This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

KING'S College LONDON

Deep Learning for Improved Quantification of Uptake in Hybrid Imaging

Dal Toso, Laura

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 31. Oct. 2023

Deep Learning for Improved Quantification of Uptake in Hybrid Imaging

Laura Dal Toso

Supervised by:
Paul K. Marsden
Julia A. Schnabel

Thesis submitted to King's College London in fulfilment
of the degree Doctor of Philosophy

September 2021

Abstract

In clinical positron emission tomography (PET) imaging, quantification of radiotracer activity in tumours is often performed using semi-quantitative measurements such as the standardised uptake value (SUV). The accuracy of SUV and related metrics is limited by the noise properties of PET images and the partial volume effect, especially when small tumours are considered. More accurate and reproducible quantification of radiotracer uptake would allow for better disease characterisation and ultimately for a better management of oncologic patients. In this thesis, a 3D deep learning approach is developed with the aim of improving quantification of tumour radiotracer activity concentration and tumour shape definition in PET images. A convolutional neural network (CNN) is trained on a set of simulated "ground truth" images, where simulated tumour shapes are assigned with given radiotracer distributions, and on a corresponding set of simulated PET images. In initial proof of concept studies, PET images are generated by applying noise to the ground truth images and by subsequently applying a smoothing filter. The procedure is then extended by using a much more realistic PET simulation based on an analytical simulator that accurately models spatial resolution and noise properties. When tested on an unseen set of simulated PET images the network predicts enhanced images which are qualitatively and quantitatively assessed. The quantitative assessment is performed by measuring the mean structural similarity index between the ground truth images and the predicted images, and by comparing the maximum, mean and median tumour radioactivity concentrations. The network yields denoised images and it improves the tumour shape definition. The network is at first trained on data generated with a single scanner geometry, that corresponds to data generated with one point spread function. The same network is then applied to PET images generated

with different point spread functions in order to test its generalisability. Overall the network is able to recover denoised images with better defined tumour shapes. The CNN trained on synthetic data is applied to three clinical studies, namely a non-small cell lung cancer (NSCLC) dataset, a Hodgkin lymphoma dataset and a prostate cancer dataset. Encouraging results are obtained for the NSCLC and the Hodgking lymphoma PET datasets. The results predicted by the network are less accurate when the CNN is applied to the prostate cancer dataset, where tumours characterised by radiotracer distributions that differ widely from the ones simulated in the training dataset.

Acknowledgements

Firstly, I would like to thank my supervisors Paul Marsden and Julia Schnabel for their continuous support and advice throughout my PhD. I would also like to thank the HYBRID 2020 consortium for the fantastic opportunities that I have been offered and that made my PhD experience unique. I would like to thank especially Irène Buvat, Claude Comtat, Zacharias Chalampalakis, Ronald Boellaard and Elizabeth Pfaehler who collaborated with me and who have helped me to achieve the results shown in this thesis. I am grateful to my colleagues at King's College: Lucy, Chris, Siti and Colm for always being available when I needed help and for making my lunch breaks more enjoyable. Thanks also to Sam and Camila who are great mentors and friends. I would like to thank Meghana Kulkarni, Gary Cook, Vicky Goh, Manil Subesinghe and Sally Barrington who shared their invaluable insights on my work, and who provided the clinical datasets and measurements that were used in this thesis. Thank you to my friends in London, who kept me company and who explored the city with me. Thanks also to my friends in Italy and Germany who supported me over these years, thanks for always being there for me even at a distance. An honourable mention goes to Anita, who knows and understands me better than most. A special thanks goes to my family, for always believing in me and for welcoming with a hug every time I fly back home. Finally, a huge thanks to Manuel, my favourite person in the world, without whom I would not have made it this far.

This PhD position was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 764458. The work was also funded in part by the Wellcome EPSRC Centre for Medical Engineering at King's College London under Grant [WT203148/Z/16/Z], and in part by the National Institute for Health Research

(NIHR) Biomedical Research Centre Award to Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Contents

Abstract **1**

Acknowledgements **3**

List of Figures **10**

List of Tables **11**

Introduction **12**

1 Background **16**

1.1 Introduction to Medical Imaging 16

1.1.1 Positron Emission Tomography 16

1.1.2 Scintigraphy 17

1.1.3 SPECT 18

1.1.4 Ultrasound 18

1.1.5 Computed Tomography (CT) 19

1.1.6 Magnetic Resonance Imaging (MRI) 20

1.1.7 Hybrid Imaging: PET/CT and PET-MR 21

1.2 Magnetic Resonance Imaging 24

1.3 Positron Emission Tomography 28

1.3.1 Clinical usage of PET in Oncology 28

1.3.2 PET Radiotracers 29

1.3.3 Radioligand Theranostics 30

1.3.4 PET Image Acquisition 31

1.3.5 Spatial Resolution of a PET Scanner 35

1.3.6 Data Corrections for Quantitative PET 38

1.3.7 Image Reconstruction 41
1.3.8 Uptake Quantification in Oncology 49
1.4 Deep Learning in Medical Imaging 52
1.4.1 Neural Networks: Components, Training and Testing . . 54
1.4.2 Convolutional Neural Networks 59
1.5 Summary 61

2 Improving Uptake Quantification on Simulated PET Images 62
2.1 Deep Learning Method 63
2.1.1 Simulation of Training and Testing Datasets 64
2.1.2 Image Evaluation Metrics 67
2.1.3 CNN Architecture 70
2.2 Results 70
2.2.1 Comparison between CNNs composed of 4 to 8 layers . . 71
2.2.2 Impact of Different Network Optimisation Algorithms . . 78
2.2.3 Impact of Different Data Normalisation Methods 80
2.2.4 Impact of the Number of Epochs Used for Training . . . 82
2.2.5 Optimised CNN: a more Detailed Analysis 85
2.2.6 Comparison with Standard Deconvolution and Denoising Methods 87
2.2.7 CNN robustness to PET Images Generated with Multiple PSFs 93
2.3 Application to Phantom PET Images 96
2.4 Summary 98

3 Improving Uptake Quantification on PET Images Generated with an Analytical Simulator 102
3.1 Simulation of Images 104
3.2 Results 107
3.2.1 Comparison between CNNs composed of 5 to 7 convolutional blocks 108
3.2.2 Small Tumours Analysis 111
3.2.3 Impact of the Number of Epochs Used for Training . . . 115
3.2.4 Impact of Different Data Normalisation Methods 117

3.2.5 Comparison with Standard Deconvolution and Denoising Methods . . . 121
3.2.6 Application to Data Generated with Different PSFs . . . 123
3.2.7 Application to Phantom Data . . . 127
3.3 Summary . . . 128

4 Clinical Applications 132
4.1 Non-Small Cell Lung Cancer PET Dataset . . . 133
4.1.1 Neural network . . . 134
4.1.2 Results . . . 135
4.2 Hodgkin Lymphoma PET dataset . . . 140
4.2.1 Neural network . . . 141
4.2.2 Results . . . 141
4.3 Prostate Cancer PET-MR Dataset . . . 148
4.3.1 Method . . . 150
4.3.2 Results . . . 155
4.4 Summary . . . 158

5 Discussion and Future Work 162
5.1 Discussion . . . 162
5.1.1 Simulation of Tumour Shapes . . . 162
5.1.2 Simulation of PET Images . . . 163
5.1.3 Application on Phantom and Real Patient Data . . . 164
5.1.4 Network Optimisation . . . 165
5.1.5 Deep Learning Limitations . . . 166
5.2 Conclusions . . . 167
5.2.1 Future Prospects . . . 169

Publications 171

Appendix 193

List of Figures

1.1 CT scanner 20
1.2 PET/CT scanner 22
1.3 PET-MR scanner 23
1.4 MR scanner 24
1.5 Positron emission and positron range 32
1.6 PET scanner and LOR definition 32
1.7 Effects degrading spatial resolution in PET 36
1.8 Attenuated, scattered and random coincidences in PET 39
1.9 Image reconstruction basics 42
1.10 Iterative image reconstruction 45
1.11 Spill-in and spill-out effects 50
1.12 Illustration of a neuron 55
1.13 Hyperbolic tangent and sigmoid function 56

2.1 Deep learning method overview 64
2.2 Examples of ground truth and simulated PET images 67
2.3 Schematic CNN architecture 70
2.4 Images obtained with different CNN architectures 72
2.5 RC_{max} curves: CNNs made of 4 and 5 convolutional layers . . . 73
2.6 RC_{max} curves: CNNs made of 6 and 7 convolutional layers . . . 74
2.7 CNNs with batch normalisation and fully connected layer 77
2.8 Loss function plots, 100 epochs 79
2.9 Loss function plots, 175 epochs 80
2.10 Impact of the number of epochs on a uniform tumour 84
2.11 Impact of the number of epochs on a hollow tumour 85
2.12 RC_{peak} plot 86

2.13 RC_{max} and RC_{median}: sphericity correlation 88
2.14 RC_{mean} and RC_{peak}: sphericity correlation 89
2.15 Standard methods applied to a hollow tumour 91
2.16 Standard methods applied to a uniform tumour 92
2.17 Dataset generated with multiple PSFs 95
2.18 Application to phantom PET images (OSEM) 98
2.19 Application to phantom PET images (ToF and PSF modelling) 99

3.1 Deep learning method overview 103
3.2 Basic CNN architecture 104
3.3 XCAT phantom 106
3.4 Ground truth image and attenuation map 107
3.5 Impact of the number of layers: uniform tumour 110
3.6 Impact of the number of layers: heterogeneous tumour 111
3.7 Impact of the number of layers: hollow tumour 112
3.8 RC curves: CNN made of 6 blocks 114
3.9 Small tumours RC_{max} and sphericity 115
3.10 Impact of the number of epochs: uniform tumour 116
3.11 Impact of the number of epochs: hollow tumour 117
3.12 Normalisation: uniform tumour 119
3.13 Normalisation: heterogeneous tumour 120
3.14 Normalisation: hollow tumour 121
3.15 Comparison between CNN and standard post-processing methods 124
3.16 Images generated with multiple PSFs 126
3.17 Small tumours RC_{max} and sphericity 127
3.18 Phantom PET images (OSEM) 129
3.19 Phantom PET images (PSF and ToF) 130

4.1 NSCLC: two representative tumours 136
4.2 NSCLC: two hollow tumours 137
4.3 NSCLC: SUV_{max} 138
4.4 NSCLC: SUV_{mean} 139
4.5 NSCLC: SUV_{peak} 139
4.6 RATHL: visual assessment 142
4.7 RATHL: visual assessment 143

4.8 RATHL: SUV_{max} . . . 145
4.9 RATHL: SUV_{mean} . . . 146
4.10 RATHL: SUV_{peak} . . . 147
4.11 PROSPER: Method . . . 151
4.12 PROSPER: UNet . . . 153
4.13 PROSPER: Convolution Block . . . 154
4.14 PROMISE: examples of prostate segmentation . . . 156
4.15 PROSPER: examples of prostate segmentation . . . 156
4.16 PROSPER: two ^{68}Ga-PSMA PET scans . . . 158
4.17 PROSPER: two ^{11}C-Choline PET scans . . . 159

5.1 NSCLC: two representative tumours . . . 194
5.2 NSCLC: two representative tumours . . . 195
5.3 RATHL: visual assessment . . . 196
5.4 RATHL: visual assessment . . . 197
5.5 PROSPER: two PSMA PET scans . . . 198
5.6 PROSPER: two ^{11}C-Choline PET scans . . . 199

List of Tables

2.1 RC obtained using different CNNs 75
2.2 MSSIM obtained using different CNNs 76
2.3 RC obtained using BN and FC layers 78
2.4 RCs: different data normalisation methods 82
2.5 RC obtained using different optimisers 84
2.6 RC measured using standard methods 93
2.7 RC measured on images generated with different PSFs 94
2.8 RC_{max} measured on phantom images 97
2.9 RC_{mean} measured on phantom images 98

3.1 RCs: CNNs with different depths 113
3.2 RCs: small tumours 113
3.3 RCs: CNNs trained for different numbers of epochs 118
3.4 RCs: different normalisation methods 122
3.5 RCs: standard post-processing methods 123
3.6 RCs: multiple PSFs 125
3.7 Small tumours RCs: multiple PSFs 127
3.8 RC_{max} phantom PET scans 128

Introduction

Motivation of this Thesis and Related Work

Positron emission tomography (PET) is an imaging modality which enables *in vivo* quantitative assessment of molecular targets for a wide range of biochemical processes. This imaging modality is widely used in oncology to obtain information about tumour metabolism. Visual and quantitative assessment of tumour metabolism is routinely used to perform tumour staging and restaging, treatment monitoring and radiotherapy planning. In PET, quantification of radiotracer activity concentration is usually performed using standardised uptake value and related metrics. However, the accuracy of these metrics is limited by the poor spatial resolution and noise properties of PET images. There is need for new methods that allow for accurate and reproducible quantification of radiotracer activity concentration. A better quantification of tumour radioactivity could lead to a more accurate patient stratification and it could thus improve disease management.

Improved quantification of tumour radioactivity can be obtained by applying image quality enhancing methods, that can be incorporated in the image reconstruction process [1, 2] or applied post-reconstruction. In clinical practice, PET raw data and list-mode data are often not available and post-reconstruction techniques have the advantage of being easily applicable retrospectively. Some approaches such as Lucy-Richardson deconvolution [3, 4]

aim to improve PET image quality by performing partial volume correction (PVC). Deconvolution-based PVC methods are commonly used [5], but these methods inevitably lead to noise amplification, and ringing artifacts can appear near sharp edges. Standard PVC methods make assumptions on the point spread function (PSF) of the imaging system or about PET image noise [6]. Other standard approaches focus on improving PET images quality by applying filters (such as Gaussian or median) to the noisy images. These methods often do not provide optimal results as the filtering operation causes blurring and thus a loss of information. Bilateral filtering [7] and non-local means (NLM) filtering [8] have been used to denoise PET images while preserving more image structures. However, the sharpness of edges in filtered images is always reduced.

In recent years deep learning (DL) algorithms have been used to enhance PET image quality. For example, Gong *et al.* [9] achieved image denoising through fine tuning of a deep convolutional neural network (CNN). Their results indicate that the method they developed is more effective in removing noise than the traditional Gaussian filtering. However, some limitations of their method are that the simulation of PET images did not include scatter and that the approach is 2D thus it does not fully exploit the 3D information provided by PET. A non-DL joint solution for PET image segmentation, denoising and partial volume correction has been developed by Xu *et al.* [10]. The idea behind this method is that denoising, segmentation and PVC are closely related to each other, and the improvement of one could simplify or improve the solution of the other. A weakness of this approach is that it involves the integration of three different frameworks, thus introducing many parameters that need to be manually tuned. Furthermore, this joint approach requires an accurate tumour delineation.

Other DL algorithms aim to enhance PET quality by mapping low count

imaging to high-count images. Lu *et al.* investigate the quantitative accuracy of DL based denoising in oncological PET [11]. In this work, a UNet was trained to recover standard dose PET images from 10% low-dose PET data. Their UNet outperformed several standardly used denoising methods including Gaussian filter and anatomical-guided non-local means filter. The UNet could achieve a superior trade-off between noise and bias at minimal computational cost. Similarly, in another work a 3D CNN was trained with real patient data with the aim of recovering high-count PET images from low-count PET images [12]. A wide range of noise levels was emulated from a set of chest PET data in patients with lung cancer and a CNN was trained to denoise the reconstructed PET images using the full-count reconstructions as the ground truth. The technique presented offered limited benefit for detection performance when using images with count levels routinely encountered in a clinical setting. Song *et al.* developed a super-resolution imaging technique for PET images based on convolutional neural networks (CNNs) [13]. Super-resolution imaging refers to the task of converting a low-resolution image to a high-resolution one. Their results indicate that CNNs outperform penalised deconvolution and PVC at the super-resolution imaging task, and that deep CNNs outperform shallow CNNs. To the best of our knowledge there is no unique, reproducible and accurate post-processing method that can simultaneously provide PVC and image denoising in clinical oncologic PET images.

Aims of this Work and Outline

The main goals of this work are:

- To develop a 3D deep learning approach that can simultaneously provide PVC and PET image denoising with the aim of improving the quantification of tumour radiotracer uptake

- To apply the proposed approach to PET images generated with different point spread functions, that simulate images acquired with different scanners
- To apply the proposed method to phantom data
- To apply the proposed method to real patient datasets and compare the results obtained with the deep learning approach to clinical measurements

A short overview of each chapter is presented here. In chapter 2 the deep learning approach is presented, together with the preliminary experiments performed training and testing a 3D CNN on a simulated dataset. In this chapter, a simplistic method is used to generate the simulated PET images used for training. The optimisation process of the network architecture and other training parameters is presented, together with a quantitative and qualitative analysis of the results.

In chapter 3, a 3D CNN is trained on more realistic PET images, generated with an analytical simulator. The images predicted by the 3D CNN are visually and quantitatively compared to the simulated PET images. This chapter also presents an application of the deep learning method to phantom images reconstructed with different algorithms.

In chapter 4, three clinical applications of the proposed deep learning method are described. The results obtained using the proposed approach are compared to clinical measurements provided by experts.

Finally, in chapter 5 a general discussion of the results achieved in this thesis is presented, together with a prospect of possible future work perspectives.

Chapter 1
Background

1.1 Introduction to Medical Imaging

Medical imaging allows the investigation of human body through images. Different modalities correspond to different types of medical images which can be used for patient screening, diagnosis, staging and treatment monitoring. Some modalities such as scintigraphy, single photon emission tomography (SPECT) and positron emission tomography (PET) provide functional information, whereas other modalities such as ultrasound (US) and computer tomography (CT) provide anatomical information. Magnetic resonance (MR) provides primarily anatomical information and is increasingly being used to extract functional information. Hybrid imaging techniques have been developed in recent years to obtain with a single device both functional and anatomical images. In the following sections, an overview of some significant examples of imaging modalities currently used in clinical practice is presented.

1.1.1 Positron Emission Tomography

Positron emission tomography (PET) is an imaging modality which enables *in vivo* quantitative assessment of molecular targets for a wide range of biochemical processes. This can be achieved by using chemical compounds labelled

with positron emitting radionuclides, which do not interfere with the biological functions of the chemical compounds. The labelled compounds, which are called radiotracers, are usually injected intravenously and spread in the whole body after a given amount of time. Due to the high sensitivity of PET scanners, small concentrations of radioactivity down to the picomolar scale [14] can be imaged. The radioactive decay events that happen in the body within the scan duration time are detected by the PET detector ring. From these data, it is possible to create an image of the radioactivity concentration in the body through the image reconstruction process. In oncology, PET is clinically used for cancer diagnosis, staging and restaging and treatment monitoring. The visual and quantitative information extracted from PET can be used for radiation therapy planning. In the field of neurology, PET enables *in vivo* examination of brain functions [15]. ^{18}F-FDG (Fluorodeoxyglucose) PET is a valuable tool for early diagnosis and differentiation of dementia, even before the onset of clinical symptoms. Alzheimer's disease (AD) patients exibit characteristic cerebral metabolic alterations which can be seen in PET imaging. PET imaging is also used to understand the role of neurotransmitters in epileptogenesis, identify epileptic foci and investigate new treatment approaches [16]. PET has been successfully applied to cardiology for quantification of myocardial perfusion and assessment of left ventricular function in patients with known or suspected coronary artery disease [17–19]. PET has been used as primary imaging modality for the development of the method proposed in this thesis. A more detailed explanation of PET imaging is given in section 1.3.

1.1.2 Scintigraphy

Scintigraphy, also known as planar single photon radionuclide imaging, is an imaging modality that provides functional information. Similar to PET, in

scintigraphy a photon-emitting labelled radiotracer is injected in the patient's blood stream. Most commonly Technetium 99m is used, a pure gamma radiation emitter emitting photons with energy of 140 keV. During the scan, the gamma rays emitted in decay events are detected in a gamma camera. 2D images can be reconstructed from the acquired data. Bone scintigraphy is used clinically to detect prostate cancer and breast cancer skeletal metastases [20, 21]. Additionally, scintigraphy can be used for diagnosing gastrointestinal disease and to assess cardiac perfusion and cardiac viability [22].

1.1.3 SPECT

SPECT is a functional imaging modality similar to scintigraphy. A photon emitting radiotracer is injected intravenously before the scan. During the scan a gamma camera rotating around the patient acquires the emitted photons from multiple angles. 3D images can be reconstructed from the acquired data using tomographic reconstruction algorithms. SPECT scans are less expensive than PET as the radioisotopes used in SPECT can be obtained more easily. However, in PET it is possible to obtain a higher spatial resolution and sensitivity thanks to the coincidence timing. SPECT can be used to complement information from scintigraphy studies, whenever 3D images are needed. It can also be used for cardiac perfusion or for brain functional imaging [23].

1.1.4 Ultrasound

Ultrasound imaging uses high-frequency sound waves to generate images. A transducer is used to transmit sound waves by converting electrical energy into mechanical energy in the form of vibration. When the sound waves enter tissues with different densities, and so with different acoustic impedance, they are reflected and subsequently detected by the transducer. Tissue with

high acoustic impedance will appear brighter in ultrasound images. Ultrasound can be used to perform elastography studies [24]. Elastography is used to assess the elastic properties of soft tissue non-invasively. It is based on the different propagation of waves in tissue with different elastic properties. At first mechanical waves are applied to deform a tissue, then the tissue response is evaluated and a quantitative map, called elastogram, of the stiffness and stiffness-related parameters is shown. Solid tumours are typically stiffer than healthy tissue. Ultrasound imaging is generally real time and low cost compared to other imaging modalities. It is widely used in cancer screening, vascular disease and for image-guided procedures. However, ultrasound images highly depend on the image acquisition protocol and on the operator. For this reason, other techniques such as CT or MRI are preferred for applications such as longitudinal oncology studies [25].

1.1.5 Computed Tomography (CT)

Computed tomography (CT) is an imaging modality based on the use of a rotating X-ray source to acquire projection data from multiple angles. Opposite to the X-ray source a set of detectors record the intensity of the beam passing through the body, along the path connecting the source to each individual detector element. The attenuation of X-rays depends on the density and the atomic number of the attenuating material. CT images can be reconstructed using tomographic reconstruction algorithms. The reconstructed CT images are maps of the linear attenuation coefficients μ of the imaged material. They are usually expressed in Hounsfield unit (HU) scale, in which the measured attenuation values μ are normalised by the attenuation coefficient of water μ_{water}. The mathematical definition of Hounsfield unit is as follows:

$$HU = (\frac{\mu - \mu_{water}}{\mu_{water}}) \cdot 1000. \tag{1.1}$$

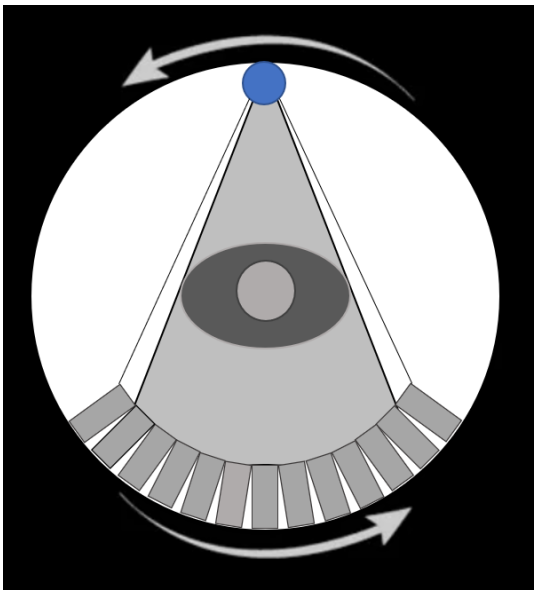


Figure 1.1: Illustration of a CT scanner. An X-ray source, shown in blue, rotates in a circular motion. Opposite to the X-ray source, a set of detectors can record at multiple angles the intensity of the beam passing through the body. The reconstructed CT images are maps of the linear attenuation coefficients μ of the imaged material.

