET image into sources of variation, specifically healthy and disease features. We proposed PET-Disentangler, a modified UNet architecture that learns to disentangle PET images into healthy and disease features in the latent space, produce segmentation predictions using the disease features, and re-entangles healthy and disease features to produce reconstructed images. A critic network is used to ensure the healthy features match the same distribution to prevent leakage of disease features. When healthy features alone are use for image reconstruction, a pseudo-healthy image is generated such that all lesions are removed and replaced with what PET-Disentangler expects the healthy anatomy to look like in those locations. Through experiments on the upper torso region, we showed learning additional features to disease features enhances lesion segmentation sensitivity. We also observed PET-Disentangler performing similarly to a segmentation only baseline, in which PET-Disentangler has the additional benefit of explainability as the disease features produced correspond to the missing features in the pseudo-healthy image estimate. Through experiments on the lower torso region, we also noted that PET-Disentangler greatly reduces the false positive segmentation of healthy uptake patterns by modelling the healthy anatomy component. Overall, we also observed that detecting and segmenting small lesions remains a challenging task as lesion sensitivity is significantly lower for lesions with 10 or fewer voxels in volume.

Throughout the development and evaluation of PET-Disentangler, it became apparent that working in 3D and requiring anatomically-aligned inputs were the major limitations of this work. As PET-Disentangler used 3D inputs and produced 3D outputs, in addition to utilizing multiple 3D model components, this produced a large computational cost on the

GPU. This computational cost limited the batch size that could be used during training in addition to the size of the input and output volumes. Furthermore, requiring anatomically aligned PET volumes is a limitation of this approach as the PET inputs much be aligned such that healthy anatomy can be accurately modeled. This limits the direct use of PET-Disentangler on full-body PET scans as more pre-processing is required to select regions that capture the whole-body scans and solutions must be developed to accommodate the differences between how much of the body is captured between scans (i.e., eyes-to-thighs, head-to-feet, etc.)

To separate healthy from disease in PET images, we require the disentanglement framework to accurately model the healthy component as it is not feasible to separate the two in other ways such as by classifying pixel-by-pixel healthy vs disease regions as this would require a reference for all healthy uptake in PET. Tools like TotalSegmentator can be leveraged to segment many anatomical landmarks in CT/MRI that can be subsequently used for PET although these tools do not account for all healthy uptake such as that from brown fat or in the urinary tract, thus a robust model that can model all these healthy variations is required.

Given the proposed method and conclusions, future work could explore the following directions:

- **Network Architecture:** Exploring the choice of architecture to use for disentanglement. In this work, a 3D UNet was extended to perform the disentanglement although as seen from the analysis of skip connections, they are required in UNet to produce high quality image reconstructions, although can introduce disease feature leakage. In contrast, other architectures without skip connections could potentially alleviate the disease leakage problem while producing high quality output. In addition, methods have been proposed to improve the UNet architecture with transformers [13] and further investigation of the use of transformers and additional state-of-the-art frameworks may alleviate the limitations seen in this work.

- **Multimodal analysis:** With the availability of hybrid PET/CT and PET/MRI scanners, PET scans are often accompanied with corresponding scans of another modality. PET scans have a low signal to noise ratio and modalities such as CT and MRI can provide additional anatomical information that is not clearly captured in PET. Future work can investigate the use of a multimodal framework in which an additional modality in combination with PET is used to enhance the disease and healthy features learned.

- **Data generation:** As public datasets for PET are limited, future work can investigate using the proposed method for data generation as a data augmentation technique. As the model produces pseudo-healthy images, these images can be used as additional data samples. Additionally, the method could produce unique disease examples by sampling various healthy features and segmentation predictions, or even artificial segmentation masks, to produce unique disease images.

- **Denoising:** Another source of variation in PET images that was not explicitly modelled via the healthy and disease features is the noise component. As the proposed method has a denoising effect, this work can further be extended to disentangle and model the noise features within PET which may subsequently enhance the disease features modelled and improve lesion segmentation. Izadi et al. [32] introduced a disentangled approach for image denoising in which smooth signal and noise features are separated in the latent space. Future work can extend upon this direction to include the disentanglement of noise in the PET-Disentangler framework and can be further modified to account for the noise properties unique to PET images.

- **Error attribution:** It can be difficult to explain in which cases segmentation fails and sources of these errors. To better understand the sources of error for segmentation, training a model to predict error can be used to highlight regions of uncertainty in addition to exploring causal discovery to address this issue.