CT is used in oncology for diagnosis, staging and restaging of lung, colorectal, oesophageal, breast, and head and neck cancers, as well as lymphoma [26].

1.1.6 Magnetic Resonance Imaging (MRI)

Magnetic resonance (MR) is an imaging technique based on the magnetic resonance of atomic nuclei. MR targets protons in the nucleus of hydrogen atoms, which are distributed in the whole body. Protons are characterised by an extremely small magnetic moment, and when placed in a magnetic field the magnetic moment precesses about its axis at an angular frequency proportional to the field strength. Radiofrequency (RF) coils within the MR scanner are used to transmit the RF energy to tissues. If RF oscillating at the resonant frequency of the proton is applied along the x and y planes, a resonance happens between the RF wave and the proton spins. When RF is applied

the magnetic moments tilt away from the external field direction. As the net magnetic moment precesses and crosses the external field direction, it induces a voltage (MR signal) that can be detected and measured by the receiver coils. In the reconstructed MR images, the brightness of different voxels is related to different properties of the underlying tissues. MRI provides excellent soft tissue contrast and enables functional imaging such as blood flow, tissue perfusion or diffusion. A more detailed description of MR and its applications is given in section 1.2.

1.1.7 Hybrid Imaging: PET/CT and PET-MR

PET scanners can be paired with other non-invasive imaging techniques, such as magnetic resonance (MR) or computed tomography (CT) which provide anatomical information. Historically, the fusion of PET, CT and MR images was at first achieved with post-processing methods, by registering retrospectively two images of the same patient, acquired in different scanners [27, 28]. These aligned images were usually visualised next to each other or superimposed, to make use of both anatomical and functional information. Software registration algorithms have been successfully applied to rigid organs such as brain, but were less successful in other parts of the body affected by more complex motion [29]. In later years, hybrid scanners have been developed with the aim of incorporating two imaging modalities in the same device.

PET/CT The first PET/CT scanner was built in the 1990s [31]. In PET/CT scanners, PET and CT systems are combined in a single device. As the time required for a CT scan is very short compared to the time needed for PET, the two scans can be acquired sequentially. As shown in Fig. 1.2 the CT components are mounted at the front of the scanner, while the PET components are mounted at the back. The acquisition and reconstruction paths are not in-

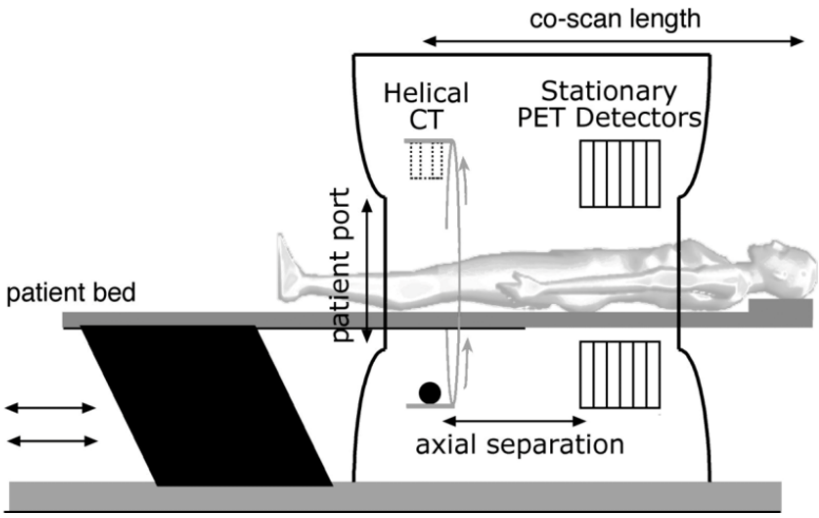


Figure 1.2: Illustration of a PET/CT scanner. The CT components are mounted at the front of the scanner, while the PET components are mounted at the back. The acquisition and reconstruction paths for PET and CT are not integrated. Image taken from Schmitz *et al.* [30].

tegrated. The CT images are reconstructed first, providing attenuation maps that can be used to correct the PET data. PET and CT images are usually fused in post-processing and displayed together. PET/CT is routinely used in oncology for diagnosis, staging and restaging of lung cancer, lymphoma, head and neck tumours and colorectal cancer [32]. The superior spatial resolution of CT images (around 0.5 mm) enables the detection and more precise delineation of small lesions that would appear blurred in PET. The PET imaging component provides information on lesion metabolism and thus helps to classify lesions as malignant or benign.

PET-MR In 2010, the first clinical whole body PET-MR scanner was released [33]. It was more complex to build this type of integrated scanner, as the magnetic fields involved in MR posed some limitations to the materials that could be used in the PET system. The first scanners, the PET and MR scanners were separated, and patients could be moved from one to the other on a revolving bed [34]. Subsequently, simultaneous PET-MR scanners have

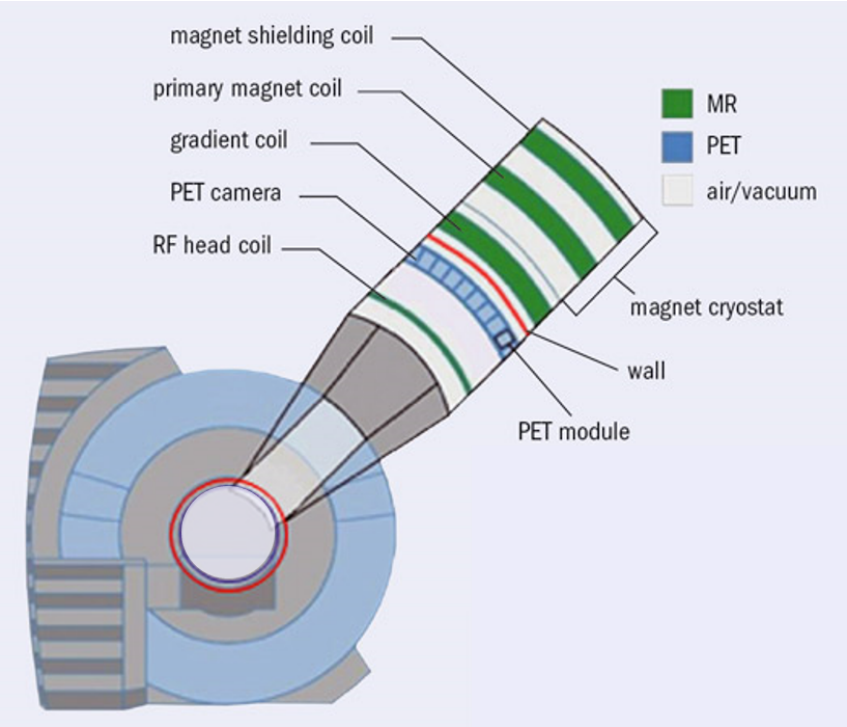


Figure 1.3: Illustration of a PET-MR scanner. PET and MR systems are integrated in the same device for simultaneous acquisition. Image adapted from: Cern courier "PET and MRI: providing the full picture".

been developed, in which PET and MR were fully integrated [35] as shown in Fig.1.3. PET-MR has the great advantage of exposing patients to a minor dose compared to PET/CT, where the CT acquisition accounts for at least 50 % of the total radiation dose of a PET/CT scan [36, 37]. The main disadvantages of PET-MR is that it is more costly, more complex to operate and less widely available than PET/CT. Unlike stand-alone PET or PET/CT, PET-MR cannot directly provide attenuation maps. New sequences have been developed for attenuation correction in the body, but they have not been widely validated or approved for use in clinical practice. Additionally, some factors such as noise generated in the acquisition of MR sequences and the small diameter of the scanner bore may cause patient discomfort and hamper compliance during the scan. PET-MR often requires acquisition times of around 1h, as

multiple sequences are usually acquired in a scan, whereas PET/CT scans are typically completed in 30 minutes. The longer scan timing of PET-MR allow for longer PET acquisitions, which may lead to higher quality PET images. Currently, PET-MR is mainly used in clinical contexts that would require sequential PET and MR, or in pediatrics where lower radiation doses might be preferable especially when serial follow-up examinations are needed [38].

1.2 Magnetic Resonance Imaging

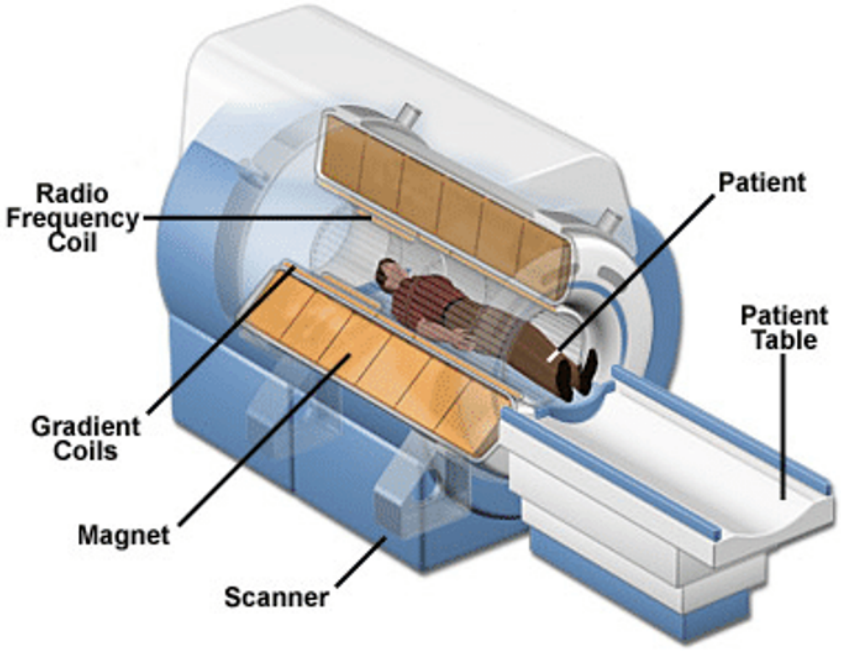


Figure 1.4: Schematic representation of a magnetic resonance scanner, adapted from [39]. A magnet coil generates a magnetic field B0, which causes the protons to align their magnetic moments along the direction of B0. Radiofrequency (RF) coils create RF waves, that are applied to the imaged object to generate the MR signal. Gradient coils are used to impose variations to B0, which in turn allow for spatial encoding and slice selection.

In this section MRI is described more in detail. As mentioned previously, MR is an imaging technique that targets protons. When protons are not

subject to an external magnetic field, their magnetic moments are randomly oriented, so there is no net magnetisation. As shown in Fig. 1.4, a major component of the MR scanner is a magnet that generates a magnetic field B0 (usually 1.5 or 3 tesla (T)). Another important component of the MR scanner are the gradient coils, which are used to impose variations to B0, which in turn allow for spatial encoding and slice selection. When B0 is applied, the protons gradually align their moments along the direction of B0. Although the moment alignment is only partial, there is a slight prevalence of parallel aligned moments so there is a net magnetisation along z. Furthermore, a precession of the magnetic moments around the direction of B0 can be observed. Magnetic moments are evenly distributed 360° around the z-axis and are precessing out of phase with each other. In the inner part of the bore, there are radiofrequency (RF) coils, which can transmit the RF energy to the tissue. If RF oscillating at the resonant frequency of the proton is applied along the x and y planes, a resonance happens between the RF wave and the proton spins. When RF is applied the magnetic moments tilt away from the B0 direction. The magnetisation along the longitudinal axis decreases, and the moments are grouped at one side of the z axis, parallel to the RF direction. As the net magnetic moment precesses and crosses the external field direction, it induces a voltage (MR signal or echo) that can be detected and measured by the receiver coils. The RF signal is received by one or more surface coils placed close to the patient's body. Surface coils are essentially made of a loop of conducting material, which can be adapted to better fit specific body parts. The amount of time between successive pulse sequences applied to the same slice is called repetition time (TR). Time to echo (TE) is the time between the delivery of the RF pulse and the detection of the echo signal. When the RF stops, the excited nuclei return to equilibrium. This happens in two separate processes, called T1 and T2 relaxations. T1 relaxation is an exponential process due

to the realignment of the net magnetisation to the direction of the external magnetic field, which can be mathematically expressed as:

$$M_z = M_0\,(1 - e^{-t/T1}). \tag{1.2}$$

T1 (longitudinal relaxation) is a time constant that expresses the rate at which excited protons realign with B0 ad thus return to equilibrium. In normal tissue the typical T1 values range from 0.5 to 5 s at 1.5 T. In parallel, when the RF pulse stops a loss of phase coherence among the nuclei spinning perpendicular to the main field takes place, as some moments speed up and others slow down. T2 relaxation is the decay or dephase process of the transverse components of magnetisation. T2 relaxation is described by a simple exponential decay with time constant T2, as shown in the following equation:

$$M_{xy} = M_0\,e^{-t/T2}. \tag{1.3}$$

T2 is defined as the time required for the transverse magnetisation to fall to approximately 37 % of its initial value. Typical T2 relaxation times associated to normal tissue are 100–300 ms at 1.5 T. In the reconstructed MR images, the brightness of different voxels is related to different tissue properties, such as T1, T2 and proton density. The TR and TE can be adjusted to generate different types of MR images, characterised by different contrasts. In T1-weighted imaging short TE and TR are used. T1-weighted images are used to visualise fat, bone marrow and white matter, which appear brighter. This type of imaging can be used to assess cerebral cortex, aortic valve disease and other heart conditions [40]. In T2-weighted, imaging long TE and TR are used. T2-weighted images enable the visualisation of water rich areas. For example, T2 images are used to study cerebrospinal fluid and inflammation (i.e. edema, infarction, demyelination). Another MR sequence that is

commonly used is called fluid attenuated inversion recovery (FLAIR). It is similar to a T2 sequence but the TR and TE used in this case are very long. FLAIR images are characterised by an excellent contrast between lesions in the white matter or grey matter and the surrounding brain tissue [41]. Diffusion weighted imaging (DWI) is a technique that allows the visualisation of random movements of water protons. It is used in clinical settings to diagnose neurological disorders such as multiple sclerosis or stoke. During ischemia, water movements become more restricted as water is accumulated intracellularly, resulting in an extremely bright signal on DWI. Diffusion-weighted sequences are usually applied in conjunction with apparent diffusion coefficient (ADC) mapping techniques [42]. ADC maps represent the actual diffusion values of the tissue without T2 effects and they appear as greyscale inverted DWI images. A decreased signal in ADC maps denotes restricted diffusion and it usually characterises pathologies such as cancer lesions.

MRI is characterised by high soft tissue contrast, good spatial resolution and it does not require the use of ionizing radiation. For these reasons, this modality is well suited for evaluation of cancer patients, that often require repeated scans. To delineate lesions in radiation oncology more than one MR sequence may be used. Multiple MR images can be subsequently acquired with different scan parameters, so that one exam can offer different visualisations for the same target area [43]. The main limitation of MRI is that it is subject to motion artifacts and it often requires prolonged acquisition times. Another aspect that needs to be considered is that MRI is sensitive to metal devices which may be implanted in the body, such as pacemakers or cochlear implants.

1.3 Positron Emission Tomography

In this section, PET is described more in detail. PET is an imaging technique that uses radionuclides that decay by positron emission to obtain information on patient metabolism. PET images reflect the radioactivity concentration in the patient body in units of kBq/ml. Oncologic PET imaging relies on the fact that most cancer cells exhibit a different metabolism than normal cells. Otto Warburg first demonstrated that tumour cells use higher levels of glucose in the presence of oxygen, with an associated increase in lactate production. The phenomenon aerobic glycolysis, which is referred to as Warburg effect, has been observed in a range of tumour types which include colorectal, breast and lung cancer [44].

1.3.1 Clinical usage of PET in Oncology

In oncology, PET is clinically used for cancer diagnosis, as it enables the detection of increased glucose metabolism that is associated with many types of tumours. Visual and quantitative assessment of tumour metabolism is also routinely used in clinical practice to perform tumour staging, which is the process of determining the extent to which a cancer has developed by growing and spreading in the body. The visual and quantitative information extracted from PET can be used for radiation therapy planning. Measurements of radiotracer uptake and tumour volume, extracted from PET images, are used in clinical settings to stage the tumour, to monitor the response to treatment and predict survival. For example, the percentage maximum SUV remaining in the primary tumour after chemotherapy has been shown to be predictive of survival for non-small cell lung cancer [45]. Residual metabolic tumour volume after treatment was also found to be correlated to lung cancer patients' prognosis. In pancreatic cancer, high metabolic tumour volume and total lesion glycolysis

prior to treatment were associated with inferior overall survival [45]. Improved measurements of tumour volume and of tumour radiotracer uptake can lead to a more accurate patient stratification and thus improve disease management.

1.3.2 PET Radiotracers

PET imaging is based on radiotracers, a type of radioligands used for diagnostic purposes. One of the most common PET radiotracers is ^{18}F-FDG, a glucose analogue that allows the visualisation of glucose metabolism in tissue. When ^{18}F-FDG enters the cells, it is phosphorylated to FDG-6-phosphate (FDG-6-P) by the hexokinase enzyme. The FDG-6-P is not further metabolised so it accumulates intracellularly in proportion to the glycolytic rate of the cells. The higher glucose metabolism that often characterises tumour cells therefore results in higher levels of FDG-6-P [46, 47]. The high activity concentration in cancer tissues corresponds to high intensity areas in PET images. The overwhelming majority of clinical PET studies use FDG. Being a tracer of glucose metabolism, ^{18}F-FDG has the disadvantage of being non-cancer specific [48]. There are several benign conditions, such as inflammation or infection, which lead to increased ^{18}F-FDG uptake [49–51]. Moreover, normal organs like brain, heart and kidneys are associated with a relatively high ^{18}F-FDG uptake which may hamper the detection of small lesions [52]. Furthermore, some types of cancers, such as carcinoid tumor, bronchoalveolar cancer and mucinous adenocarcinoma, have low glucose avidity (comparable to normal surrounding tissue) so ^{18}F-FDG may fail to detect these cancer lesions [48,53]. To address these pitfalls of ^{18}F-FDG, other tracers have been developed. Carbon-11-choline (^{11}C-choline) [54, 55] or Fluorine-18-choline (^{18}F-choline) can be used as tracers of the lipid metabolism, and they are recognised imaging agents for the diagnosis of prostate cancer. A range of tracers based on amino-acid metabolism, such as Carbon-11-methionine (11C-MET), fluorine-18-fluoroethyl tyrosine (^{18}F-FET)

and fluorine-18-fluoromethyl tyrosine (^{18}F-FMT), have shown high sensitivity in detecting low-grade brain tumours. ^{18}F–fluciclovine (FACBC) is another imaging agent based on amino-acid metabolism, and it can be used clinically when recurrent prostate cancer is suspected. New tracers are currently being investigated in first-in-human studies. One example is ^{18}F-tetrafluoroborate, which is a PET tracer for the sodium/iodide symporter (NIS), a membrane glycoprotein that is in charge of the active transport of iodide. This radiotracer would enable the visualisation of accumulation of radioiodide at sites where the NIS is expressed, such as thyroid and salivary glands [56]. Another new radiotracer called ^{18}F-fluoro-pivalic acid (^{18}F-FPIA) is a promising candidate radiotracer for tracing short chain fatty acid metabolism. Another novel radiotracer is ^{227}Th-BAY2287411, a Thorium-227 labelled antibody-chelator conjugate, for which the first in-human study in now being conducted in patients with tumours known to express mesothelin [56].

1.3.3 Radioligand Theranostics

Radioligand theranostics (RT) focus on the molecular features that characterise different tumour cells, rather than the organ where the tumour originates. The goal of RT is to use a single compound for the diagnosis, drug delivery, and treatment-response monitoring of cancer lesions, thus potentially achieving improved tumour control with minimal toxicity. In radioligand therapy, radiation is delivered directly to selected types of tumour cells. This is gradually emerging as a precision therapeutic approach for cancer care. This type of treatment has been shown to improve overall survival and quality of life for patients affected by neuroendocrine cancers and castration-resistant prostate cancer that has metastasised to the bone [57]. RT is based on molecules such as somatostatin, that can be conjugated to different chelators, including DTPA and DOTA. These chelators can accommodate radioisotopes

either for diagnosis (e.g., ^{68}Ga, ^{111}In) or therapy (e.g., ^{177}Lu or ^{90}Y) [58]. The resulting compounds can target the corresponding receptors which are found in various tumour cell types. In medical imaging, ^{68}Ga-DOTATOC and ^{68}Ga-DOTATATE are being used for the diagnosis, staging and treatment monitoring of neuroendocrine tumours [59]. Furthermore, ^{68}Ga-prostate-specific membrane antigen (^{68}Ga-PSMA) can be used to target a cell-surface protein that is overexpressed in some prostate cancers, thus enabling primary diagnosis and staging [60]. PSMA yields an good tumour-to-background radioactivity ratio leading to an improved detection rate. ^{177}Lu-PSMA has been used for the treatment of metastatic castration-resistant prostate cancer [61].

1.3.4 PET Image Acquisition

Radioactive tracers that are injected in the body in preparation for a PET scan undergo radioactive decay. In particular, positron-emitting radionuclides undergo β^+ decay, a process in which a proton inside the nucleus of an unstable nuclide is converted into a neutron while releasing a positron and a neutrino ν_e. In the specific case of ^{18}F, this reaction can be schematised as:

$$^{18}F \rightarrow ^{18}O + \beta^+ + \nu_e. \tag{1.4}$$

As shown in Fig. 1.5, the positron emitted in a β^+ decay travels through tissue for a short distance, called positron range, before annihilating with an electron. The positron produced by ^{18}F has a maximum positron range in water of 2.4 mm and a mean positron range of 0.6 mm [62]. The result of the annihilation is the emission of two back-to-back 511 keV gamma rays according to Einstein's law $E = mc^2$. During a PET scan, these gamma rays are detected by a detector ring that surrounds the patient. As represented in Fig. 1.6, the line that connects the two detectors where the gamma rays are

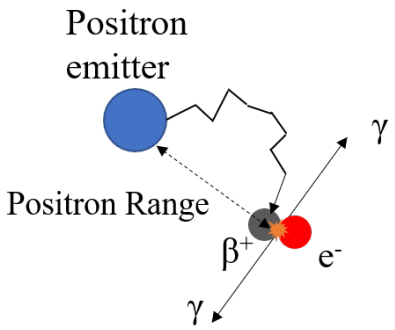


Figure 1.5: Schematic illustration of positron emission. In β^+ radioactive decay, a positron is emitted by a radionuclide. It then travels for short distances in the surrounding material before annihilating with an electron. Two back-to-back photons are produced as a result of the annihilation. The distance between the positron emitter and the annihilation location is referred to as positron range.

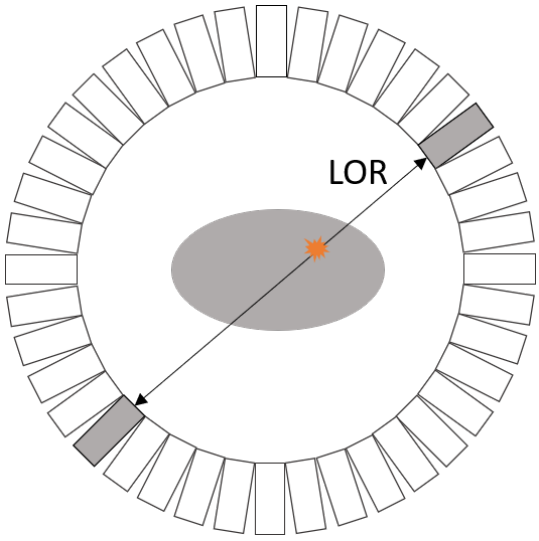


Figure 1.6: Schematic representation of a PET scanner detector ring. Back-to-back photons, emitted in an annihilation event in the imaged object, are detected in coincidence by individual detector elements. The line connecting the two detectors where the photons are detected is called line-of-response (LOR).

detected is called line-of-response (LOR). Each event is stored together with a digital time stamp that indicates when each photon was detected. Unless the decay happens in the centre of the LOR, the two gamma rays hit the detector ring at two different times, with a timing difference of a few nanoseconds. A coincidence processor analyses the time stamps of events acquired in each opposing detector. A coincidence event is recorded when two photons within a selected energy window are collected at opposing detectors within a coincidence window of few nanoseconds [63]. Such coincident events are called prompt events. Time-of-flight (ToF) PET [64] allows the measurement of the time delay Δt between the detection of two coincident gamma rays and so restricts the location of the annihilation to a segment of the LOR. The location of the annihilation event with respect to the midpoint of an LOR can be expressed a:

$$\Delta d = \frac{\Delta t \times c}{2}, \tag{1.5}$$

where c is the speed of light ($\sim 3 \times 10^{10}\, cm\, s^{-1}$). Typically, 10^6 to 10^8 coincidence events are detected in a PET scan. There are inherent fluctuations in the numbers of photons detected and used to generate PET images. These fluctuations, which are the primary source of noise in PET images, are due to the fact that the positron emission process is characterised by Poisson statistics [65,66]. The emitted photons are detected by multiple adjacent detector rings, which compose the PET scanner detection system. First generation PET scanners only acquired coincidences between elements in the same detector ring (direct) and coincidences between elements in adjacent detector rings (cross-plane). Tungsten shields called "septa" were inserted between detector rings to limit the detection of photons scattering at larger angles. With this setting, data collection was restricted to almost parallel 2D planes, thus this acquisition mode is called 2D data acquisition. A higher sensitivity can be

achieved in newer PET systems equipped with smaller detector elements, by increasing the acceptance angle over which the coincident events are acquired. Additional coincidences can be recorded between planes separated by up to 5 or 6 detector rings. A conventional PET camera with septa in 2D acquisition mode detects around 4,000–5,000 coincidence events per 10^6 ($\sim 0.5\%$) positron emissions [67]. The sensitivity of a PET scanner can be further improved using 3D mode data acquisition, in which coincident events obtained from all possible LORs are accepted. For 3D acquisition it is necessary to remove the septa. The 3D acquisition typically yields an improvement of 5 to 7 times in sensitivity compared to 2D acquisition [68]. 3D acquisition mode does not always offer an improvement over 2D mode as several factors hamper image quality [69–71]. 3D acquisition requires more memory for data storage and more sophisticated 3D image reconstruction algorithms.

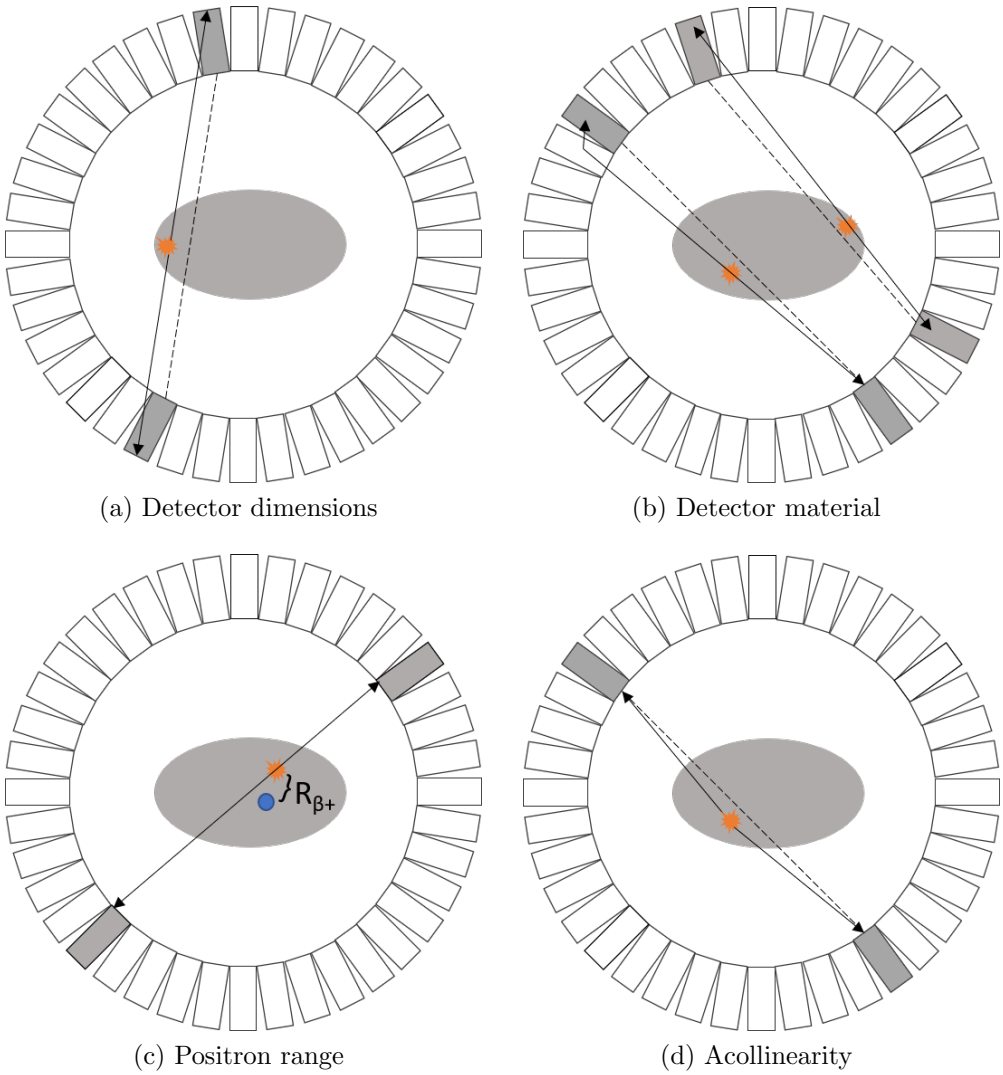
PET Acquisition Protocols The simplest protocol for PET data acquisition is to collect a single dataset over a fixed time frame. In this case, the reconstructed images represent the average tissue activity concentration observed during the acquisition. This type of protocol is used for FDG studies, for which the tracer concentration is relatively stable for the duration of the scan. In other studies there is need to monitor changes of radiotracer distribution in the body over time, so it is more convenient to acquire PET data as a sequence of dynamic time frames. These studies often require monitoring the arterial plasma radioactivity concentration through blood samples. The majority of modern PET scanners are characterised by an axial field of view (FOV) of around 25 cm, which makes them suitable for covering a small region of the body. Whole body scans can be obtained by moving the patient through the PET scanner, allowing the sequential scan of different areas of the body. In recent years, a total body PET system has been developed. In some cases, such as in oncologic PET applications, there is need to monitor patients over

long periods of time. Datasets that comprise multiple scans acquired over time to monitor changes are usually referred to as repeat scans. Research studies may refer to scans acquired over time as "longitudinal" scans.

1.3.5 Spatial Resolution of a PET Scanner

Spatial resolution is a measurement of the smallest distance two object can be placed apart and still be distinguished. It is determined by scanner features but it also depends on the reconstruction parameters. The spatial resolution of a scanner is usually measured according to the NEMA standards [72], by measuring a point source at different locations. Due to the limited spatial resolution, a point source in the PET scanner is visualised in the final image as a blurred object. This blurring, which is approximated by a Gaussian function, is referred to as the point spread function (PSF) of the system [73,74]. Practically, the spatial resolution of a system is defined as the full-width-half-maximum (FWHM) of its PSF. In the next paragraphs the intrinsic limitations to the spatial resolution that can be attained in modern PET scanners are illustrated.

Detector Dimension and Material The first effect deteriorating spatial resolution is associated with the detector dimensions. The geometric resolution at the centre of a LOR connecting two detector elements can be described by a triangular shaped coincidence response function with FWHM equal to half the detector width w. When the source is moved radially towards one of the detectors, the coincidence response function becomes trapezoidal and eventually becomes a square function at the front surface of the detector. The intrinsic resolution is therefore influenced by the width of the detector elements. Furthermore, photons may interact at any depth in the detectors and typically the location of the interaction is not measured. As a result, in order



(a) Detector dimensions

(b) Detector material

(c) Positron range

(d) Acollinearity

Figure 1.7: This figure illustrates some effects that cause mispositioning of line-of response (LOR, shown as dashed lines in the figures), thus degrading the spatial resolution of PET imaging. Fig. (a) shows how the finite width and depth of the individual detector elements can cause mispositioning of a LOR. Fig. (b) shows how intra-detector scatter or penetration of photons to adjacent detector elements can cause LOR mispositioning. Mispositioning of a LOR can also be due to the positron range (c), which is the distance between the locations of positron emission (in blue) and positron annihilation (in orange). Finally, the positron-electron pair may have a non-zero momentum, causing acollinearity of the photons produced in the annihilation, as shown in (d).

to locate a LOR the position of energy deposition is projected to the entrance surface of the detector or to some fixed depth. When photons hit the detectors at larger angles of incidence this projection causes a mispositioning of the LOR towards the centre of the scanner, as shown in Fig. 1.7 (a). Considering a ring of detectors, at the centre of the FOV the photons will enter the detector perpendicular to its surface [75]. However, photons emitted further away from the centre of the FOV hit the detectors at different angles of incidence. As illustrated in Fig. 1.7 (b) these photons may travel across the first detector and then penetrate into the adjacent detector before being stopped. Furthermore, photons could scatter within the detector before depositing all their energy in an adjacent detector element. Both effects result in a mispositioning of the LOR.

Positron Range A second effect that degrades spatial resolution is positron range, which is defined as the distance between the location where the positron is emitted and the location where the annihilation takes place. The positron range is expected to depend on the energy of the original particle and on the characteristics of the tissue where the decay takes place. The maximum energy of positrons emitted by radionuclides standardly used in nuclear medicine is in the range of 0.5 to 5 MeV. According to Levin and Hoffman [76] the FWHM of the positron range is 0.10 mm for ^{18}F in water-equivalent tissues. As illustrated in Fig. 1.7 (c), the positron range effect causes a mispositioning of the LOR which in turn results in a blurring effect on the final PET images. The direction of positron movement is isotropic in a uniform medium, resulting in a corresponding isotropic point spread function (PSF). The positron range distribution could be deconvolved from the PET images, but due to the limited number of counts usually recorded in a PET scan this deconvolution would lead to noise amplification. More commonly, positron range distribution is incorporated in iterative reconstruction algorithms.

Photon Acollinearity Another factor that further degrades the spatial resolution is photon acollinearity, which originates from the fact that positron and electron are not at rest when they annihilate. Due to the small momentum of the positron - electron pair, the two gamma rays are not emitted at exactly 180° but with an angular distribution with FWHM $\sim 0.5°$ [63]. This results in an error in the location of the LOR, as shown in Fig. 1.7 (d). The blurring effect due to photon acollinearity can be estimated as:

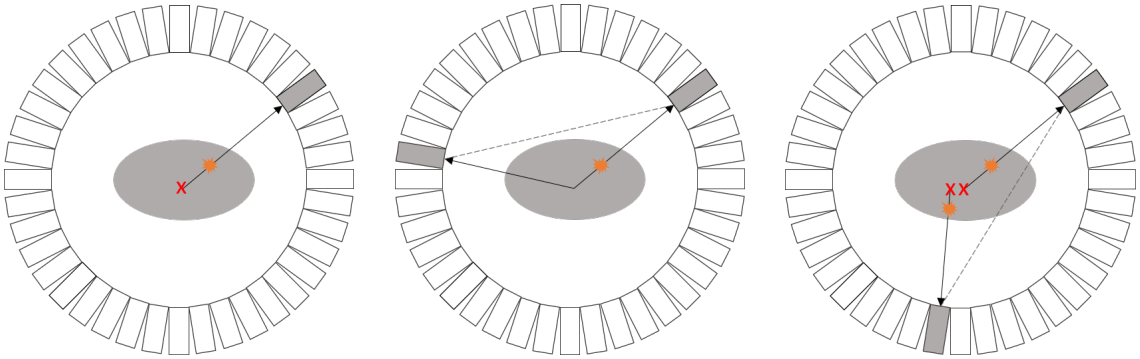
$$\Delta_{acoll} = 0.0022 \cdot D, \tag{1.6}$$

where D is the PET scanner diameter.

1.3.6 Data Corrections for Quantitative PET

One of the main advantages of PET imaging is its ability to provide quantitative measurements of radiotracer concentration in the body *in vivo*. Ideally only true coincident events, that is pairs of back-to-back 511 keV photons generated by a positron annihilation, should be detected by the scanner. However, in reality many "undesired" events are acquired together with the true coincident events, resulting in noisy PET images. Furthermore, limitations in the detector efficiency and in the electronics further degrade the final image quality. Corrections need to be applied to balance for these effects, in order to obtain quantitative PET images in which the voxel intensities are proportional to the true radioactivity concentrations. In the following paragraphs the sources of image degradation are presented.

Photon Attenuation When 511 keV photons are emitted in the body, they interact with the surrounding tissues mainly through Compton scatter. In this process the photon interacts with an electron and is deflected from its original trajectory. This deflection causes a loss in the photon's energy, which



(a) Attenuation (b) Scatter (c) Random

Figure 1.8: Attenuation of photons can prevent annihilation events from being recorded, as one of the two back-to-back photons may be stopped in the imaged object as indicated by the red cross in (a). Another process that degrades PET image quality is photon scatter. As shown in (b) one of the photons emitted in positron-electron annihilation may scatter in the object, thus changing its trajectory. As a result, an incorrect LOR is assigned to the photon pair. An incorrect LOR assignment may also result from random events, which happen when two photons emitted in different annihilations are detected within the allowed timing and energy windows. As shown in (c), these two photons can be incorrectly recorded as one annihilation event.

is transferred to the recoil electron. Each material is characterised by a linear attenuation coefficient μ $[m^{-1}]$ which measures the fraction of photons that are attenuated in that material per unit distance. Linear attenuation coefficients for 511 keV photons are 0.096 cm^{-1} for water and 0.172 cm^{-1} for bone [77]. The photons produced in inner regions of the body have to travel through tissue for longer distances, thus having a larger probability of being attenuated than the photon produced near the skin of the patient. This effect can be corrected for by creating a map of the attenuation coefficients for the whole FOV, called attenuation maps, which can be obtained with different methods [78–81].

Scatter Most scatter events arise from Compton interactions between photons and electrons, in which scattered photons are deflected from their original trajectory and lose some of their original energy. At 511 keV most Compton

events are characterised by forward scatter, in which a small amount of energy is lost and the photons are scattered at angles which are less than 60°. The finite energy resolution of scintillators makes it difficult to discriminate these scatter events from true events. It is even more difficult to make this distinction based on the photon energy, because most 511 keV photons only deposit part of their energy in the detector. In order to detect higher number of true events, large acceptance energy windows between 350 and 650 keV are often used in PET scanners. This in turn causes a larger number of scatter events to be detected. The fraction of scatter events can range from 15% to 40% [67]. Scatter correction methods include empirical approaches, methods based on two or more energy windows, convolution methods and simulations to model scatter distribution [82].

Random Coincidences Random coincidences can happen when two photons belonging to two separate events are detected within the coincidence timing window and recorded as a single annihilation event. Random events usually cause an approximately uniform background in the measured data. The rate of random coincidences is proportional to the square of the activity in the FOV. This effect can be corrected for before PET image reconstruction [83] or during image reconstruction [84]. Both approaches require an estimate of the number of random events which occur along each LOR [85].

Dead Time and Pile-up In a perfect system, the count rate is expected to increase linearly with increasing activities in the FOV. However, the detectors and the connected electronics are affected by a dead time as the activity increases. Correction for this effect includes modelling the dead time of the system at different count rates. At high count rates two photons belonging to different annihilations may deposit an amount of energy that falls within the accepted energy window. This leads to a mispositioning of the LOR and/or

to loss of counts, as two photons are detected as one.

Normalisation In standard PET scanners the detector ring is composed of 10000-20000 individual detector elements. In practice, there are many factors that cause differences in the sensitivity of different detector pairs. For example, detector elements may have different efficiencies based on physical variations in the crystals or different photomultiplier gains. The geometry of the PET scanner causes a systematic variation of the detector sensitivity, as the solid angle coverage differs for detectors located at different positions. Furthermore, the gaps located between detector elements contribute to a further reduction of the sensitivity. The correction for these effects is called normalisation and it results in a series of normalisation coefficients which are applied to each LOR before or during reconstruction.

1.3.7 Image Reconstruction

The aim of image reconstruction is to generate quantitatively accurate images from the data acquired by the detectors. The events recorded in a PET scan can be stored as raw data, that is as single photon detections relative to individual detector elements, together with a time stamp that indicates when each event was detected. In modern scanners data are usually stored in list-mode format, by adding to the raw data information about the coordinate of the LOR along which the event was detected and some information about timing and energy of the detected photon pair. Alternatively, PET data can be stored as sinograms. As shown in Fig. 1.9, an LOR can be parametrised using a radial and an angular variable. The radial variable r measures the distance between the LOR and the centre of the detector ring, whereas the angular variable θ specifies the orientation of the LOR. Each event can then be stored by adding a count to the corresponding pixel in the (r, θ) space. The value

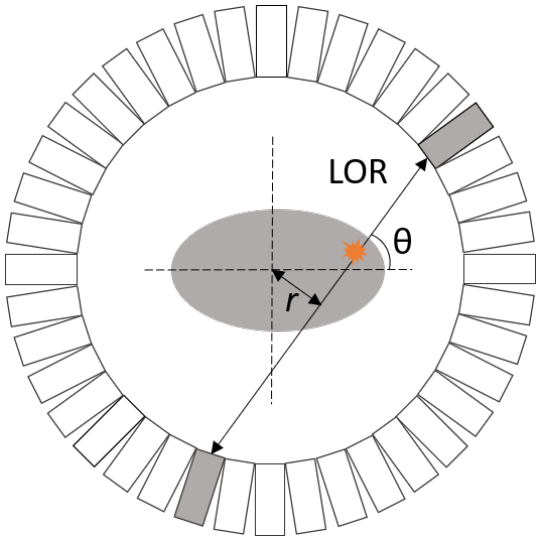


Figure 1.9: Parametrisation of a LOR using a radial variable r measuring the distance between the LOR and the centre of the detector ring, and an angular variable θ that specifies the orientation of the LOR. This parametrisation allows the storage of PET data as sinograms. This 2D parametrisation can be extended to 3D.

stored in each sinogram pixel corresponds to the number of events acquired by the detector pair that is associated to a specific LOR. If all the sinogram elements corresponding to emission from a single point source are plotted, the resulting graph in the (r, θ) space is a sine wave turned on its side, thus the name 'sinogram' [86]. The number of counts detected by a detector pair is proportional to the integrated activity (i.e. line integral) along the LOR joining the two detectors.

Backprojection is a basic algorithm that is commonly incorporated into many reconstruction methods. An estimate of the original source distribution can be obtained from sinogram data through the pixel-driven backpropagation algorithm. The first step of this method is the definition an image matrix, usually made of 128×128 pixels. For an image pixel (x, y) and a projection angle θ it is possible to calculate the sinogram coordinate r that will contribute to that pixel using the equation:

$$r = x\ cos(\theta) + y\ sin(\theta). \tag{1.7}$$

The measured counts in the sinogram corresponding to the calculated r are added to the (x, y) pixel in the image matrix. This process is repeated over all the projection angles for each image pixel. Mathematically, the backprojected image $a'(x, y)$ is calculated using the expression:

$$a'(x, y) = \frac{1}{N} \sum_{n=1}^{N} s(r, \theta_n), \tag{1.8}$$

where N is the number of valid projection angles and $s(r, \theta_n)$ is the number of counts in the sinogram element at angle θ_n and radial offset r. This simple backprojection algorithm generates blurred representations of the original objects because the backprojection maps the measured counts back into the image by equally distributing them along the LORs. The blurring is proportional to 1/r as the sampling frequency decreases as 1/r, where r is the distance from the source. The backprojected image can be mathematically described as:

$$a'(x, y) = a(x, y) \circledast \frac{1}{r}, \tag{1.9}$$

where $\circledast$ represents a convolution operation. The weighting term 1/r amplifies low frequencies and attenuates high frequencies. The blurring effect can be minimised by applying a filter to the acquired data, before backprojecting them to produce an image. This method is called filtered backprojection and it is based on the Fourier slice theorem. According to the Fourier theorem, the measured data in the sinogram are related to the activity distribution in an object through the Fourier transform. The Fourier transform (FT) allows the conversion of data from the spatial domain to the frequency domain. The Fourier transform of a function $f(x)$ can be expressed as:

$$F(\nu) = FT\,[f(x)], \tag{1.10}$$

where ν represents the spatial frequencies. The Fourier slice theorem states that the one-dimensional (1D) Fourier transform of a projection at angle θ (one row in the sinogram) is equal to the 2D Fourier transform of the image evaluated along a radial profile at angle θ with respect to the x-axis [87]. This can be expressed as:

$$S(\nu_r, \theta) = A(\nu_x, \nu_y)|_{\nu_x = \nu_r cos\theta\,, \nu_y = \nu_r sin\theta}, \tag{1.11}$$

where $S(\nu_r, \theta)$ is the 1D Fourier transform of $s(r, \theta)$ with respect to r, and $A(\nu_x, nu_y)$ is the 2D Fourier transform of the radioactivity distribution $a(x, y)$. This theorem enables the direct Fourier reconstruction of a cross-sectional image $a(x, y)$. To do so, we take the Fourier transform of the first sinogram row. Then, these results need to be interpolated into a 2D rectangular grid and the inverse Fourier transform can be applied. After applying this process to all sinogram rows the inverse 2D Fourier transform can be used to form the image $a(x, y)$. The reconstructed image represents the original radioactivity distribution in the imaged object, but its accuracy is hampered by the statistical noise in the measured data by the resolution and sampling limitations of the data. Eq. 1.11 can be rewritten in the spatial domain as:

$$a(x, y) = \frac{1}{N} \sum_{n=1}^{N} s^*(r, \theta_n), \tag{1.12}$$

where s^* represents the sinogram data modified applying a filter function in the frequency domain. The process described by Eq. 1.12 is called filtered backprojection. The ramp filter used in filtered backprojection amplifies the higher spatial frequencies, thus reversing the effects of the 1/r blurring. Filtered backbrojection is practically implemented by taking the 1D Fourier

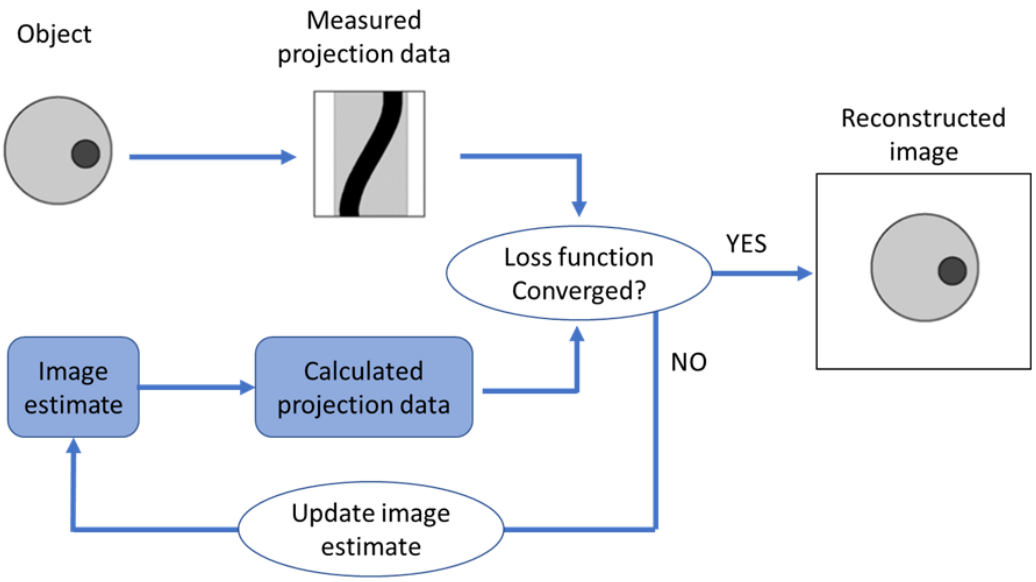


Figure 1.10: Illustration of the iterative image reconstruction method. An initial estimation of the image distribution is made at first. Then, the projection data are calculated from the estimated data through forward projection or through more sophisticated models of the imaging system. The projection data can be compared to the measured projection data, through an objective function. Based on the differences measured by the objective function, the initial guess can be updated, and the whole process is repeated until the estimated data converge to the measured data. This image has been adapted from [87].

transform of the first projection angle in the sinogram, then multiplying this by the filter function. Subsequently, the inverse Fourier transform is calculated and the filtered projection is backprojected. This process is then repeated for all angles. Variations of the reconstruction filter are implemented to optimise the trade- off between signal-to-noise ratio and spatial resolution in the reconstructed image. Until now a 2D reconstruction process was described, but nowadays most scanners are made of multiple rings of detectors. Using these scanners, many contiguous 2D image slices can be reconstructed and stacked to generate a 3D image. Fully 3D reconstruction algorithms have also been developed.