- **Shape priors:** As discussed by Nosrati and Hamarneh [48], introducing prior knowledge into segmentation algorithms, such as appearance, shape, topology, distance between regions, etc., enhances the segmentations produced. Mirikharaji and Hamarneh [46] also showed the benefit of integrating shape priors into a deep learning segmentation framework. Future work can investigate the use of prior knowledge, particularly shape priors, to enhance lesion segmentation in PET.

- **Management prediction:** As introduced by Abhishek et al. [2], segmentation of skin lesions is commonly performed to identify a diagnosis, and other works may produce a diagnosis based directly off the input image, although there has not been a focus on predicting the management of disease itself. Future work can investigate the clinical applications of PET lesion segmentation and whether there are opportunities to directly predict the management of disease and additional clinical tasks without first obtaining lesion segmentations.

Bibliography

[1] M. Abdoli, R. Dierckx, and H. Zaidi, “Contourlet-based active contour model for PET image segmentation,” *Medical Physics*, vol. 40, no. 8, p. 082507, 2013.

[2] K. Abhishek, J. Kawahara, and G. Hamarneh, “Predicting the clinical management of skin lesions using deep learning,” *Scientific Reports*, vol. 11, no. 1, p. 7769, 2021.

[3] S. Afshari, A. BenTaieb, and G. Hamarneh, “Automatic localization of normal active organs in 3D PET scans,” *Computerized Medical Imaging and Graphics*, vol. 70, pp. 111–118, 2018.

[4] S. Afshari, A. BenTaieb, Z. Mirikharaji, and G. Hamarneh, “Weakly supervised fully convolutional network for PET lesion segmentation,” in *Medical Imaging 2019: Image Processing*, vol. 10949. SPIE, 2019, pp. 394–400.

[5] P. Ahmadvand, N. Duggan, F. Bénard, and G. Hamarneh, “Tumor Lesion Segmentation from 3D PET Using a Machine Learning Driven Active Surface,” in *Machine Learning in Medical Imaging: 7th International Workshop, MLMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Proceedings 7.* Springer, 2016, pp. 271–278.

[6] V. Andrearczyk, V. Oreiller, M. Vallières, J. Castelli, H. Elhalawani, M. Jreige, S. Boughdad, J. O. Prior, and A. Depeursinge, “Automatic Segmentation of Head and Neck Tumors and Nodal Metastases in PET-CT scans,” in *Medical Imaging with Deep Learning.* PMLR, 2020, pp. 33–43.

[7] M. Aristophanous, B. C. Penney, M. K. Martel, and C. A. Pelizzari, “A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography,” *Medical Physics*, vol. 34, no. 11, pp. 4223–4235, 2007.

[8] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein Generative Adversarial Networks,” in *International Conference on Machine Learning.* PMLR, 2017, pp. 214–223.

[9] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. J. Mollura, “Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images,” *Medical Image Analysis*, vol. 17, no. 8, pp. 929–945, 2013.

[10] S. Belhassen and H. Zaidi, “A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET,” *Medical Physics*, vol. 37, no. 3, pp. 1309–1324, 2010.

[11] P. Blanc-Durand, S. Jégou, S. Kanoun, A. Berriolo-Riedinger, C. Bodet-Milin, F. Kraeber-Bodéré, T. Carlier, S. Le Gouill, R.-O. Casasnovas, M. Meignan *et al.*, "Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network." *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, pp. 1362–1370, 2021.

[12] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, A. Myronenko, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M. Z. Darestani, C. Budd, M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B. S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P. F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L. A. D. Cooper, H. R. Roth, D. Xu, D. Bericat, R. Floca, S. K. Zhou, H. Shuaib, K. Farahani, K. H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, and A. Feng, "MONAI: An open-source framework for deep learning in healthcare," 2022. [Online]. Available: https://arxiv.org/abs/2211.02701

[13] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, M. P. Lungren, S. Zhang, L. Xing, L. Lu, A. Yuille, and Y. Zhou, "TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, p. 103280, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841524002056

[14] A.-S. Dewalle-Vignion, N. Betrouni, R. Lopes, D. Huglo, S. Stute, and M. Vermandel, "A New Method for Volume Segmentation of PET Images, Based on Possibility Theory," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 409–423, 2010.