One limitation of filtered backprojection algorithms is that they require adequate sampling. Data should be sampled using a projection sampling interval Δr which is at least half of the highest spatial resolution of the reconstructed image. Inadequate linear or angular sampling causes a reduction in spatial resolution and it can generate artifacts in the reconstructed images. Another assumption that is made in filtered backprojection is that the PET detectors are point detectors. This is not true in practice, and real detectors do not measure data along LORs but rather along a volume joining two detectors. Finally, quality degrading effects such as noise, positron range and random events are not incorporated in analytic reconstruction algorithms.

Iterative reconstruction techniques can be used as an alternative to analytic approaches [88]. A flowchart describing iterative image reconstruction approaches is shown in Fig. 1.10. An initial guess $a^*(x, y)$ of the image distribution is made at first. Then, the projection data are calculated from the initial guess through forward projection, which is the opposite of backprojection, or through more sophisticated models of the imaging system. This set of projection data can be compared to the measured projection data through an objective function. Based on the differences measured by the loss function, the initial guess can be updated and the whole process is repeated until the estimated data converge to the measured data. These methods are more computationally intensive than analytic approaches as they require many iterations. Maximum-likelihood (ML) based methods are among the most commonly used iterative approaches. In statistics, likelihood is a function that measures how well a model fits a sample of data. In image reconstruction, likelihood is maximised when the difference between the measured and the estimated data is minimised. Expectation-maximisation (EM) models aim to maximise the likelihood of Poisson-distributed projection data. To better explain how this algorithm is implemented in image reconstruction, we first

mathematically define the reconstruction problem as:

$$q_i = \sum_j a_{ij}\,\theta_j, \tag{1.13}$$

where θ_j is the activity in a pixel j and q_i is the number of counts in the projection element i. a_{ij} represents an element of the system matrix A, which expresses the probability that gamma rays emitted in pixel j will be detected in the projection bin i. The system matrix describes the relation between the image space and the projection space and it can be factorised into different components that represent the effects that deteriorate PET image quality such as detector geometry and sensitivity, positron range and blurring effects. Using the ML-EM method, the activity of a PET image pixel at iteration $k+1$ is calculated based on the estimated pixel intensity at iteration k. This is expressed mathematically by the following equation:

$$\theta_j^{k+1} = \frac{\theta_j^k}{\sum_i a_{ij}} \sum_i a_{ij} \left(\frac{m_i}{\sum_n a_{in}\theta_n^k} \right), \tag{1.14}$$

where θ_j^k is the j^{th} image pixel at the iteration k, m_i is the number of radioactive emissions measured in the projection bin i, and a_{ij} is the contribution of the j^{th} image pixel to the i^{th} projection bin. The summation over the index n is the projector, and the summation over the index i is the back-projector. An advantage of iterative reconstruction is that it models more accurately the emission and detection processes, taking into account physical effects such as noise, photon scatter and photon attenuation. Secondly, in iterative approaches missing data result in lower quality images rather than in striking artifacts as is the case for filtered backprojection. Because there is no embedded filtering operation, iterative algorithms are expected to generate images with better signal-to-noise ratios. One drawback of iterative approaches is the high computational burden. To speed up the computation

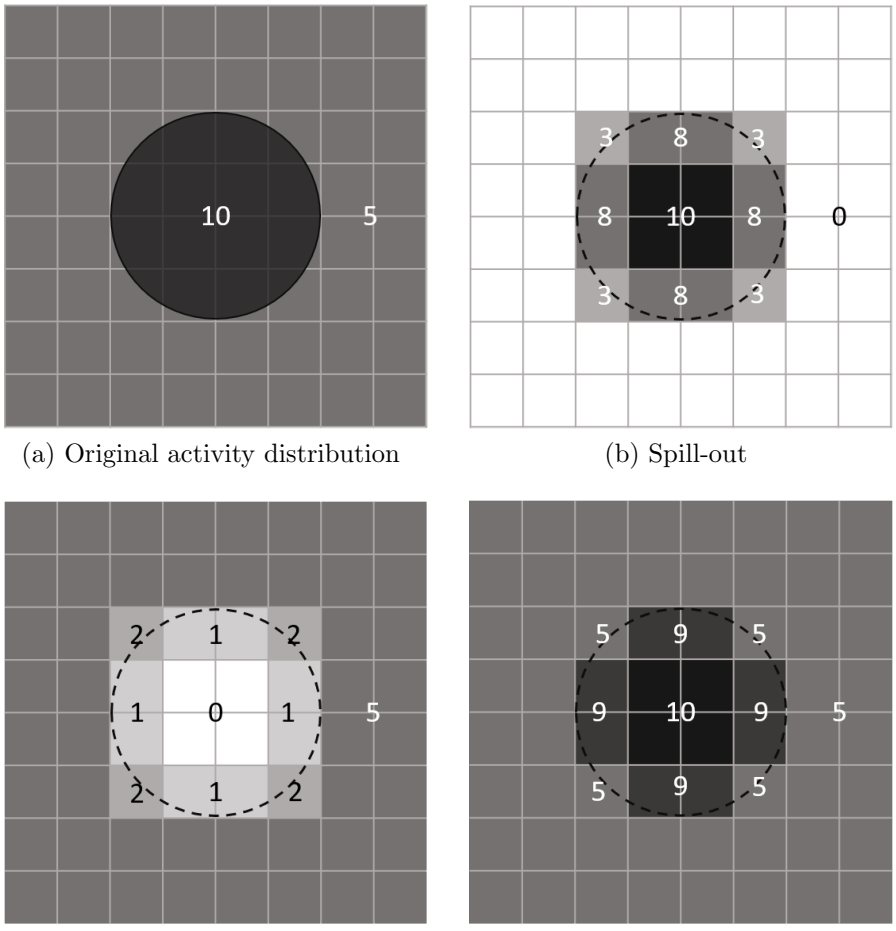
time, in ordered subsets expectation maximisation (OSEM) algorithm only a subset of all the projection angles are used in any iteration. To further improve PET image quality a model of the system resolution (see section 1.3.5) can be integrated in the reconstruction algorithm. These models have been developed to describe the effects that degrade PET image quality with the aim to reduce the partial volume effect and recover high frequency signals. The PSF model can result from Monte Carlo or analytical simulations, or from point source measurements. Monte Carlo simulations require long computational times and a detailed knowledge of the imaging system. On the other hand, technical difficulties are often encountered when performing point source measurements. The PSF model can be applied in the image space or in the projection space. Image based methods are only approximating the resolution degrading effects, so they are expected to be less accurate. Alternatively, the PSF information can be incorporated into the system matrix and thus applied in the sinogram space. Most PSF models focus on the description of the blurring effects which include crystal penetration and inter-crystal scattering [89]. Resolution modelling typically results in qualitatively enhanced images where small lesions and narrow structures are characterised by a better contrast. Although resolution modelling can significantly improve image resolution and contrast, some high frequency signals can never be recovered. The frequency spectrum abruptly drops to zero for frequencies that exceed the scanner cut-off. This cut-off generates Gibbs ringing artifacts near any discontinuity in the PET image. Another effect of PSF-based reconstruction algorithms is an over-enhancement of the edges in the reconstructed images. Furthermore, resolution modelling can significantly modify the noise structure of an image [89] thus affecting quantitative measurements especially for small lesions. Iterative reconstruction approaches are often limited by the high number of iterations that are needed to obtain significant recovery, thus by their high computational

burden which can be impractical in a clinical setting.

1.3.8 Uptake Quantification in Oncology

Quantification of uptake in PET images is hampered by noise, by partial volume effect (PVE) and by the limited spatial resolution of the system which results in blurred images. As described in the previous sections, noise arises primarily from the statistical fluctuations of the number of detected counts. Additional sources of noise are the detection of random and scatter events which are wrongly classified as coincidence events (see section 1.3.6) and the reconstruction algorithms [90, 91]. The PVE is related to the limited spatial resolution of PET systems, and it causes small structures in the reconstructed PET images to be associated with lower activity than their original one [92]. The finite spatial resolution of the system causes image blurring, so small hot lesions appear larger and with lower activity in the reconstructed PET images. Some of the activity of the lesion "spills out" in the adjacent voxels, as shown in Fig. 1.11 (b). Equivalently, tumours characterised by a lower activity may seem more aggressive due to spill in of activity from the surrounding higher activity tissue as illustrated in as shown in Fig. 1.11 (c). Another phenomenon that contributes to the PVE is image sampling. To generate the PET images, the events detected in the PET scanner are stored in different bins, which correspond to voxels in the final PET images. The boundaries of each voxel do not match the real contours of the radiotracer distribution. Many voxels include different types of tissue, and the intensity assigned to each voxel is an average of the intensities of all the tissues included in that voxel. PVE particularly affects the measurement of uptake for small lesions, with diameter smaller than 2 or 3 times the FWHM of the system resolution [92, 93]. Various methods have been proposed for partial volume correction, operating on the reconstructed images or incorporating the correction in the reconstruction

algorithm [6, 94].



(a) Original activity distribution (b) Spill-out

(c) Spill-in (d) Measured activity distribution

Figure 1.11: Illustration of spill-in and spill-out effects. Due to the limited spatial resolution of the system, in the reconstructed PET images some activity originally belonging to small hot lesions may "spill out" in the adjacent voxels as shown in (b). Equivalently, tumours characterised by a lower activity suffer from spill-in of activity from the surrounding higher activity tissue as shown in (c). This figure was originally published by Soret et al. [92].

In clinical practice, qualitative visual assessment is the standard technique for image interpretation. However, this approach is affected by high inter- and intra- observer variability. Semi-quantitative measurements of radiotracer uptake are often provided in order to support diagnosis and to assess response to treatment [95]. One of the most commonly used semi-quantitative metrics

is the standardised uptake value (SUV), which is defined as the ratio between radioactivity concentration in a region of interest (ROI) or volume of interest (VOI) at a given time, divided by the average radioactivity concentration in the whole body as shown in Eq. 1.15.

$$SUV = \frac{Radioactivity\ concetration_{ROI}\ (kBq/ml)}{Injected\,activity\,(MBq)/\ Body\,weight\,(kg)} \tag{1.15}$$

The measured activity in a given ROI can be normalized to the average radioactivity in the body based on of the patient's body weight (SUV_{bw}), lean body mass (SUV_{lbm} or SUL), or body surface area (SUV_{bsa}). Independent of the normalisation method, different SUV metrics can be defined. The most commonly used are SUV_{max} and SUV_{mean}, which are obtained by considering respectively the maximum or the mean of all the voxels in the VOI. SUV_{max} has the advantage of being more reproducible as it does not require an accurate VOI definition, but it is highly affected by the noise levels of the PET image. SUV_{mean} is less dependent on image noise, but it depends on the definition of the VOI which is prone to inter- and intra-reader variability. SUV_{peak} has been introduced to combine the lack of need for accurate VOI definition of SUV_{max} and the noise reduction of SUV_{mean}. SUV_{peak} measures the highest average tumour radioactivity concentration, calculated using a given VOI which is usually spherical. Although guidelines have been developed for the PET response criteria of solid tumours (PERCIST) [96], there is no unique definition of the size, shape and location of the VOI used to measure the SUV_{peak} value. SUV_{peak} has some limitations when uptake in small lesions is measured, especially when using a VOI bigger than the lesion itself. Although commonly used, SUV measurements can be significantly affected by biological and physical factors and by reconstruction parameters [97].

Other volumetric measurements that are used to characterise tumours are metabolic tumour volume and total lesion glycolysis. Metabolic tumour vol-

ume (MTV) represents the volume of the tumour which is metabolically active. Total lesion glycolysis (TLG) can be obtained by multiplying the total tumour SUV_{mean} by the metabolic tumour volume and it gives information on both the tumour size and the radiotracer uptake [98]. As there is no consensus on how to best segment tumours from PET images, MTV and TLG are not practically used in most clinical settings [99]. There is need for new methods that provide more accurate, precise and reproducible information from PET images.

1.4 Deep Learning in Medical Imaging

In recent years, deep learning (DL) algorithms have been widely applied in medical imaging to perform a variety of tasks. DL approaches are based on neural networks that can analyse data at different levels of abstraction. Through the training process, DL models can learn the features that best fit the data they are presented with. The final aim is to obtain generalisable models that can yield accurate predictions when presented with an unseen set of data. DL has the great advantage of offering time-efficient solutions to a variety of tasks. Furthermore, the automated data analysis may lead to more reproducible results by reducing inter- and intra- observer variations. DL algorithms can process large datasets in a short time, which may not be feasible for a human interpreter. On the other hand, DL algorithms often require large amounts of data for training, which may not be available in medical imaging. The use of large training datasets is often computationally intensive, thus requiring access to GPU servers. Many DL approaches are optimised for a specific task or on a specific dataset, and extensive external validation across different centres, scanners, populations is needed to test the applicability of DL techniques in clinical practice. Another factor that hampers the wide use of DL algorithms

in clinical routine is that many models work as a 'black box' and there is little understanding of the decision-making process of the machine [100, 101]. In recent years, there has been an increasing effort in developing understandable AI models, through the use of feature visualisation or saliency maps, that can give insights into the AI black box.

DL models can be split into two broad categories: supervised and unsupervised. Unsupervised algorithms can learn useful representations from unlabelled data. In the field of medical imaging, unsupervised DL algorithm have been developed to perform tasks such as denoising [102–104], data clustering [105] and classification [106]. In supervised learning, networks are trained on labelled datasets to learn how to map inputs to outputs. Supervised learning approaches can be used to perform classification tasks when the labels are discrete variables (classes), or regression tasks when labels are a set of continuous values. A review of some applications of supervised learning in medical imaging is presented in section 1.4.2.

In between these two categories there is semi-supervised learning, in which the networks are trained with both labelled and unlabelled data. This approach is particularly useful in medical imaging where it may be too time consuming and costly for experts to manually label an entire dataset. Generative adversarial networks (GANs) have been extensively applied to medical imaging to perform tasks such as image segmentation, reconstruction, detection, denoising, registration and classification [107].

Finally, another approach that may be used to train networks is reinforcement learning, in which the machine learns in an interactive environment by trial and error. In reinforcement learning, rewards and punishments are associated to positive and negative decisions and used as feedback to optimise the machine's performance. Reinforcement learning has been applied to various healthcare tasks characterised by noisy, multi-dimensional and incomplete

data [108].

Deep Learning in Oncology DL algorithms have been applied to oncology to solve a range of tasks. For example, automated detection and classification lesions can help diagnosis by improving the detection accuracy or decreasing the reading time of human experts. Different network architectures have been successfully used for automated lung nodule detection and classification reaching high accuracies comparable to the ones obtained by radiologists [109–112]. DL applications to chest radiography gave encouraging results towards lung cancer screening [113]. DL has been applied to breast cancer to detect and classify breast malignancies [114–116] and skin lesions, demonstrating that the network was capable of classifying skin cancer with a level of competence comparable to dermatologists [117]. Another task that can be solved with DL is automated lesion segmentation, which is the process of assigning a label to every tumour voxel to automatically identify its location and boundaries. This can be particularly helpful to quantify in a reproducible way clinical parameters such as tumour volume and shape. For example, DL algorithms have been successfully used for segmentation of brain tumours [118], skin lesions [119] and prostate lesions [120]. In prostate studies, DL can be used to automatically grade cancer lesions [121, 122]. In the next section some basic concepts about deep learning architecture, training and optimisation are introduced.

1.4.1 Neural Networks: Components, Training and Testing

During supervised training, a neural network receives as an input a training dataset composed of two sets of images, often called input image and label. The labels correspond to the final target or 'ground truth' images. Through the training process, DL models learn the features that best fit the input data they

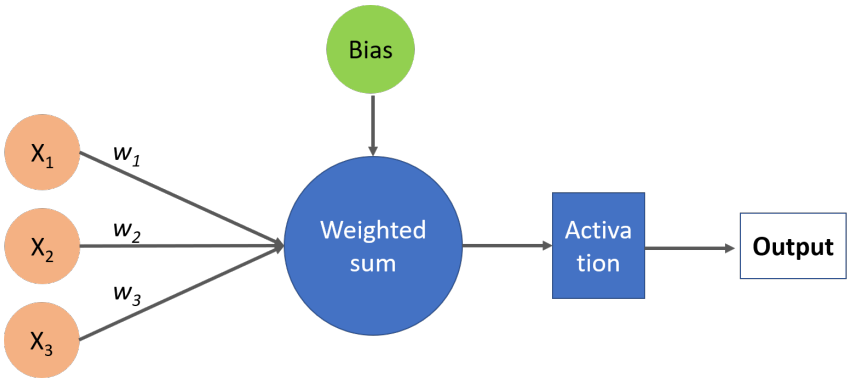


Figure 1.12: Illustration of a neuron. In a neural network, neurons receive the input data and sum them to produce an output. Each neuron further transforms the input by adding a bias.

are presented with. To illustrate how the training process work, we start from the description of the network's components. A neural network is composed of a series of layers. Each layer is formed of a typically very large number of neurons, which receive the input data and sum them to produce an output. As shown in Fig. 1.12, each neuron transforms the input by applying weights and by adding a bias. The vectors of weights and the bias are called filters (or kernels) and they are usually randomly initialised when training starts.

Activation functions At each forward pass, the result generated by each filter is passed into an activation function that transforms the data by applying a linear or non-linear operation to the input data before it is sent to the next layer. In general, neural networks are more easily optimised when their behaviour is linear or close to linear. Linear activation allows for a faster convergence and it avoids the vanishing gradient problem which affects other activation functions such as tanh or sigmoid. These two functions are shown in Fig. 1.13 and they are mathematically described by the following equations:

$$Sigmoid : y = \frac{1}{1 + e^{-x}} \tag{1.16}$$

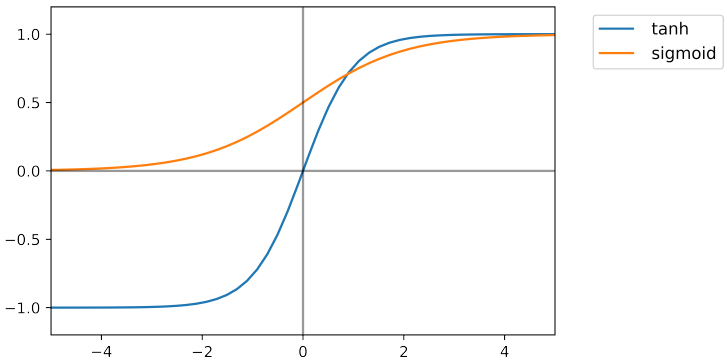


Figure 1.13: Plots of the hyperbolic tangent and sigmoid function.

$$Tanh : y = \frac{2}{1 + e^{-2x}} - 1 \tag{1.17}$$

The sigmoid function is defined between 0 and 1, so it is especially used for models where a probability is yielded as an output. The tanh function is defined between -1 and 1 and it is mainly used for binary classification. The gradient of these functions rapidly decreases as x diverges from 0, thus preventing the weights from changing their value during training. This is called the 'vanishing gradient' problem. An alternative activation function that partly solves the vanishing gradient problem is the rectified linear unit (ReLU) activation function. This function, mathematically defined as $y = max(0, x)$, returns the input value x if x is positive, otherwise it returns 0. ReLU is less computationally expensive than tanh and sigmoid because it involves simpler mathematical operations. When using ReLU activation only a few neurons are activated at a time making the computation more efficient.

Network training The simplest neural network architectures are composed of an input layer, a hidden layer and an output layer. Deep networks are formed of several hidden layers. At each 'pass' all neurons in a neural network perform the operations described above and one such 'pass' is known as forward propagation. The full output of each layer is generated by stacking the

activation maps for all filters. During training the network iteratively adjusts the biases and weights through backpropagation, with the aim of minimising the difference between the estimates provided by the network at each forward pass and the labels. The function used to evaluate the difference between these two images is called loss function. At the output layer, the network calculates the total error, which measures the total difference between actual and predicted values for all data points. Then the network calculates the derivative of the loss function with respect to the weights, henceforth referred to as the gradient. The gradient can be used to minimise the loss function as it gives information on how the weights can be changed to obtain a small improvement in the predicted image. The weights can be updated using the gradient itself or a factor of it. This factor is known as the learning rate and it determines the size of the step that are used to change the weights. A too small learning rate results in very long training, with the risk of the algorithm getting stuck in local minima. A too large learning rate may result in completely missing the global minimum. During supervised training, convergence is achieved when all the elements of the gradient are close to zero. The update process starts from the output layer and it continues until the first layer is reached.

Optimisers Nowadays, most deep learning algorithms use stochastic gradient descent (SGD) to optimise the parameters. This algorithm allows optimisation to be performed faster when using large datasets. Other optimisation techniques have been developed to accelerate training. One of these is momentum, an algorithm in which the learning rate is tuned taking into account past gradients as well. The step size is larger when many subsequent gradients point in the same direction, and it is smaller when subsequent gradients point in different directions. The learning rate has a significant impact on the results yielded by the network. Optimisers such as AdaGrad, RMSprop and Adam have been developed to automatically adapt the learning rates of differ-

ent model parameters throughout the training process. AdaGrad is one of the first algorithms used for training optimisation. It adapts the learning rates of all the training parameters by scaling them inversely proportional to the square root of the sum of all their historical squared values [123]. This means that the training parameters characterised by a large partial derivative of the loss function are associated with a rapid decrease of the learning rate. Considering all the squared gradients from the beginning of training can result in excessive decrease of the learning rate, thus negatively impacting the network's performance. Root mean square error propagation (RMSprop) algorithm is based on the AdaGrad algorithm, but in this case the learning rate is divided by an exponentially decaying average of squared gradients. With this modification, the information stored at the beginning of training is discarded, allowing for faster convergence. RMSProp is currently one of the preferred optimisation methods, and it is standardly used in deep learning applications [123]. Finally, adaptive moment estimation (Adam) [124] is an optimisation algorithm that combines RMSprop and momentum. Similar to RMSprop, an exponentially decaying average of past squared gradients is stored to rescale the learning rates, together with an exponentially decaying average of past gradients similar to momentum. Adam has been proven fairly robust to the choice of hyperparameters, and in many cases the default parameters are sufficient to obtain a good performance. Although many optimisers have been developed through the years, a best candidate has not been found yet. The choice of optimiser is often based on the characteristics of the training dataset.

Validation and testing As explained above, during the training process the model is fitted to the input data so that its predicted results are as close as possible to the labels. During training, the fitted model is applied to a validation dataset which helps to determine the optimal design of the deep learning model. Finally, a test data set composed of data that have not been

used for training or validation is used to provide an unbiased evaluation of a final model fit on the training data set.

1.4.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) are among the most popular algorithms for medical imaging applications [125]. The first real world application of CNNs was LeNet, an algorithm that could automatically perform hand-written digit recognition [126]. In 2012, the advancements in technology and the availability of large labelled datasets such as ImageNet, enabled researchers to develop complex CNNs that could outperform other algorithms that had been used within the ImageNet challenge to perform image classification [127]. Strategies such as transfer learning [128] and data augmentation [129] have been developed to mitigate the amount of labelled data required to train a CNN. Multi-stream CNN architectures have been proposed to make use of complementary information from multiple imaging modalities [130] or from multi-planar images [131]. As medical images can have a large size, which may exceed the GPU memory, patch-based approaches in which CNNs are trained on small regions cropped from the original images (patches) have been proposed to lower the amount of memory needed for training. CNNs are able to learn data in multiple levels of abstraction, by using three basic types of layers: convolution, pooling and fully connected layers. In convolutional layers, the filters are convolved with the input images. Each filter is slid across the width, height and depth of the input image and the dot products between the input and filter are computed at every spatial position. Each neuron in the resulting activation map is connected to a small region of the input volume, with dimensions equal to the filter size [132]. Pooling layers perform non-linear downsampling of the input data. Pooling operations reduce the output of multiple neurons to a single neuron in the next layer. A commonly used technique

is max pooling, in which the maximum value in the region seen by each filter is taken to generate the output matrix. Pooling helps to make the network more robust to small translations in the input images. This is useful in some image recognition tasks, but when a task relies on preserving precise spatial information pooling operations may lead to an increased training error [123]. For this reason, pooling layers are not used in our proposed algorithm. In fully connected layers all neurons in one layer to be connected to every neuron in the following layer. They are typically located at the end of CNNs. Recent work has shown that fully connected layers do not always improve the performance of the network, and that in some cases better results can be estimated from the last convolutional layer directly [133]. In addition to the layers presented, batch normalisation and dropout layers can be used. Batch normalisation enables the standardisation of the mean and variance of each input to a layer. Batch normalisation results in a reduction of the number of iterations over the dataset (also called epochs) that are needed to train deep networks. Dropout layers helps prevent overfitting by randomly deactivating a given number of neurons on a particular layer. For example, a dropout rate of 50 % means that 50% of the neurons in a layer will be deactivated. Dropout is only used during training, and it improves the generalisation performance of the model by making the weights learnt in different nodes more independent.

Examples of CNN Architectures applied to Oncology In 2015 a new architecture called UNet was introduced [134]. UNets are characterised by an encoder-decoder structure, made of a series of downsampling and upsampling layers. The downsampling and upsampling operations allow the processing of medical images in one go, instead of splitting them in patches. Skip connections can be added between opposing convolution and deconvolution layers to concatenate features from the downsampling to the upsampling paths. UNets are mainly used for segmentation tasks, and their output is usually a binary

segmentation map. Some successful applications of UNets in oncology include brain tumour and intra-tumour segmentation [135], pulmonary nodule segmentation [136] and breast cancer segmentation [137].

Another architecture that has been developed is ResNet [138], which is similar to traditional CNNs with the addition of skip connections between subsequent layers. Skip connections make it possible to train much deeper networks. The operation of copying the activations from layer to layer preserve information as data goes through the layers. ResNet has been used for skin lesion analysis [139] and lung nodules detection [140].

Finally, DenseNet [141] is a convolutional neural network in which each layer is connected to every other layer in a feed-forward fashion. DenseNets require fewer parameters than an equivalent traditional CNN, as there is no need to learn redundant feature maps. Furthermore, the DenseNet structure strengthens feature propagation and encourages feature reuse. DenseNet has been successfully applied to breast cancer segmentation [142], gastric cancer recognition [143] and response to therapy prediction in lung cancer [144].

1.5 Summary

In this chapter different medical imaging modalities were introduced, with focus on MRI and PET. An overview of deep learning models was presented, together with a range of successful applications of supervised deep learning models to oncology. In the next chapter, a deep learning algorithm developed with the aim to improve tumour radiotracer uptake quantification in PET imaging will be introduced.

Chapter 2
Improving Uptake Quantification on Simulated PET Images

The first part of this work is focussed on the implementation of a deep learning approach to improve quantification of tumour radiotracer uptake in PET images. Supervised deep learning methods often need a large amount of labelled data for training. It is difficult to find such datasets in PET imaging, as PET studies often involve a limited number of patients. Furthermore, when large datasets are acquired the labelling procedure performed by experts may be difficult and time consuming. One way to overcome these issues is to make use of simulated data. Although PET data simulations are a valuable tool, they inevitably involve a series of assumptions and approximations that need to be considered. First of all, the true radionuclide distribution within the object imaged in a PET scan is rarely known and very difficult to obtain. This results in a lack of "ground truth" data to be used as gold standard for the assessment of PET data processing techniques. This intrinsic limitation affects not only deep learning-based approaches, but also other PET data processing methods such as image reconstruction algorithms. Secondly, real tumours are characterised by a high degree of inter- and intra- tumour variability, both in terms of biological composition and spatial structure. Most mathematical and computational models of cancer developed in recent years mainly aim to characterise

a few specific aspects of the disease [145], such as tumour growth, metabolism or heterogeneity. No unique model has been found that encompasses the wide range of tumour compositions and characteristics. Biopsy is the clinical standard examination that allows for cancer characterisation and diagnosis. This technique makes use of one or more samples taken across the region of interest and thus only provides limited information on the lesion characteristics.

In this chapter, PET images are simulated with a simple method that aims to reproduce the two main effects that degrade image quality: PVE and noise. The dataset generated with this method is used to train and test a 3D CNN. Different network architectures are compared, and the results are evaluated qualitatively and quantitatively. The impact of training parameters such as the optimiser and the number of epochs is also investigated in this chapter. Afterwards, the 3D CNN is trained on a more complex dataset, in which the simulated PET images are generated with a range of PSFs. This network is finally applied to phantom PET scans.

2.1 Deep Learning Method

The method developed in this thesis is composed of a number of steps which are illustrated in Fig. 2.1. At first, spheres with different sizes and radioactivity patterns are generated. The words radioactivity and activity will henceforth be used interchangeably when referring to simulated or real radioactivity. The simulated spheres are then warped in different ways to obtain less spherical and more tumour-like shapes. The deformed tumour shapes are assigned with a range of radioactivity values. The resulting images are henceforth referred to as "ground truth" images. A set of simulated PET images is then generated by applying a Gaussian filter and by adding noise to the ground truth images. Finally, a 3D CNN is trained on the pairs of ground truth images and simulated

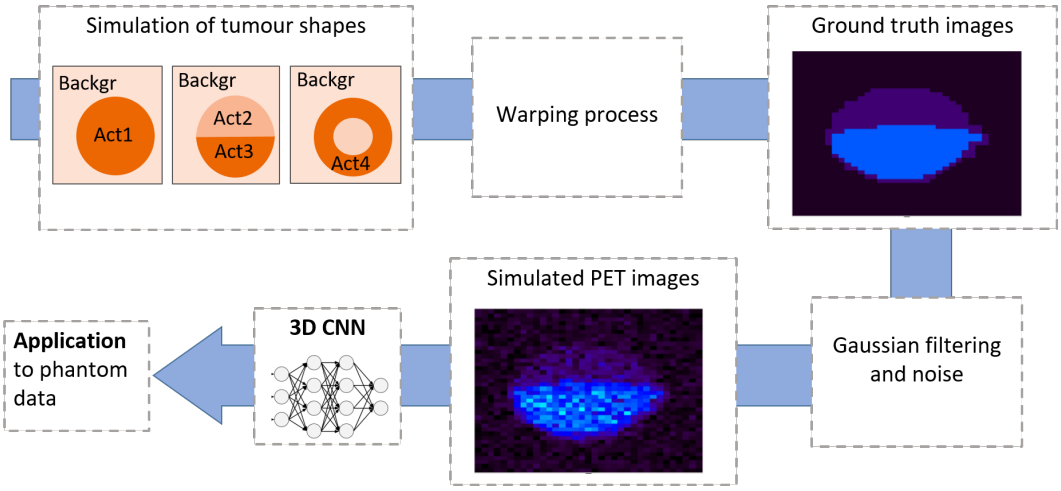


Figure 2.1: Overview of the proposed deep learning approach. At first, spherical shapes are generated with three different activity patterns. These spheres are then warped in different ways to create more tumour-like shapes. The images created with this process are henceforth referred to as ground truth images. To generate the simulated PET images, the ground truth images are filtered with a spatially invariant Gaussian filter, and Gaussian noise is added. A 3D CNN is trained on pairs of ground truth images and simulated PET images, with the aim to recover the ground truth activity distribution. Finally, the CNN trained on simulated data is applied to phantom data.

PET images. When tested on an unseen set of simulated PET images, the network generates a prediction of the corresponding ground truth. A more detailed description of each step of the proposed deep learning approach is provided in the next sections.

2.1.1 Simulation of Training and Testing Datasets

As shown in Fig. 2.1, the first step of the proposed approach involves the simulation of a large number of tumour shapes. The tumours are initially generated as spheres with different sizes, with radius spanning 2 to 12 voxels. These activity patterns simulated in this step can be split into three groups:

- uniform: these spheres are assigned with a uniform activity distribution. This group represents focal tumours composed of one or more tissues

with similar properties.

- bimodal: these spheres are split into halves, each assigned with a different activity with ratio 3:1. This specific ratio was chosen so that the two areas with different tumour activities could be well distinguished visually. This group represents tumours which may have non-homogenous uptake due to different underlying tissue properties.

- hollow: these spheres are characterised by a hollow inner part, assigned with background activity. This pattern mimics tumours with a necrotic core.

The spheres are then warped to generate more tumour-like shapes. The deformation process starts with the definition of a small cube of edge length $(a + b * 2)$. The parameter a controls the local variations, whereas b is used to identify the inner volume of the cube. Random values spanning from -10 to 10 are assigned to the inner portion of the cube, leaving the edges at 0, then the cube is resized to the dimension of the input images. The cube is zoomed using second order spline interpolation, so the randomly assigned values are smoothly interpolated to create a smooth uniform deformation field for each dimension. Each dimension is deformed by a^3 random values and larger a values will add more local variation. The deformed tumour shapes are thresholded to clearly identify the tumour voxels, and subsequently assigned with a range of radioactivity values. The activities assigned to the warped tumours range from 2 to 50 kBq/ml, and the background is assigned with an activity equal to $1/10^{th}$ of the maximum activity. This foreground to background ratio is chosen to replicate the settings used by Pfaehler *et al.* in their phantom study [146], as our goal is to ultimately apply the CNN to the same phantom PET scans. The resulting images are henceforth referred to as "ground truth" images, and they are used as labels to train the network. The first dataset

used to train and test the 3D CNN is composed of 4200 ground truth images, each made of 35×35×40 voxels with voxel size 2.09×2.09×2.03 mm^3. The tumour shapes generated in this dataset are somewhat simplified in that they do not fully represent the wide variety of possible tumour shapes that can be observed in real patients. Analogously, the three activity patterns that have been assigned to the simulated tumours do not represent the highly heterogeneous activity patterns that may characterise real tumours. Nonetheless, the simulated dataset offers sufficient variability to reliably test the proposed deep learning approach. As described in section 1.3.4, the two main effects that degrade PET image quality are partial volume effect and noise. In order to reproduce these effects, in this dataset the simulated PET images are generated by convolving the ground truth images with a Gaussian filter, which results in a blurring of the images similar to that due to the PSF of the system. After blurring with a Gaussian filter, noise is added to the blurry images. The noise generated by the statistical fluctuations in the positron emission process can be described by a Poisson distribution. During PET data acquisition and image reconstruction there are several other sources of noise that further contribute to the degradation of PET image quality. As a result, many post-processing approaches assume that overall the noise in PET images can be characterised by a Gaussian distribution [91]. In this experiment, the noisy PET images are obtained by randomly drawing each voxel value from a Gaussian distribution with standard deviation $\sigma = k \cdot \sqrt{N}$. In this equation, N is the number of counts stored in each voxel in the filtered images, and k is a constant value. The PET images are generated with the same number or voxels and voxel size as the ground truth images. The PET images are generated with a Gaussian filter with FWHM = [4, 4, 4] mm^3, and setting the constant k to 0.2. Three examples of ground truth images and simulated PET images are shown in Fig 2.2. The simplistic method used to generate the simulated PET images does

not incorporate the many effects that degrade the quality of original PET images, but it enables to generate a training dataset in a short amount of time, and to test different network architectures and training parameters in a simplified environment.

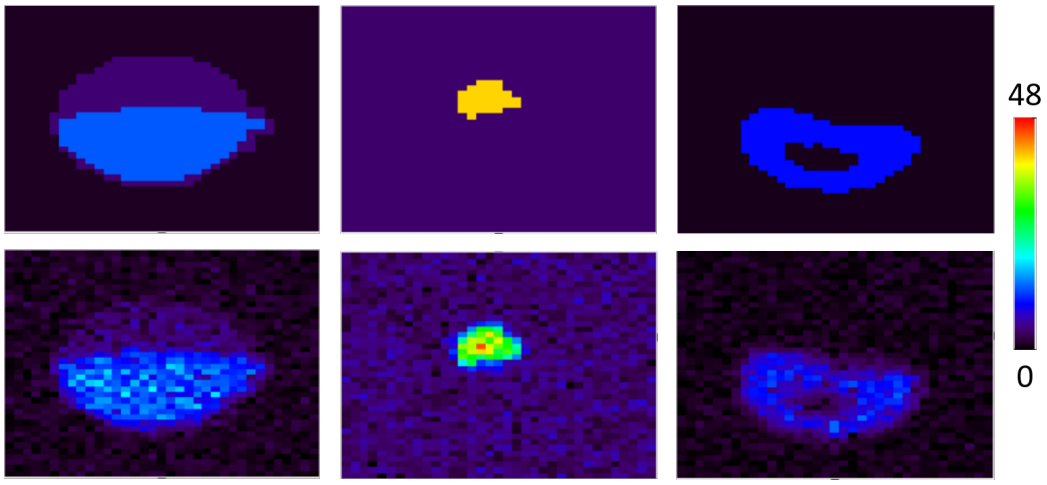


Figure 2.2: Examples of ground truth and simulated PET images. The first column shows transverse views of a tumour split in halves, assigned with different activities with ratio 1:3. In the second and third column images of a uniform tumour and of a hollow tumour are shown. Top row: ground truth images. Bottom row: Corresponding simulated PET images of more blurry and noisy appearance. All figures are normalised to the same colour scale, expressed in kBq/ml.

2.1.2 Image Evaluation Metrics

Different metrics are used in order to assess the performance of the network. At first, visual inspection of a subset of the predicted images allows for a preliminary assessment of image quality. The 3D images are visualised using the open source software AMIDE, a free software tool for multimodality medical image analysis [147]. Using AMIDE, multiple images can be overlaid and 3D contours can be drawn around a VOI by automatically selecting all the voxels that exceed a chosen threshold. The contours of the predicted tumours can be

compared to the contours of the ground truth tumours. AMIDE also allows the measurement of the maximum and mean activity within a VOI. This preliminary assessment is used for each simulation, especially to guide the choice of network architecture. After visual assessment, a quantitative analysis is performed. The maximum tumour activity is measured for all the simulated tumours. The mean and median activities are measured only for the uniform and hollow tumours. The mean and median activity concentrations are not measured for the tumour split in halves as, especially when analysing small tumours, it is difficult to precisely draw the contours of the two tumour regions assigned with different activity concentrations. A more compact way to report these quantitative measurements is by using the recovery coefficients (RC). The maximum recovery coefficient is defined as the ratio between the maximum activity concentration measured in the predicted images or in the simulated PET images, divided by the maximum activity concentration assigned to the ground truth images. Eq. 2.1 and Eq. 2.2 show how the RC are estimated using the predicted images and the simulated PET images. The comparison between pairs of RC values indicates whether the network was able to improve the quantification of the maximum activity. The mean and median recovery coefficients can be measured analogously, using the mean and median tumour activity concentration.

$$RC_{max}\left(pred\right) = \frac{Max_{\,predicted\,image}}{Max_{\,ground\,truth\,image}} \tag{2.1}$$

$$RC_{max}\left(sim.PET\right) = \frac{Max_{\,sim.\,PET\,image}}{Max_{\,ground\,truth\,image}} \tag{2.2}$$

To further assess the similarity between predicted images and ground truth images, the structural similarity index (SSIM) is measured [148]. This metric is commonly used to assess image quality in the context of image reconstruction

or restoration. As shown in Eq. 2.3, three separate components contribute to the similarity measurement: luminance l, contrast c and structure s.

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{2.3}$$

In Eq. 2.3, x and y are two non-negative image signals, which could be for example two image patches. The luminance comparison function $l(x,y)$ depends on the mean intensities of the two image signals μ_x and μ_y. After estimating the luminance, the mean intensity is removed from the initial image signal. At this point, contrast comparison between the two signals is performed by measuring the standard deviations σ_x and σ_y. Finally, each signal is normalised by its standard deviation and the structure comparison is made on the normalised signals. For image quality assessment, it is useful to apply the SSIM index locally rather than globally as image statistics may be spatially variant. Measurements such as μ_x and σ_x are usually computed within a local square window, which is then moved pixel-by-pixel over the entire image. At each step, the local statistics and SSIM are calculated within the local window. In practice, a single overall quality measure of the entire image may be needed so a mean SSIM (MSSIM) can be used. The MSSIM is defined as:

$$MSSIM\,(X,Y) = \frac{1}{M}\sum_{j=1}^{M}(SSIM(x_j, y_j)) \tag{2.4}$$

where X and Y are the reference image and the distorted image, x_j and y_j are the image contents at the j-th local window and M is the number of local windows in the image. In this thesis a conservative approach was adopted for the assessment of the network's predictions. As a result, we decided not to use mean squared error and related metrics (such as the residual sum of squares) as a measurement of image quality, as mean squared error was also used as a loss function to train the neural network.

2.1.3 CNN Architecture

In this chapter, we use different CNN architectures which have some common features. The basic CNN architecture, illustrated in 2.3, is composed of a series of convolutional layers each followed by a batch normalisation layer, except for the final one. The convolutional layers are composed of 32 filters with dimension 3×3×3 voxels, and assigned with ReLU activation. The final layer has one filter and it has a linear activation function. In all the experiments presented in this thesis 65% of the simulated data are used for training, 15% for validation and the remaining 20% for testing. The loss function used for training is mean squared error, a function that is often used in deep learning to perform regression tasks.

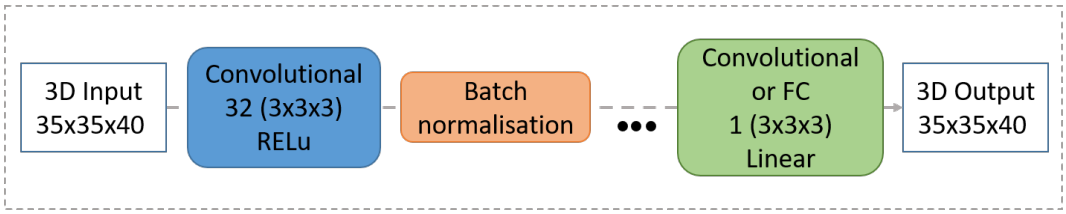


Figure 2.3: Representation of a CNN architecture. The CNNs used in this thesis are composed of a series of convolutional layers each followed by batch normalisation. The convolutional layers are typically made of 32 3x3x3 filters, and they are assigned with RELu activation. The output layer can be a convolutional layer of a fully connected layer, with one filter and linear activation.

2.2 Results

As the performance of a network strongly depends on the number of layers (depth) that compose its architecture [123], networks with different depths are therefore tested to find the architecture that performs best on the simulated dataset. Other parameters that affect the performance of a network, and that are here investigated, are the number of epochs used for training and the

optimiser.

2.2.1 Comparison between CNNs composed of 4 to 8 layers

The first experiments aim to compare the performance of 3D CNNs made of different layers. For simplicity, the depth of the CNN is described in terms of convolutional layers only. In this section, five networks composed of 4 to 8 convolutional layers are trained and tested on the same dataset. Batch normalisation layers are used in all the CNNs tested in the following experiments, unless explicitly stated. The dataset used for training and testing is composed of 4200 pairs of ground truth and simulated PET images. The simulated PET images are generated with the method described in section 2.1.1, using a spatially invariant Gaussian filter with FWHM = [4, 4, 4] mm^3. All the CNNs tested in this section are trained on the same set of non-normalised images for 200 epochs, using the RMSprop optimiser. Fig. 2.4 shows a comparison between the predictions obtained for a representative uniform tumour. The ground truth image is shown in Fig. 2.4 (a) and the corresponding simulated PET is shown in Fig. 2.4 (b). As expected, the simulated PET image is noisier and the tumour shape has blurry edges. All the images predicted by the CNNs are denoised, and the tumours in these images are characterised by sharper edges. Although the activity distribution in the predicted tumours looks more uniform, there are tumour voxels that exhibit slightly higher or lower activity than the ground truth. At this stage errors are expected as the network has not been optimised yet. Visually, the tumours in Fig. 2.4 (d) and Fig. 2.4 (e) appear more similar to the ground truth in terms of tumour shape, and they are characterised by higher activities than the other tumours.