[15] V. Duclos, A. Iep, L. Gomez, L. Goldfarb, and F. L. Besson, "PET Molecular Imaging: A Holistic Review of Current Practice and Emerging Perspectives for Diagnosis, Therapeutic Evaluation and Prognosis in Clinical Oncology," *International Journal of Molecular Sciences*, vol. 22, no. 8, p. 4159, 2021.

[16] Y. E. Erdi, O. Mawlawi, S. M. Larson, M. Imbriaco, H. Yeung, R. Finn, and J. L. Humm, "Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 80, no. S12, pp. 2505–2509, 1997.

[17] 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.

[18] M. Früh, M. Fischer, A. Schilling, S. Gatidis, and T. Hepp, "Weakly supervised segmentation of tumor lesions in PET-CT hybrid imaging," *Journal of Medical Imaging*, vol. 8, no. 5, pp. 054 003–054 003, 2021.

[19] S. Gatidis, T. Hepp, M. Früh, C. La Fougère, K. Nikolaou, C. Pfannenberg, B. Schölkopf, T. Küstner, C. Cyran, and D. Rubin, "A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions," *Scientific Data*, vol. 9, no. 1, p. 601, 2022.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[21] E. Grossiord, H. Talbot, N. Passat, M. Meignan, and L. Najman, "Automated 3D lymphoma lesion segmentation from PET/CT characteristics," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 174–178.

[22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved Training of Wasserstein GANs," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[23] D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, "Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-Based Co-Segmentation Method," in *Information Processing in Medical Imaging: 22nd International Conference, IPMI 2011, Kloster Irsee, Germany, July 3-8, 2011. Proceedings 22*. Springer, 2011, pp. 245–256.

[24] L. Han, Y. Lyu, C. Peng, and S. K. Zhou, "GAN-based disentanglement learning for chest X-ray rib suppression," *Medical Image Analysis*, vol. 77, p. 102369, 2022.

[25] H. Hanzouli-Ben Salah, J. Lapuyade-Lahorgue, J. Bert, D. Benoit, P. Lambin, A. Van Baardwijk, E. Monfrini, W. Pieczynski, D. Visvikis, and M. Hatt, "A framework based on hidden Markov trees for multimodal PET/CT image co-segmentation," *Medical Physics*, vol. 44, no. 11, pp. 5835–5848, 2017.

[26] M. Hatt, F. Lamare, N. Boussion, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest *et al.*, "Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET," *Physics in Medicine & Biology*, vol. 52, no. 12, p. 3467, 2007.

[27] 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.

[28] M. Hatt, C. C. Le Rest, P. Descourt, A. Dekker, D. De Ruysscher, M. Oellers, P. Lambin, O. Pradier, and D. Visvikis, "Accurate Automatic Delineation of Heterogeneous Functional Volumes in Positron Emission Tomography for Oncology Applications," *International Journal of Radiation Oncology* Biology* Physics*, vol. 77, no. 1, pp. 301–308, 2010.

[29] F. Hofheinz, J. Langner, J. Petr, B. Beuthien-Baumann, J. Steinbach, J. Kotzerke, and J. van den Hoff, "An automatic method for accurate volume delineation of heterogeneous tumors in PET," *Medical Physics*, vol. 40, no. 8, p. 082503, 2013.

[30] H. Hu, L. Shen, T. Zhou, P. Decazes, P. Vera, and S. Ruan, "Lymphoma Segmentation in PET Images Based on Multi-view and Conv3D Fusion Strategy," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1197–1200.

[31] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[32] S. Izadi, Z. Mirikharaji, M. Zhao, and G. Hamarneh, "WhiteNNer-Blind Image Denoising via Noise Whiteness Priors," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 0–0.

[33] S. Jemaa, J. Fredrickson, R. A. Carano, T. Nielsen, A. de Crespigny, and T. Bengtsson, "Tumor Segmentation and Feature Extraction from Whole-Body FDG-PET/CT Using Cascaded 2D and 3D Convolutional Neural Networks," *Journal of Digital Imaging*, vol. 33, pp. 888–894, 2020.

[34] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET Volumes by Iterative Image Thresholding," *Journal of Nuclear Medicine*, vol. 48, no. 1, pp. 108–114, 2007.

[35] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random Walk and Graph Cut for Co-Segmentation of Lung Tumor on PET-CT Images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5854–5867, 2015.