This visual assessment is supported by a quantitative analysis. The maxi-

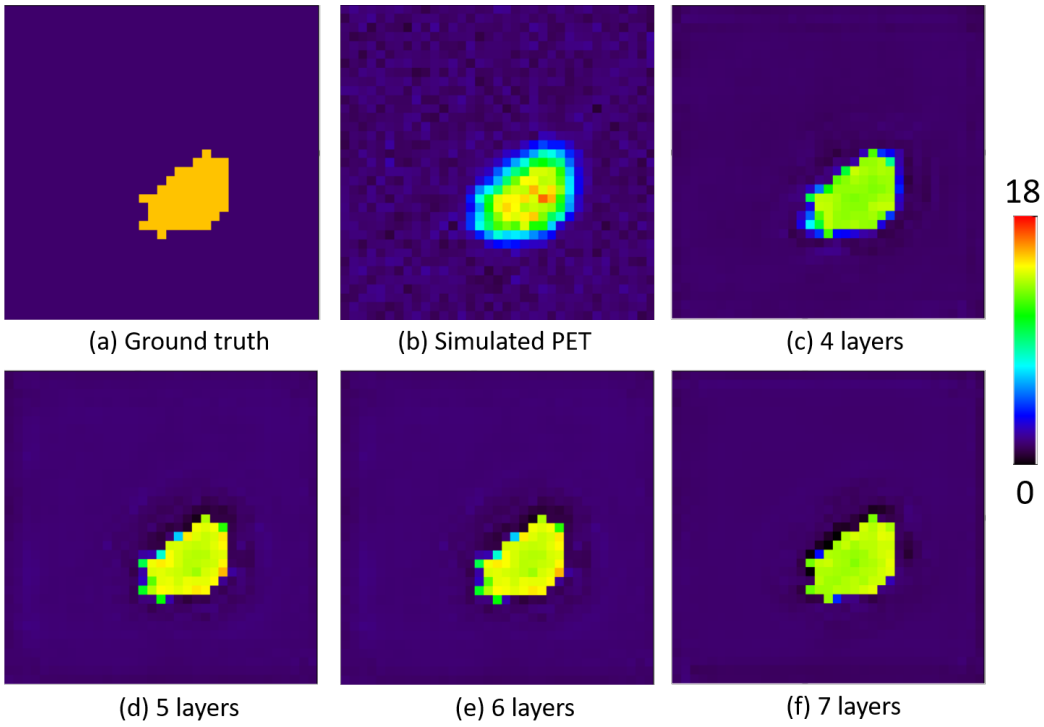
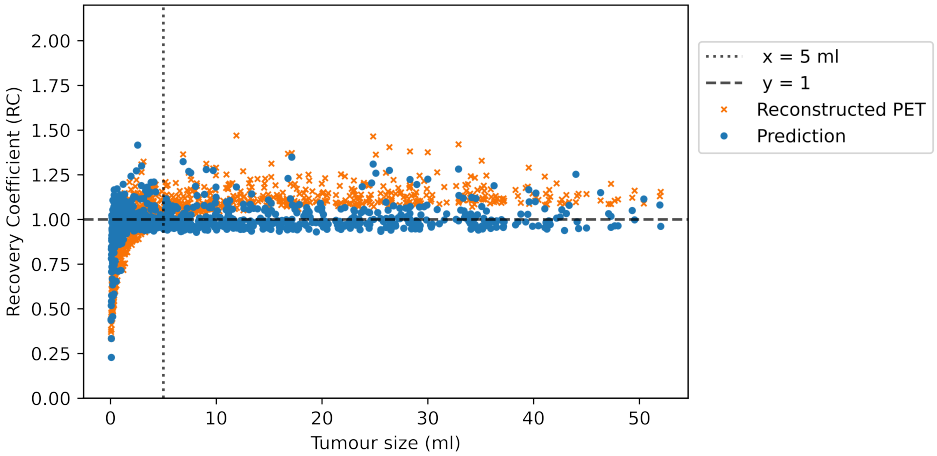
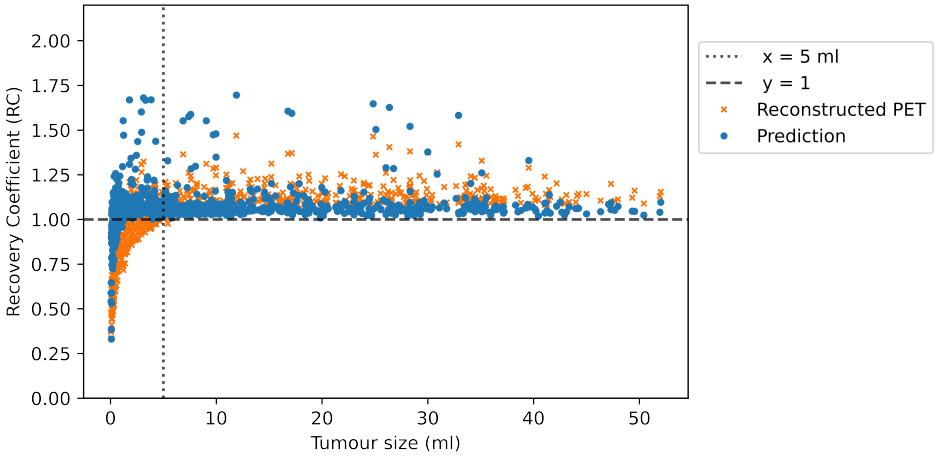


Figure 2.4: Transverse views of a uniform tumour. The images in (a) and (b) represent the ground truth image and the corresponding simulated PET image. The following images are predicted by CNNs made of 4 layers (c), 5 layers (d), 6 layers (e) and 7 layers (f). The tumours in (d) and (e) appear more similar to the ground truth in terms of tumour shape and activity concentration. All the images are shown with the same colour scale, expressed in kBq/ml.

mum recovery coefficients, estimated for all the simulated PET images in the test dataset and the corresponding CNN's predictions, are plotted against tumour volumes in Fig. 2.5 and Fig. 2.6. None of the CNNs tested in this experiment could accurately recover the maximum tumour activity for tumours with volumes smaller than approximately 5 ml. This can be seen in the presented plots, which show RCs as low as 0.25 for the smallest tumours. In this chapter the optimisation of the neural networks is guided by the quantitative analysis performed using tumour volumes larger than 5 ml. We separately analyse the recovery of small tumours for the optimised network. Table 2.1 presents the max, mean and median RC, averaged over all the images belonging to the

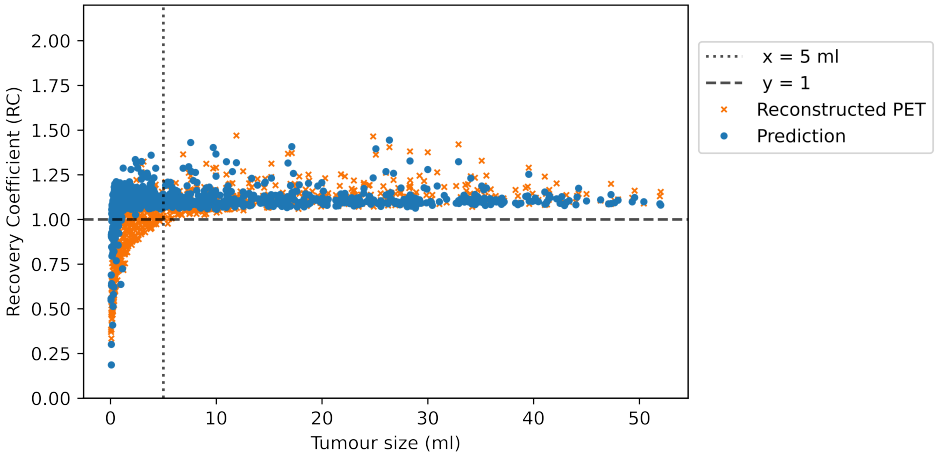


(a) 4 layers

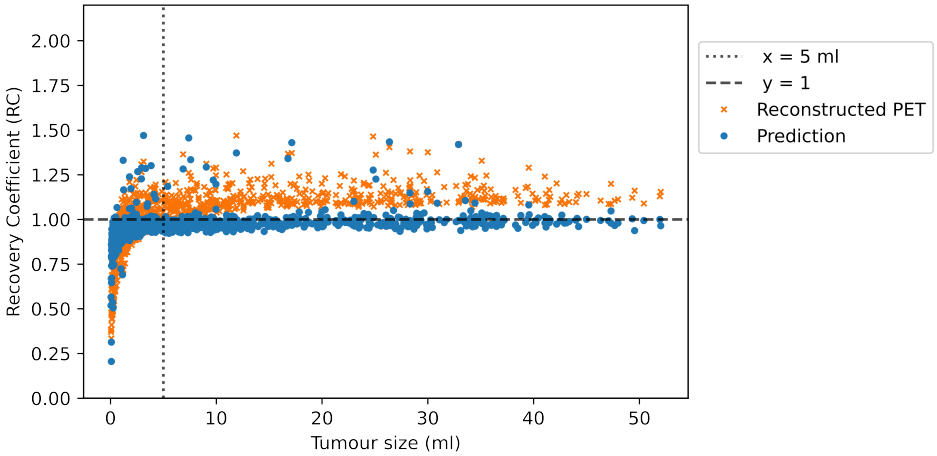


(b) 5 layers

Figure 2.5: RC_{max} values, measured for all the images in the test set, are here plotted against the tumour size expressed in ml. The coefficients measured using the simulated PET images are shown in orange, the ones measured using the CNN's predictions are shown in blue. CNNs composed of 4 and 5 convolutional layers yield the results shown in (a) and (b).



(a) 6 layers



(b) 7 layers

Figure 2.6: RC_{max} values, measured for all the images in the test set, are plotted here against the tumour size expressed in ml. The coefficients measured using the simulated PET images are shown in orange, the ones measured using the CNN's predictions are shown in blue. CNNs composed of 6 and 7 convolutional layers yield the results shown in (a) and (b).

test dataset. In this table, and in all the tables presented in this thesis, the mean and median RC are only estimated for uniform and hollow tumours. As expected, the CNNs composed of 4 and 8 layers have the worst performance. The poor performance of the network composed of 4 layers is probably due to the fact that too few layers are not sufficient for the CNN to learn the complex relationships between image voxels, whereas too many layers may result in over-fitting of the training dataset, so the network made of 8 layers is not able to generalise well to the test dataset. Improved RC estimates are measured in the images predicted by the networks with 5 and 6 and convolutional layers. The CNN with 6 convolutional layers predicts the best RC_{mean} and RC_{median}, whereas the CNNs with 7 and 4 convolutional layers yields the best RC_{max}. The RC_{max} provided by this CNN is higher than the estimates provided by other CNNs, nevertheless it is better than the RC_{max} measured using the simulated PET images. The MSSIM measurements, evaluated for the simulated PET and for the predictions obtained with different CNN configurations, are presented in table 2.2. All CNNs predict an improved MSSIM and the best performance is achieved by the CNN made of 6 convolutional layers.

convolutional layers	RC_{max}	RC_{median}	RC_{mean}
4	**1.01±0.07**	0.87±0.01	0.82±0.02
5	1.09±0.10	0.92±0.02	0.89±0.08
6	1.10±0.07	**0.93±0.03**	**0.92±0.01**
7	**0.99±0.07**	0.89±0.02	0.88±0.02
8	0.95±0.08	0.82±0.04	0.77±0.16
Simulated PET	1.13±0.07	0.88±0.06	0.85±0.04

Table 2.1: Average recovery coefficients obtained using CNNs with different depths. The coefficients measured using the simulated PET images are shown in the last row of the table. The CNN with 6 convolutional layers yielded the best estimate for RC_{mean} and RC_{median}, whereas the CNNs with 4 and 7 convolutional layers yielded the best RC_{max} average value. These values are measured for tumours larger than 5 ml.

convolutional layers	MSSIM
4	0.67±0.08
5	0.65±0.07
6	**0.70±0.08**
7	0.68±0.09
8	0.54±0.19
Simulated PET	0.20±0.09

Table 2.2: Average MSSIM obtained using CNNs with different depths. The MSSIM measured using the simulated PET images are shown in the last row of the table. All CNNs yield an improved MSSIM, the best performance is achieved by the CNN with 6 convolutional layers.

We now focus on the optimisation of the CNN with 6 layers as it is characterised by higher MSSIM, RC_{mean} and RC_{median}. The 3D CNN composed of 6 convolutional layers and 5 batch normalisation layers is compared to a 3D CNN composed of 6 convolutional layers only, and to a 3D CNN with 5 convolutional layers each followed by batch normalisation apart from the final fully connected (FC) layer. Visual results are shown in Fig. 2.7 for the same uniform tumour shown previously. The prediction yielded by the network containing a FC layer has blurrier edges and lower activity than the ground truth image. The images predicted by the CNNs with BN layers and without BN layers look similar, so a quantitative analysis is performed to further compare the three networks. The RC measured for the three networks, shown in table 2.3, confirm that the network with a final FC layer has the worst performance in terms of mean and median RC. The mean and median RC estimated by the CNN containing BN layers are higher than the ones yielded by the CNN that does not contain BN layers. For this reason, in the next experiments a CNN with 6 convolutional layers and 5 BN layers is used.

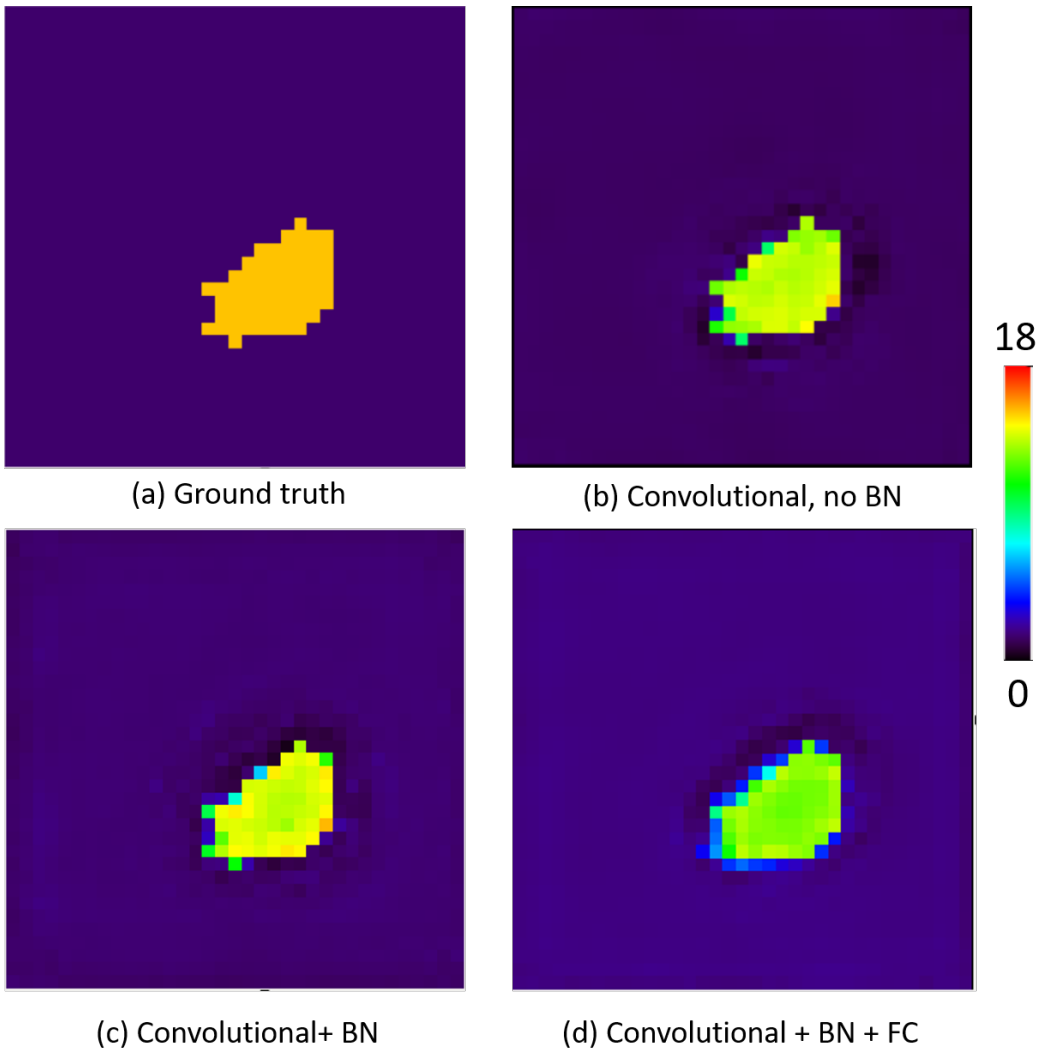


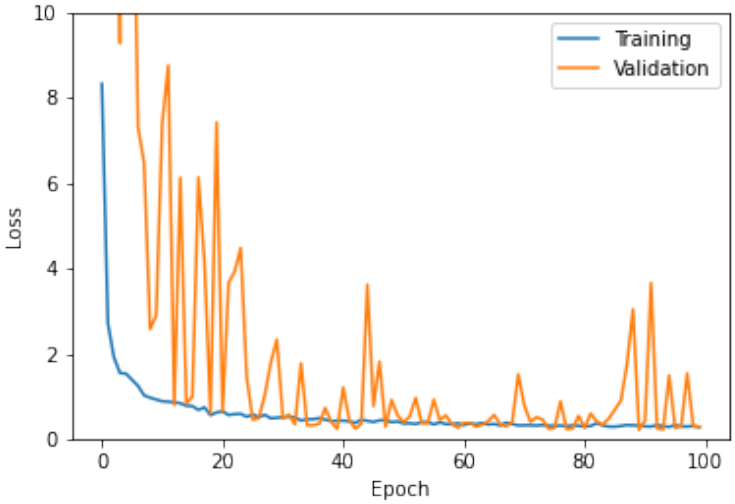
Figure 2.7: Comparison between transverse views of a uniform tumour, obtained using two different CNNs. The image in (b) is predicted by a CNN made of 6 convolutional layers, whereas the image in (c) is predicted by a CNN with 6 convolutional layers and 5 BN layers. The images yielded by the CNNs with and without BN look similar. The CNN made of 5 convolutional layers, each followed by BN, and by a final fully connected layer predicts the image in (d). In this image the tumour has blurrier edges and lower activity than the ground truth. The images are shown with the same colour scale, expressed in kBq/ml.

CNN architecture	RC_{max}	RC_{median}	RC_{mean}
convolutional, no BN	**1.01±0.04**	0.89±0.01	0.88±0.01
convolutional + BN	1.10±0.07	**0.93±0.03**	**0.92±0.01**
convolutional + BN + FC	0.98±0.13	0.84±0.04	0.83±0.04

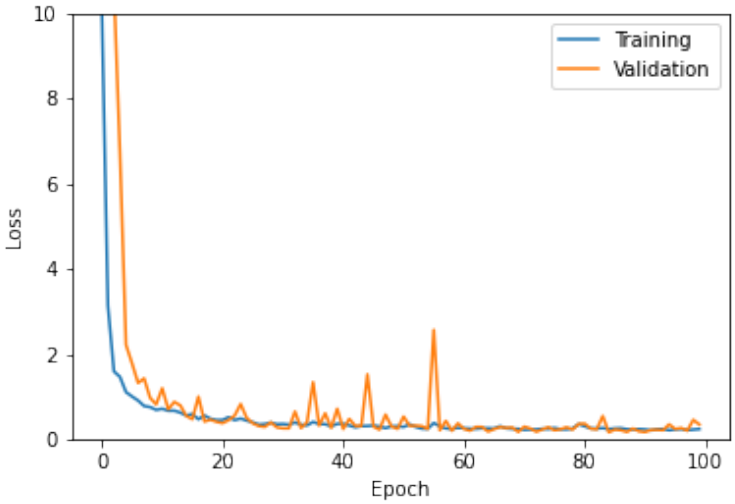
Table 2.3: Average recovery coefficients obtained using CNNs with different layers. The first row shows the results obtained using a CNN made of 6 convolutional layers only. The second row shows the results obtained using a CNN with 6 convolutional layers, each followed by batch normalisation except for the output layer. The results in the third row are yielded by a CNN with 5 convolutional layers, each followed by batch normalisation, and a final fully connected layer. The mean and median RC estimated by the CNN with BN are higher than the ones yielded by the CNN that does not contain BN layers. These values are measured for tumours larger than 5 ml.

2.2.2 Impact of Different Network Optimisation Algorithms

In this section we compare a 3D CNN made of 6 convolutional layers and BN layers trained with Adam optimiser, to the same network trained using RMSprop. Both networks are trained at first for 100 epochs, and the loss function is monitored through training. In Fig. 2.8, the training and validation loss functions obtained using RMSprop and Adam are plotted at each epoch. The network trained with Adam optimiser reaches convergence in a smaller number of epochs, as shown in Fig. 2.8 (b). When using Adam, after around 60 epochs the loss function converges to a value close to 0. In contrast, when RMSprop is used the loss function acquires values above 2 even after 60 epochs. To better investigate the behaviour of the loss functions, the 3D CNNs are also trained for 175 epochs and the corresponding plots are presented in Fig. 2.9. Training the network for a longer time confirmed that using Adam a faster and more stable convergence can be achieved. For this reason, Adam is used in the next experiments.

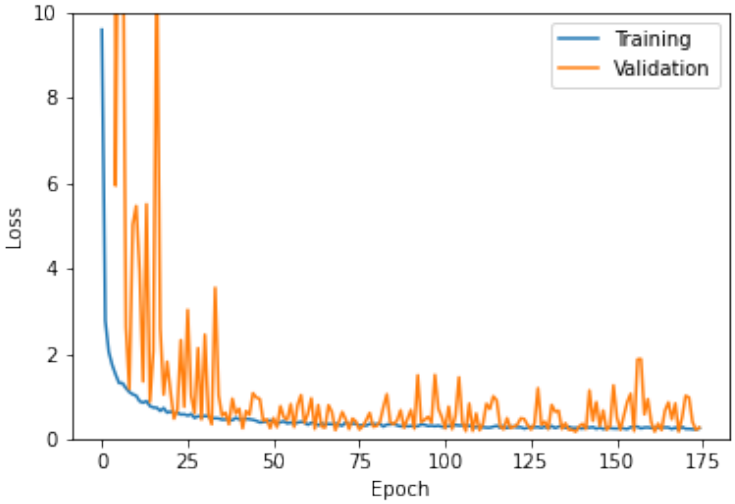


(a) RMSprop, 100 epochs

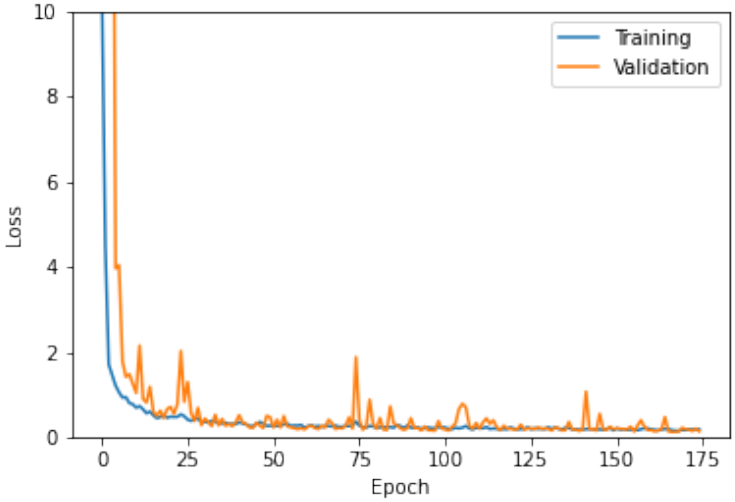


(b) Adam, 100 epochs

Figure 2.8: Illustration of loss functions, plotted against the number of epochs used for training. The training loss is shown in blue, whereas the validation loss is shown in orange. The plots in (a) and (b) are obtained training the CNN for 100 epochs, using RMSprop optimiser and Adam respectively. The network trained with Adam optimiser reaches convergence in a smaller number of epochs.



(a) RMSprop, 175 epochs



(b) Adam, 175 epochs

Figure 2.9: Illustration of loss functions, plotted against the number of epochs used for training. The training loss is shown in blue, whereas the validation loss is shown in orange. The plots in (a) and (b) are obtained training the CNN for 175 epochs, using RMSprop optimiser and Adam respectively. The network trained with Adam optimiser reaches convergence in a smaller number of epochs.

2.2.3 Impact of Different Data Normalisation Methods

In deep learning, data are usually normalised before being given as an input to neural networks. Normalisation is a pre-processing step that changes the activ-

ities of the input image voxels so that they lie within the same range of values. Without data normalisation, the weights can become large thus making the network more unstable and less generalisable. Furthermore, as the learning rate is determined by the feature with the largest range, normalising the training data can speed up the training process when the features in the training dataset have drastically different ranges. As PET is a quantitative imaging technique, the impact of normalising data before training the network needs to be carefully considered. For example, a normalisation method that changes the relations between voxels within the same image may negatively impact the CNN training leading to less accurate predictions. As quantification is a key aspect of this thesis, different normalisation techniques are investigated and compared. The predictions obtained training a 3D CNN with 6 convolutional layers and 5 batch normalisation layers for 175 epochs using non-normalised data are used as a benchmark. The RC measured using the images predicted by the network trained with these settings are shown in the first row of table 2.4.

At first the simulated PET images in the training dataset are normalised by dividing each image voxel by the maximum activity recorded in the entire dataset of simulated PET images. The results measured on the rescaled images obtained using this method, called "overall max", are shown in the second row of table 2.4. Overall, the RCs are lower than the ones obtained training the network on non-normalised data. One possible explanation for this is that when using the overall maximum to normalise the data, most values within the single images characterised by lower activity are scaled to very low values so the network is not as accurate in its predictions. A second simulation is performed scaling each simulated PET image individually, by dividing each voxel by the maximum activity measured in each image. The RCs derived from the rescaled predicted images obtained with this method, henceforth referred

to as "Individual max", are shown in the third row of table 2.4. In this case all RCs are overestimated. One possible explanation for this is that when dividing each simulated PET image by its maximum value, the relation between the ground truth and simulated PET intensity values is lost. Finally, normalisation is performed scaling each simulated PET image individually, using the 'sklearn' normalise function that scales the input vectors individually to unit norm. The results are shown in the bottom row of table 2.4. This normalisation method did not improve the RC values. This quantitative analysis suggests that the best results are obtained training the network on non-normalised data. For this reason, non-normalised data are used in all further experiments in this chapter.

Normalisation	RC_{max}	RC_{median}	RC_{mean}
None	1.11±0.06	**1.00±0.03**	**0.99±0.03**
Overall Max	**1.06±0.06**	0.97±0.02	0.96±0.03
Individual Max	1.33±0.20	1.10±0.10	1.10±0.10
Sklearn	1.11±0.13	0.96±0.04	0.95±0.03

Table 2.4: Average RCs obtained normalising the simulated PET data used for training with different methods. The best RC_{max} is obtained when normalising the training dataset by the overall max. The best $RC_{mean}$and RC_{median} are estimated by the CNN trained on non-normalised data. These values are measured for tumours larger than 5 ml.

2.2.4 Impact of the Number of Epochs Used for Training

The number of epochs used for training indicates how many times the entire training dataset is passed through the network. Multiple passes of the dataset through the network allow the weights to be better adjusted and the features characteristics of the chosen data to be better learnt. When the network is

trained for a small number of epochs the small number of training data seen by the network hampers the optimal tuning of the high number of training parameters, leading to under-fitting. On the other hand, when the model is trained for too many epochs over-fitting of the training data occurs. This happens because the model learns details and noise information of the training dataset, so it has a poor performance when applied to an unseen set of data. There are techniques which help deciding what could be an appropriate number of epochs for a given dataset. Monitoring the training loss and the validation loss can guide this decision. The validation loss function gives information on the generalisation error, and by monitoring it we can have an idea of if and when over-fitting begins to occur. The experiments in section 2.2.2 suggest that when training with Adam optimiser, the loss functions converge at approximately 100 epochs. In this section, the 3D CNN is trained on the same dataset with increasingly high number of epochs, starting from 100 and up to 200, to investigate how this parameter affects the final predictions.

A visual comparison of the predictions obtained training the same CNN for different number of epochs, is presented in Fig. 2.10. In this figure, transverse views of a uniform tumour are shown. Similarly, the predictions obtained for a representative hollow tumour are shown in Fig. 2.11. All the tumours predicted by the networks are characterised by better defined tumour edges, and the predicted images are denoised. Both in Fig. 2.10 and Fig. 2.11, the predictions obtained training the CNN for 175 epochs are characterised by activity values that are more similar to the ground truth. This visual assessment is followed by a quantitative analysis, which is presented in table 2.5. As expected, the best estimates of RC mean and median are obtained when training the CNN for 175 epochs. This indicates that training the network for 175 epochs yields the best results when using this specific training dataset.

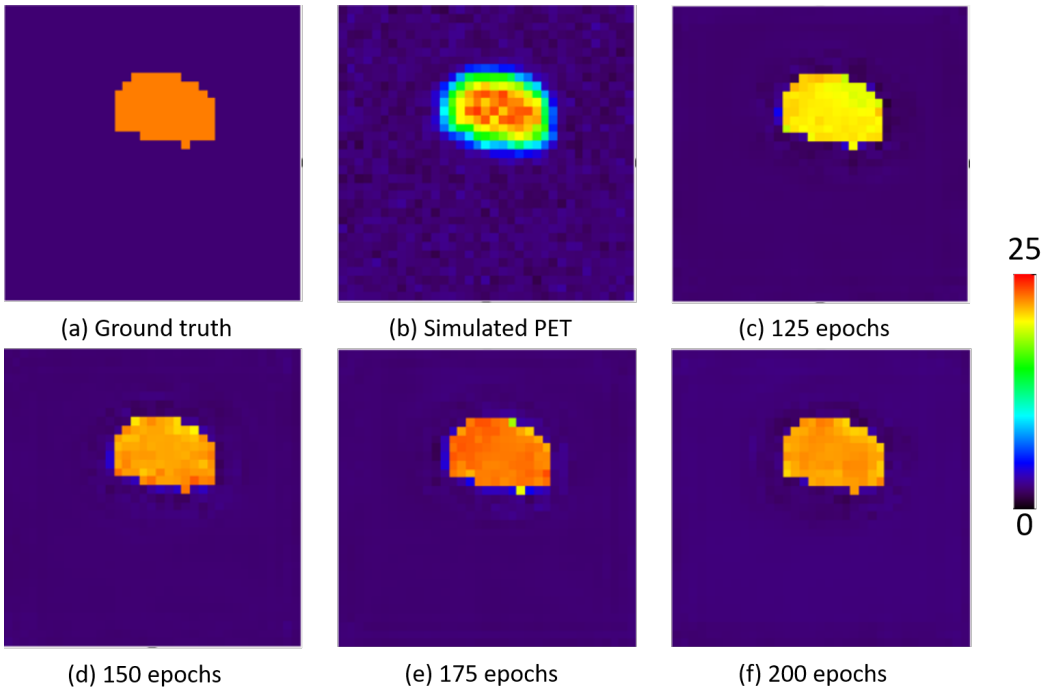


Figure 2.10: Transverse views of a uniform tumour. In (a) and (b) the ground truth image and the corresponding simulated PET image are shown. The following images are obtained training the same 3D CNN for increasingly high number of epochs, starting from 125 in (c) up to 200 (f). The CNN predicts denoised images with sharper tumour edges. All the images are shown with the same colour scale, expressed in kBq/ml.

	RC_{max}	RC_{median}	RC_{mean}
100 epochs	1.06±0.05	0.89±0.09	0.90±0.03
125 epochs	**1.02±0.10**	0.89±0.08	0.87±0.08
150 epochs	1.10±0.08	0.94 ±0.09	0.93±0.09
175 epochs	1.11±0.06	**1.00±0.03**	**0.99±0.03**
200 epochs	1.05±0.09	0.97±0.01	0.94±0.08
Simulated PET	1.13±0.07	0.88±0.06	0.85±0.04

Table 2.5: Average RCs obtained training a CNN with Adam optimiser, with increasingly high numbers of epochs. The best RC_{max} is obtained when training the CNN for 125 epochs. The best RC_{mean} and RC_{median} are estimated by the CNN trained for 175 epochs. These values are measured for tumours larger than 5 ml.

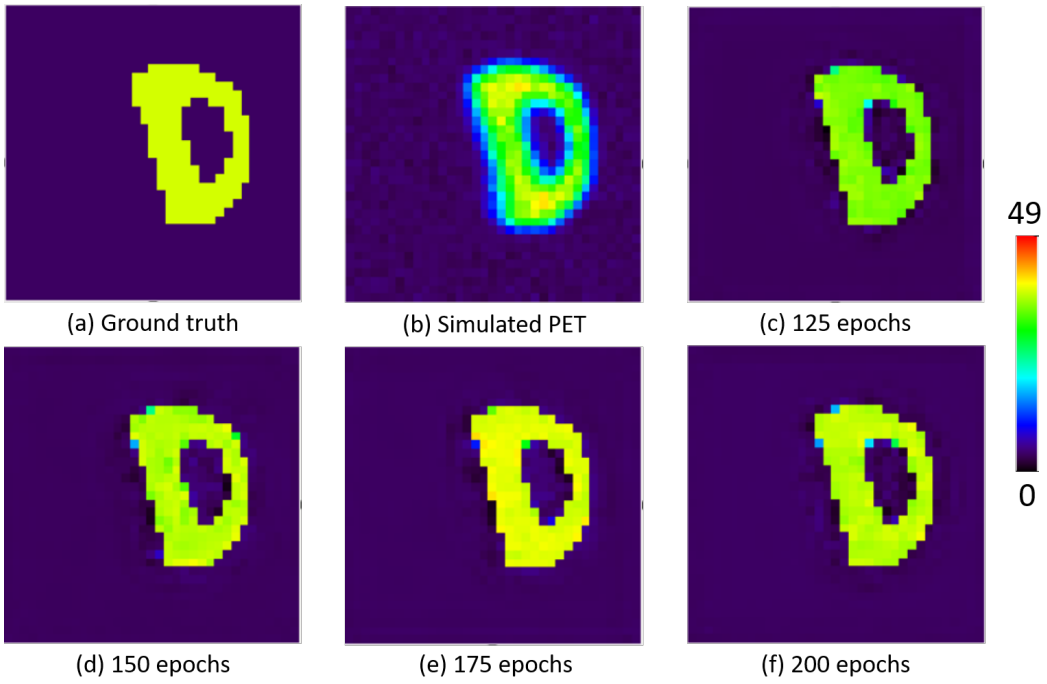


Figure 2.11: Transverse views of a hollow tumour. In (a) and (b) the ground truth image and the corresponding simulated PET image are shown. The following images are obtained training the same 3D CNN for increasingly high number of epochs, starting from 125 in (c) up to 200 (f). The CNN predicts denoised images with sharper tumour edges. The network trained for 175 epochs predicted the best image in terms of tumour shape and activity values. All the images are shown with the same colour scale, expressed in kBq/ml.

2.2.5 Optimised CNN: a more Detailed Analysis

In this section, the results predicted by the CNN composed of 6 layers and trained for 175 epochs with Adam optimiser are further analysed. A useful measurement that is often provided together with the maximum and mean tumour activity is the peak activity. Various definitions have been proposed to measure the peak value, using different VOI shapes, sizes and locations [149]. In this work, for each tumour voxel the average activity concentration is measured within a 1 cm^3 spherical VOI centred in that voxel. The peak tumour activity is defined as the highest average value measured using the 1

cm^3 sphere. A plot of the RC_{peak} is presented in 2.12. As observed for the other recovery coefficients, the RC_{peak} is not accurately recovered for small tumours. The quantitative analysis, performed for tumour volume larger than 5 ml, provided an average $RC_{peak} = 1.00 \pm 0.02$ for the images predicted by the CNN, compared to an average $RC_{peak} = 0.93 \pm 0.08$ for the simulated PET images. Thus, the optimised 3D CNN also improved the RC_{peak} estimates.

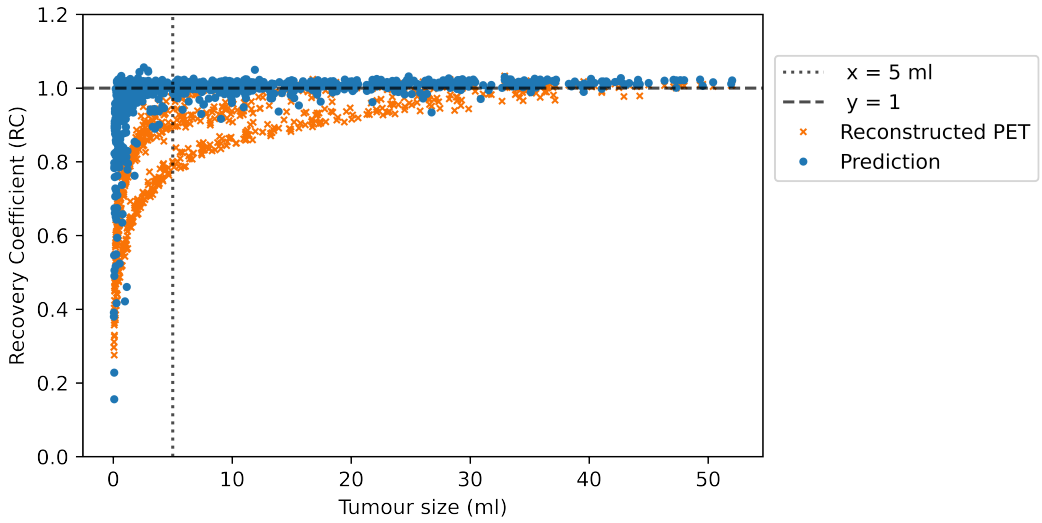


Figure 2.12: The RC_{peak} values are here plotted against the tumour volume, expressed in ml. The coefficients measured using the simulated PET images are shown in orange, the ones measured using the CNN's predictions are shown in blue.

Up to this point the quantitative analysis was only performed using tumour volumes larger than 5 ml, for which the RC values are more accurately recovered. A more detailed analysis is performed on small tumours, using the images predicted by the optimised CNN. Many simulated tumours present an elongated shape and are thus more affected by partial volume effect. In order to assess if there is any correlation between the tumour shape and the network's performance, the tumour sphericity was measured. The sphericity is a dimensionless metric defined as:

$$sphericity = \frac{\sqrt[3]{36\pi \cdot V^2}}{A} \tag{2.5}$$

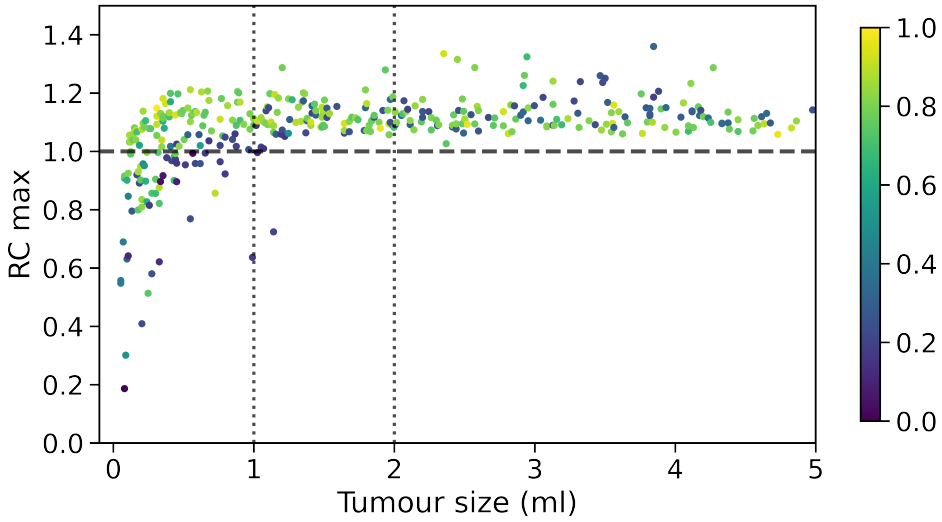
where V is the tumour volume and A is the tumour surface area. It describes the roundness of the tumour shape relative to a sphere and its values range from 0 to 1. A sphericity value of 1 indicates that the tumour has a perfect spherical shape. In Fig. 2.13 and Fig. 2.14, the recovery coefficients measured using the CNNs predictions are plotted for tumour volumes smaller than 5 ml. In these plots the tumour sphericity is colour-coded, so that less spherical tumour shapes are shown in darker colours. In these plots, more elongated tumours are often associated with lower RCs. This suggests that the network recovers the activity concentration less accurately for small elongated tumours than for small spherical tumours.

2.2.6 Comparison with Standard Deconvolution and Denoising Methods

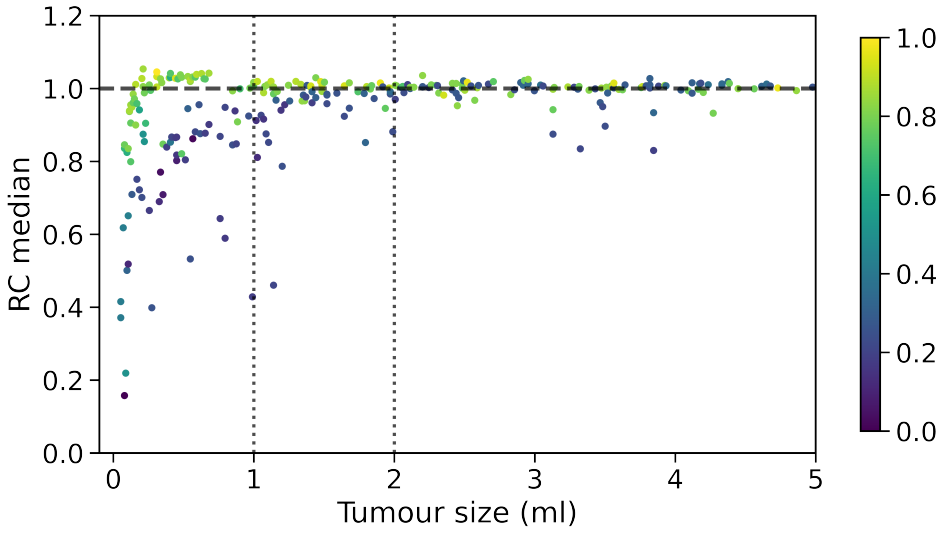
In this section, the images predicted by the CNN are compared to non-deep learning based techniques that are commonly used to deconvolve and denoise PET images. These techniques are:

- Lucy-Richardson deconvolution
- Gaussian filtering
- Median filtering

Lucy-Richardson (LR) is an iterative algorithm standardly used to restore images that have been degraded by a known point spread function. The algorithm is based on maximising the likelihood that the blurry image is an instance of the original image under Poisson statistics. The Gaussian filtering



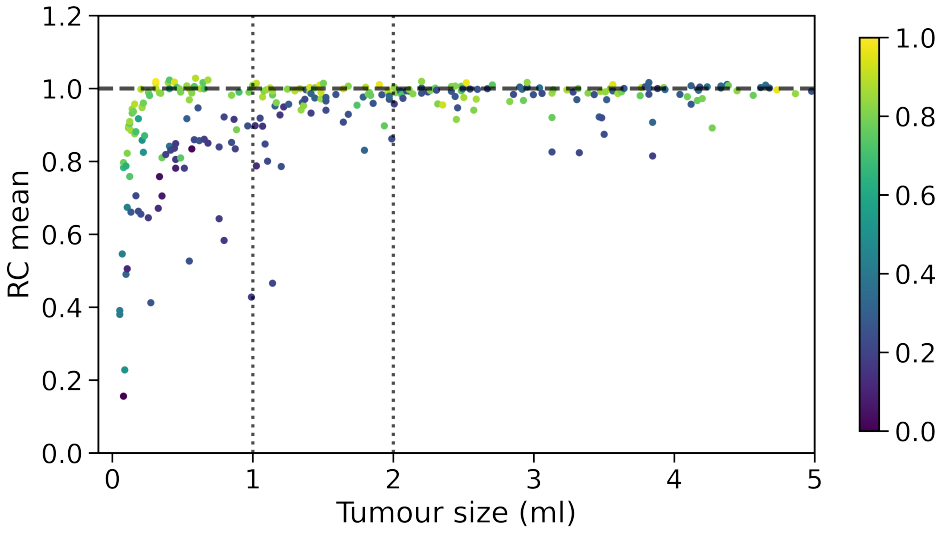
(a) RC_{max}



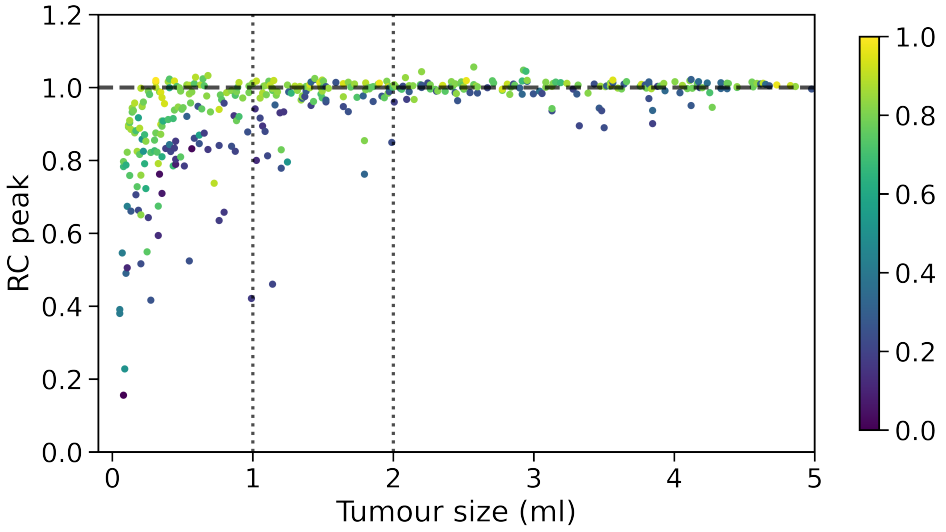
(b) RC_{median}

Figure 2.13: The RC_{max} and RC_{median} and measured using the CNN's predicted images are plotted here against the tumour volume, for tumours smaller than 5 ml. The tumour sphericity is colour-coded so that more spherical tumours are represented in yellow whereas less spherical tumours are shown in dark blue. Worse RCs are correlated with lower tumour sphericity.

operation consists of convolving the noisy image with a Gaussian function, with the aim of removing noise. A median filter can also be used to remove image noise. It is usually applied through the image voxel by voxel, replacing



(a) RC_{mean}



(b) RC_{peak}

Figure 2.14: The RC_{mean} and RC_{peak} measured using the CNN's predicted images are here plotted against the tumour volume, for tumours smaller than 5 ml. The tumour sphericity is colour-coded so that more spherical tumours are represented in yellow whereas less spherical tumours are shown in dark blue. Worse RCs are correlated with lower tumour sphericity.

each voxel activity with the median of neighbouring activities.

A visual comparison between images of a hollow tumour, post-processed with different non-DL methods is presented in Fig. 2.15. In (d) Lucy-Richardson

deconvolution is performed using 30 iterations and a Gaussian kernel with FWHM = [4, 4, 4] mm^3 and truncated at 4 σ, where σ is the standard deviation of the Gaussian distribution. The resulting image is noisy and, although the tumour shape definition is improved, the tumour shape still looks blurry around the edges. The simulated PET image post-processed by applying a Gaussian filter with FWHM = [4, 4, 4] mm^3 is shown in 2.15 (e). This image is denoised and the tumour activity is more uniform, but the tumour edges are blurry. Finally, the image in Fig. 2.15 (f) is obtained by subsequently applying Lucy-Richardson algorithm and Gaussian filtering. This image is denoised and the tumour edges are sharper than in image obtained only applying the Gaussian filter. Nevertheless, a range of tumour activities can be observed within the tumour, with a number of voxels showing higher activity than the ground truth activity. The prediction obtained using a 3D CNN composed of 6 convolutional layers and 5 BN layers, trained for 175 epochs with Adam optimiser, is shown in Fig. 2.15 (c). The CNN outperforms all other methods as it is able to yield a denoised image with sharp tumour edges and more uniform tumour activity distribution. Similar considerations can be made looking at Fig. 2.16, in which the same comparison is shown for a uniform tumour. The quantitative analysis of the images obtained with these different techniques is presented in table 2.6. Lucy-Richardson deconvolution followed by Gaussian filtering provides on average the best RC_{max} coefficients, but it underestimates the mean and median RC. The Lucy-Richardson algorithm alone provides slightly higher RC_{mean} and RC_{median} values, but it also overestimates the RC_{max}. The CNN provides better mean and median RCs than all other image post-processing techniques.

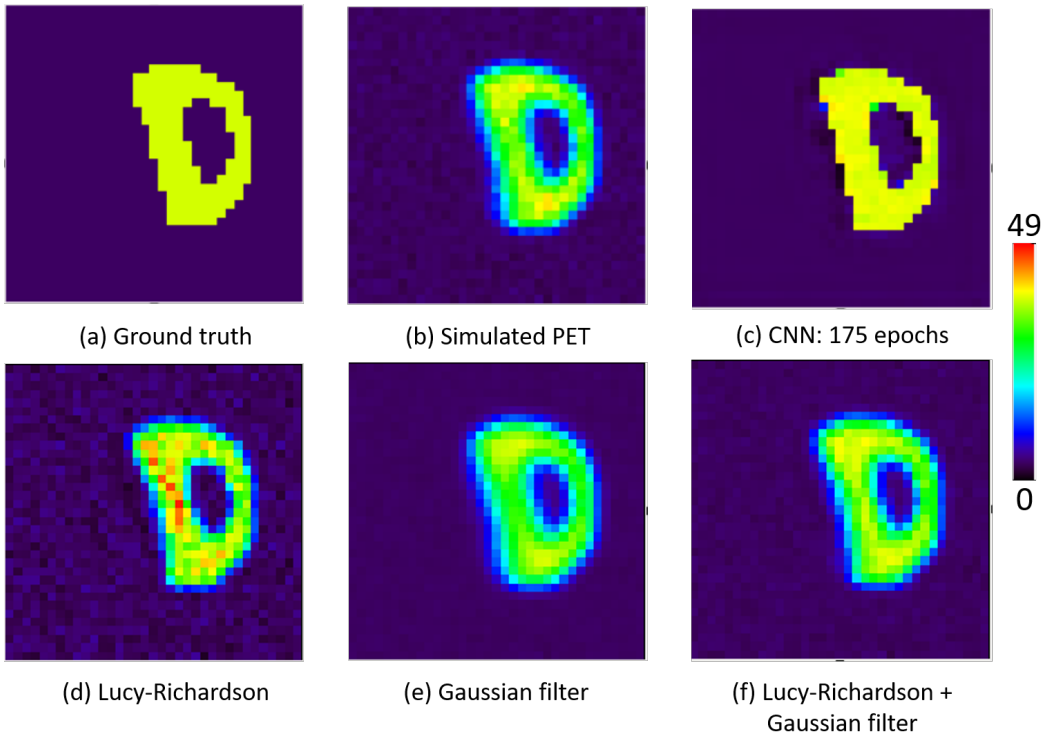


Figure 2.15: Comparison between transverse views of a hollow tumour, obtained with the proposed deep learning method other standard deconvolution and denoising methods. In (a) and (b) the ground truth and simulated PET images are shown. The prediction yielded by a CNN trained for 175 epochs is shown in (c). The image in (d) was obtained by applying LR deconvolution to the simulated PET image. The image in (d) was obtained by applying a Gaussian filter to the simulated PET image and the image in (e) was obtained by applying LR deconvolution and subsequently a Gaussian filter. The CNN provides better images than all other image post-processing techniques, both in terms of tumour shape and activity distribution. All the images are shown with the same colour scale, expressed in kBq/ml.