[36] J. S. Karp, V. Viswanath, M. J. Geagan, G. Muehllehner, A. R. Pantel, M. J. Parma, A. E. Perkins, J. P. Schmall, M. E. Werner, and M. E. Daube-Witherspoon, "PennPET Explorer: Design and Preliminary Performance of a Whole-Body Imager," *Journal of Nuclear Medicine*, vol. 61, no. 1, pp. 136–143, 2020.

[37] K. Kobayashi, R. Hataya, Y. Kurose, M. Miyake, M. Takahashi, A. Nakagawa, T. Harada, and R. Hamamoto, "Decomposing normal and abnormal features of medical images for content-based image retrieval of glioma imaging," *Medical Image Analysis*, vol. 74, p. 102227, 2021.

[38] J. Lapuyade-Lahorgue, D. Visvikis, O. Pradier, C. Cheze Le Rest, and M. Hatt, "SPEQTACLE: An automated generalized fuzzy C-means algorithm for tumor delineation in PET," *Medical Physics*, vol. 42, no. 10, pp. 5720–5734, 2015.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[40] H. Li, H. Jiang, S. Li, M. Wang, Z. Wang, G. Lu, J. Guo, and Y. Wang, "DenseX-Net: An End-to-End Model for Lymphoma Segmentation in Whole-Body PET/CT Images," *IEEE Access*, vol. 8, pp. 8004–8018, 2019.

[41] L. Li, X. Zhao, W. Lu, and S. Tan, "Deep learning for variational multimodality tumor segmentation in PET/CT," *Neurocomputing*, vol. 392, pp. 277–295, 2020.

[42] P. Liu, M. Zhang, X. Gao, B. Li, and G. Zheng, "Joint Lymphoma Lesion Segmentation and Prognosis Prediction From Baseline FDG-PET Images via Multitask Convolutional Neural Networks," *IEEE Access*, vol. 10, pp. 81 612–81 623, 2022.

[43] X. Liu, P. Sanchez, S. Thermos, A. Q. O'Neil, and S. A. Tsaftaris, "Learning disentangled representations in the imaging domain," *Medical Image Analysis*, vol. 80, p. 102516, 2022.

[44] S. R. Meikle, V. Sossi, E. Roncali, S. R. Cherry, R. Banati, D. Mankoff, T. Jones, M. James, J. Sutcliffe, J. Ouyang *et al.*, "Quantitative PET in the 2020s: a roadmap," *Physics in Medicine & Biology*, vol. 66, no. 6, p. 06RM01, 2021.

[45] K. Mertens, D. Slaets, B. Lambert, M. Acou, F. De Vos, and I. Goethals, “PET with ^{18}F-labelled choline-based tracers for tumour imaging: a review of the literature,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, pp. 2188–2193, 2010.

[46] Z. Mirikharaji and G. Hamarneh, “Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11.* Springer, 2018, pp. 737–745.

[47] C. Nanni, S. Fanti, and D. Rubello, “^{18}F-DOPA PET and PET/CT,” *Journal of Nuclear Medicine*, vol. 48, no. 10, pp. 1577–1579, 2007.

[48] M. S. Nosrati and G. Hamarneh, “Incorporating prior knowledge in medical image segmentation: a survey,” *arXiv preprint arXiv:1607.01092*, 2016.

[49] A. R. Pantel, D. Ackerman, S.-C. Lee, D. A. Mankoff, and T. P. Gade, “Imaging Cancer Metabolism: Underlying Biology and Emerging Strategies,” *Journal of Nuclear Medicine*, vol. 59, no. 9, pp. 1340–1349, 2018.

[50] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, “Semantic Image Synthesis with Spatially-Adaptive Normalization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.

[51] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18.* Springer, 2015, pp. 234–241.

[52] L. Sibille, R. Seifert, N. Avramovic, T. Vehren, B. Spottiswoode, S. Zuehlsdorff, and M. Schäfers, “^{18}F-FDG PET/CT Uptake Classification in Lymphoma and Lung Cancer by Using Deep Convolutional Neural Networks,” *Radiology*, vol. 294, no. 2, pp. 445–452, 2020.

[53] C. D. Soffientini, E. De Bernardi, F. Zito, M. Castellani, and G. Baselli, “Background based Gaussian mixture model lesion segmentation in PET,” *Medical Physics*, vol. 43, no. 5, pp. 2662–2675, 2016.

[54] Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J. E. Bayouth, J. M. Buatti, and X. Wu, “Optimal Co-Segmentation of Tumor in PET-CT Images With Context Information,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1685–1697, 2013.

[55] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, “Combo loss: Handling input and output imbalance in multi-organ segmentation,” *Computerized Medical Imaging and Graphics*, vol. 75, pp. 24–33, 2019.

[56] S. Tan, L. Li, W. Choi, M. K. Kang, W. D D’Souza, and W. Lu, “Adaptive region-growing with maximum curvature strategy for tumor segmentation in ^{18}F-FDG PET,” *Physics in Medicine & Biology*, vol. 62, no. 13, p. 5383, 2017.

[57] Y. Tang, Y. Tang, Y. Zhu, J. Xiao, and R. M. Summers, "A disentangled generative model for disease decomposition in chest X-rays via normal image synthesis," *Medical Image Analysis*, vol. 67, p. 101839, 2021.

[58] T. G. Turkington, "Introduction to PET Instrumentation," *Journal of Nuclear Medicine Technology*, vol. 29, no. 1, pp. 4–11, 2001.

[59] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep Image Prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.

[60] S. Vandenberghe, P. Moskal, and J. S. Karp, "State of the art in total body PET," *EJNMMI Physics*, vol. 7, pp. 1–33, 2020.

[61] T. Wang, Y. Lei, E. Schreibmann, J. Roper, T. Liu, D. M. Schuster, A. B. Jani, and X. Yang, "Lesion segmentation on ^{18}F-fluciclovine PET/CT images using deep learning," *Frontiers in Oncology*, vol. 13, 2023.

[62] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang *et al.*, "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, 2023.

[63] A. J. Weisman, M. W. Kieler, S. B. Perlman, M. Hutchings, R. Jeraj, L. Kostakoglu, and T. J. Bradshaw, "Convolutional Neural Networks for Automated PET/CT Detection of Diseased Lymph Node Burden in Patients with Lymphoma," *Radiology: Artificial Intelligence*, vol. 2, no. 5, p. e200016, 2020.

[64] X. Wu, L. Bi, M. Fulham, and J. Kim, "Unsupervised Positron Emission Tomography Tumor Segmentation via GAN based Adversarial Auto-Encoder," in *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2020, pp. 448–453.

[65] T. Xia, A. Chartsias, and S. A. Tsaftaris, "Pseudo-healthy synthesis with pathology disentanglement and adversarial learning," *Medical Image Analysis*, vol. 64, p. 101719, 2020.

[66] F. Yousefirizi, A. K. Jha, J. Brosch-Lenz, B. Saboury, and A. Rahmim, "Toward High-Throughput Artificial Intelligence-Based Segmentation in Oncological PET imaging," *PET Clinics*, vol. 16, no. 4, pp. 577–596, 2021.

[67] F. Yousefirizi, I. S. Klyuzhin, J. H. O, S. Harsini, X. Tie, I. Shiri, M. Shin, C. Lee, S. Y. Cho, T. J. Bradshaw *et al.*, "TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images—a multi-center generalizability analysis," *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–18, 2024.

[68] H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, "Automated Radiation Targeting in Head-and-Neck Cancer Using Region-Based Texture Analysis of PET and CT Images," *International Journal of Radiation Oncology* Biology* Physics*, vol. 75, no. 2, pp. 618–625, 2009.

[69] Y. Zhang, X. Lin, Y. Zhuang, L. Sun, Y. Huang, X. Ding, G. Wang, L. Yang, and Y. Yu, “Harmonizing Pathological and Normal Pixels for Pseudo-Healthy Synthesis,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2457–2468, 2022.

[70] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss Functions for Image Restoration With Neural Networks,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[71] X. Zhao, L. Li, W. Lu, and S. Tan, “Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network,” *Physics in Medicine & Biology*, vol. 64, no. 1, p. 015011, 2018.

[72] Z. Zhong, Y. Kim, L. Zhou, K. Plichta, B. Allen, J. Buatti, and X. Wu, “3D fully convolutional networks for co-segmentation of tumors on PET-CT images,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 228–231.

[73] M. Zhuang, R. A. Dierckx, and H. Zaidi, “Generic and robust method for automatic segmentation of PET images using an active contour model,” *Medical Physics*, vol. 43, no. 8Part1, pp. 4483–4494, 2016.