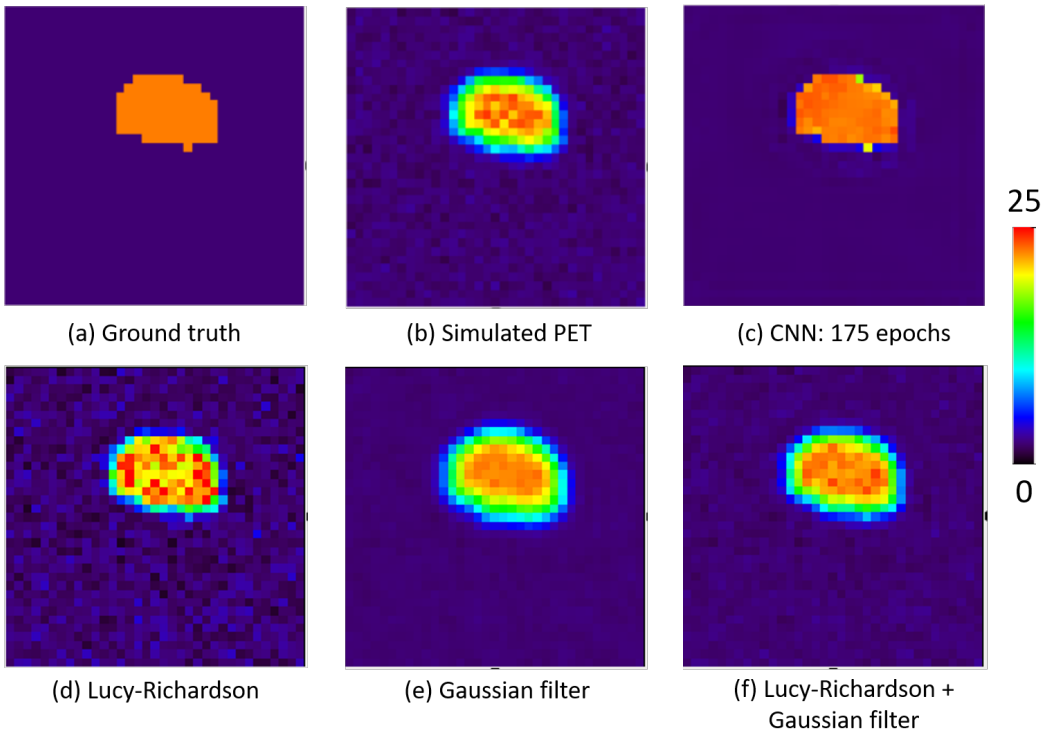


Figure 2.16: Comparison between transverse views of a uniform tumour, obtained with the proposed deep learning method other standard deconvolution and denoising methods. In (a) and (b) the ground truth and simulated PET images are shown. The prediction yielded by a CNN trained for 175 epochs is shown in (c). The image in (d) was obtained by applying Lucy-Richardson deconvolution to the simulated PET image. The image in (d) was obtained by applying a Gaussian filter to the simulated PET image and the image in (e) was obtained by applying Lucy-Richardson deconvolution and subsequently a Gaussian filter. The CNN provides better images than all other image post-processing techniques, both in terms of tumour shape and tumour activity distribution. All the images are shown with the same colour scale, expressed in kBq/ml.

	RC_{max}	RC_{median}	RC_{mean}
Gaussian filter	0.90±0.19	0.71±0.18	0.69±0.16
Lucy-Richardson	1.41±0.22	0.85±0.13	0.83±0.11
Gaussian filter + Lucy-Richardson	**1.02±0.16**	0.78±0.16	0.75±0.14
CNN, 175 epochs	1.11±0.06	**1.00±0.03**	**0.99±0.03**
Simulated PET	1.13±0.07	0.88±0.06	0.85±0.04

Table 2.6: Average recovery coefficients obtained post-processing the simulated PET images with Gaussian filtering, Lucy-Richardson algorithm and a combination of the two. The results are compared to the recovery coefficients extracted from the simulated PET images and from the CNN's predictions. The CNN yields the best RC_{mean} and RC_{median} estimates.

2.2.7 CNN robustness to PET Images Generated with Multiple PSFs

One of the problems that affect quantification in PET imaging is that PET images acquired with different systems or reconstructed with different algorithms are characterised by different parameters (such as the PSF) and they cannot be quantitatively compared directly. One goal of the proposed method is to improve quantification of tumour radiotracer uptake in data generated with different parameters. A new experiment is performed to test the robustness of the proposed method to data generated with different PSFs. The new dataset used in this experiment is composed of 4200 images made of 35×35×40 voxels, with voxel size 3.18×3.18×2.00 mm^3. The new voxel size matches the voxel dimension of the phantom PET scans on which the CNN will be applied, after being trained and tested on simulated data. The simulated PET images are generated by assigning different PSFs to tumours with different activity patterns. The PET images of the uniform tumours are generated using a PSF with FWHM = [4, 4, 4] mm^3. The PET images of the tumours split in halves are generated with a PSF with FWHM = [5, 5, 5] mm^3 and the PET images

relative to the hollow tumours are generated with a PSF with FWHM = [6, 6, 6] mm^3. A secondary aim of this experiment is to test the performance of the network on a noisier dataset, so the simulated PET data in this dataset are generated setting the constant k to 0.66. According to the results obtained previously, the 3D CNN used in this experiment is trained for 175 epochs on non-normalised data, using Adam. Three representative tumours and the corresponding CNN predictions are shown in Fig. 2.17. The ground truth images and the corresponding simulated PET images are shown in the first and second column. The CNN predictions, in the third column, are denoised and they result more similar to the ground truth images both in terms of tumour shape and tumour activity distribution. This suggests that the network trained on data generated with one PSF can recover improved tumour shapes and activity distributions when tested on PET images generated with a range of PSFs. The quantitative analysis performed on this dataset shows that when using the simulated PET images and the ground truth images, a MSSIM equal to 0.14 ±0.11 is measured. When using the predicted images, the measurement of MSSIM is improved to 0.63 ±0.11. This confirms that the CNN predicts images that are more similar to the ground truth in terms of shape and contrast. As shown in table 2.7, the CNN also provides improved average RCs. The RC_{mean} and RC_{median} are improved, although these values are not as good as the estimates obtained when training and testing the CNN on a dataset in which PET data are generated with a single PSF.

	RC_{max}	RC_{median}	RC_{mean}
CNN	**1.07±0.15**	**0.90±0.13**	**0.87±0.13**
Simulated PET	1.29±0.32	0.76±0.16	0.76±0.15

Table 2.7: Average recovery coefficients measured on simulated PET images generated with different PSFs, and the corresponding CNN predictions. The CNN improves all RC estimates.

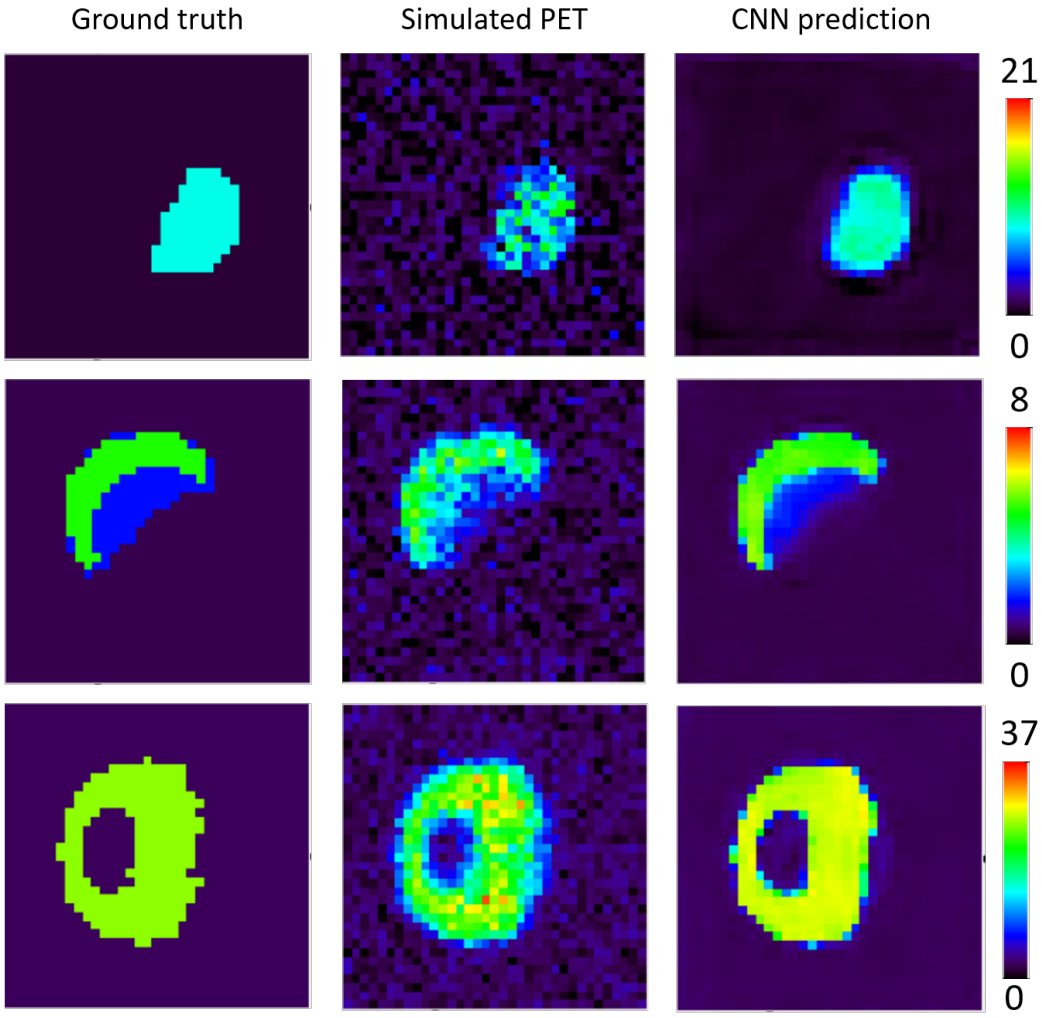


Figure 2.17: Transverse views of three tumour shapes. A uniform tumour is shown in the first row, a heterogeneous tumour is the second row, and a hollow tumour is presented in the bottom row. The first column shows the ground truth images. The simulated PET images, in the middle column, are generated with PSF = [4, 4, 4] mm^3 for the uniform tumour, PSF = [5, 5, 5] mm^3 for the heterogeneous tumour, and PSF = [6, 6, 6] mm^3 for the hollow tumour. The final column contains the images predicted by the CNN. The network trained on data generated with one PSFs improves the tumour shape and activity when tested on PET images generated with a range of PSFs. Images of the same tumour are shown with the same colour scale, expressed in kBq/ml.

2.3 Application to Phantom PET Images

After testing the deep learning method on simulated data, the CNN is applied to real phantom data. The phantom used for this application has been developed in Groningen by E. Pfaehler and colleagues. It is a NEMA NU 2-2012 IQ phantom that has a shape similar to a torso, in which the spheres are replaced by three realistic 3D printed tumour inserts. These inserts are designed according to non-small cell lung cancer lesions observed in real patient studies [146]. The top row of Fig. 2.18 shows coronal views extracted from a PET image of the three inserts, reconstructed with OSEM algorithm. The first tumour, shown in the first column, has a volume of 46.05 ml and it is filled with an activity solution of 19.49 kBq/ml. The second insert is composed of two parts, the upper part (10.75 ml) filled with an activity solution of 10.94 kBq/ml and the lower part (13.12 ml) filled with an activity solution of 19.49 kBq/ml. A coronal view of this insert is shown in the second column of Fig. 2.18. The third column of Fig. 2.18 shows the third tumour insert, which is hollow. The outer part (65.35 ml) is filled with an activity solution of 19.49 kBq/ml and the inner core (7.80 ml) is filled with water. The activity solution injected in the background compartment of the phantom is 1.94 kBq/ml. The PET images of this phantom are acquired on a PET/CT system (Siemens Biograph mCT-40 PET/CT, Siemens, Knoxville, TN, USA). The data are reconstructed to obtain a frame of 300 s, using OSEM algorithm (3 iterations, 24 subsets) or a time-of-flight (TOF) iterative reconstruction method (3 iterations, 21 subsets) provided by the vendor. All PET images are reconstructed with and without PSF modelling. The final PET images have a size of 256×256×256 voxels, with voxel size 3.18×3.18×2 mm^3. A region made of 35×35×40 voxels is cropped around each tumour and given as input to the 3D CNN. Visual results obtained applying the CNN to the phantom

PET images reconstructed with OSEM algorithm are shown in the bottom row of Fig. 2.18. The CNN's predictions are denoised and the tumour inserts are characterised by better defined shapes. Similar results are obtained when testing the network on 35×35×40 voxels regions cropped from the PET images reconstructed with ToF iterative reconstruction and PSF modelling, as shown in Fig. 2.19. The images predicted by the CNN, shown in the bottom row, are denoised and tumours are characterised by sharper edges. This indicates that the same network, trained on simulated data, can improve the recovery of phantom images reconstructed with different algorithms. The quantitative analysis presented in table 2.8 shows the RC_{max} measurements performed on the phantom PET images and on the CNN's predictions. The RCs presented in this table are averaged over the three tumour inserts. The network yields improved RC_{max} values for all the reconstruction algorithms, the best result is obtained when using phantom images reconstructed with ToF iterative reconstruction. The RC_{mean} values measured on the same set of images are presented in table 2.9. The network improves the RC_{mean} when applied to phantom PET images reconstructed with different techniques. The best performance is achieved when applying the network on PET images reconstructed with ToF algorithm and PSF modelling.

RC max	CNN prediction	PET images
OSEM	1.25 ± 0.15	1.78 ± 0.04
PSF + TOF	1.33 ± 0.14	1.32 ± 0.08
PSF	1.31 ± 0.09	1.37 ± 0.04
TOF	**1.14 ± 0.07**	1.65 ± 0.15

Table 2.8: The RC_{max} values measured using the CNN's predicted images are shown in the second column. The RC_{max} measured on the phantom images reconstructed with different algorithms are presented in the third column. For each reconstruction algorithm, the RCs are averaged over the three phantom inserts. The best estimate is obtained applying the CNN on ToF PET images.

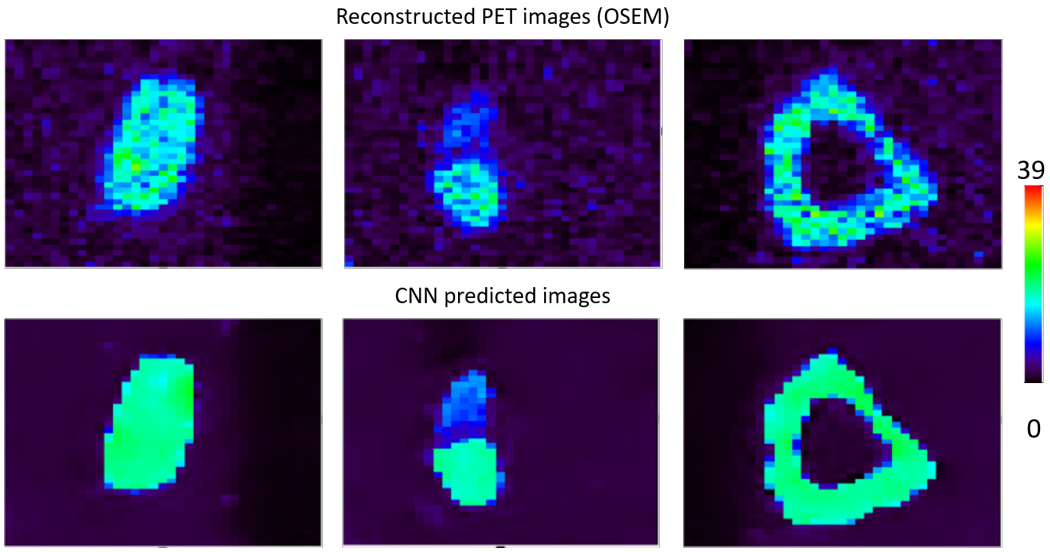


Figure 2.18: Coronal views of three phantom inserts. In the top row, three images cropped from the phantom PET scan reconstructed with OSEM algorithm are presented. The corresponding predictions yielded by the CNN are shown in the bottom row. The CNN's predictions are denoised, and the tumour inserts are characterised by better defined shapes. All images are shown using the same colour scale, expressed in kBq/ml.

RC mean	CNN prediction	PET images
OSEM	0.94 ± 0.02	0.77 ± 0.01
PSF + TOF	**0.99 ± 0.07**	0.81 ± 0.01
PSF	0.96 ± 0.06	0.80 ± 0.01
TOF	0.92 ± 0.04	0.77 ± 0.01

Table 2.9: RC_{mean} values measured using the CNN's predicted images are shown in second column. The RC_{mean} measured on the phantom images reconstructed with different algorithms are presented in the third column. For each reconstruction algorithm, the RC values are averaged over the three phantom inserts. The best estimate is obtained applying the CNN on PET images reconstructed with ToF algorithm and PSF modelling.

2.4 Summary

In this chapter we have presented a DL method that improves quantification of radiotracer uptake in simulated data. Several 3D CNN architectures have been

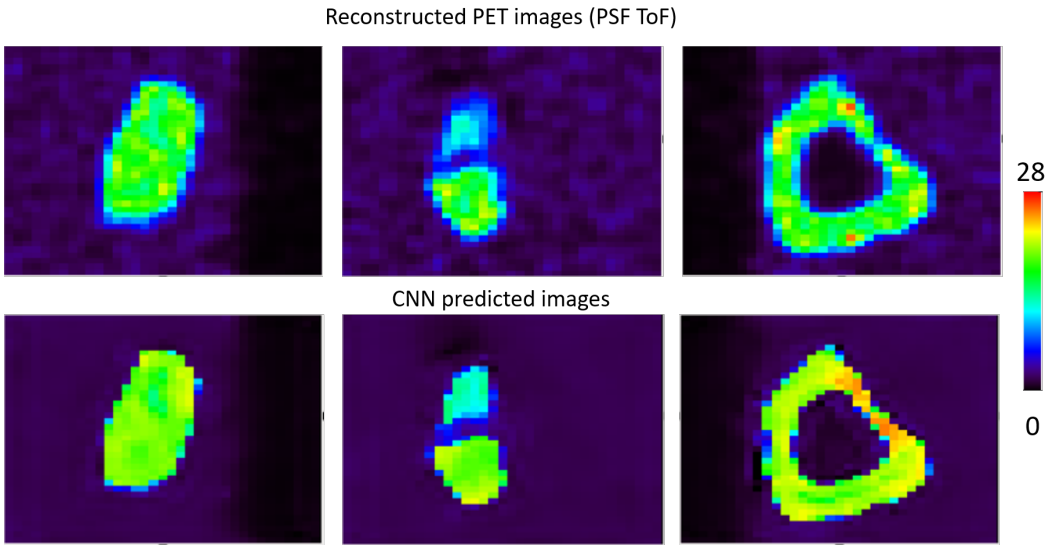


Figure 2.19: Coronal views of three phantom inserts. In the top row, three images cropped from the phantom PET scan reconstructed with ToF iterative reconstruction and PSF modelling are presented. The corresponding predictions yielded by the CNN are shown in the bottom row. The images predicted by the CNN are denoised, and the tumours are characterised by sharper edges. All images are shown using the same colour scale, expressed in kBq/ml.

trained and tested on a simulated dataset, composed of 4200 ground truth images and 4200 simulated PET images. This dataset was generated with voxel size 2.09×2.09×2.03 mm^3 and the PET images were generated with a single PSF with FWHM = [4, 4, 4] mm^3. CNNs composed of four to eight convolutional layers were trained on the same dataset, and the resulting predictions were analysed qualitatively and quantitatively. The 3D CNN composed of six convolutional layers, each followed by batch normalisation except for the final layer, could best improve the mean and median RC. Tumours smaller than 5 ml were not accurately recovered, which may be due to the fact that in these cases the tumours were very pixelated in the ground truth images, so the network could not learn a well defined shape. When the tumours were very small and they had low activity concentration, they appeared very blurry in the simulated PET images so it was even more difficult for the network to lo-

calise the tumour boundaries. Another experiment was performed to compare this architecture to a CNN only composed of 6 convolutional layers and to a CNN made of 5 convolutional layers, each followed by BN, and of a final fully connected layer. A quantitative analysis showed that the CNN composed of 6 convolutional layers and 5 BN layers could best recover the mean and median tumour RCs. For this reason, this configuration was used in all the subsequent simulations.

Next, the impact of the optimiser used for training on the network's performance was assessed. The CNN composed of 6 convolutional layers and 5 BN layers was trained for 175 using RMSprop or Adam as optimisers. A faster and more stable convergence of the training and validation loss functions was achieved when using Adam as optimiser. The same network was then trained on input images normalised with different methods. The best results were obtained when training the 3D CNN on non-normalised data, which were used in further experiments.

The 3D CNN was then trained for increasingly high number of epochs, to assess how this parameter would impact on the network's predictions. The number of epochs used for this experiment ranged from 100 to 200. Visual and quantitative analysis showed that the best results in terms of tumour shape, RC_{mean} and RC_{median} were achieved training the 3D CNN for 175 epochs. Finally, the results obtained using the proposed deep learning approach were compared to non-deep learning methods commonly used to improve quantification in PET images. The images yielded by the 3D CNN were compared to simulated PET images smoothed with a Gaussian filter and to the same simulated PET images post-processed using the Lucy-Richardson algorithm. A combination of these two techniques, Gaussian filtering and Lucy-Richardson, was also used for further comparison. The proposed deep learning method outperformed the other techniques, yielding better defined tumour shapes and

better RC_{mean} and RC_{median} estimates. A small number of standard deconvolution and denoising methods was considered, each tested with small set of parameters (i.e. filter size, number of iterations). Although the analysis that we presented is limited, the proposed DL method offers a superior performance in terms of tumour shape recovery. Furthermore, the proposed DL method is not affected by quality degrading effects such as noise amplification and artifacts that are common in standard deconvolution methods.

One goal of this thesis was to develop a deep learning approach that could generalise well to data generated with different parameters. A new dataset was generated by using three different PSFs to simulate the PET images. The noise level was also increased in this dataset, compared to the one used previously. Visual and quantitative analysis showed that the network could predict better defined tumour shapes and improved RCs. The CNN trained on simulated data was then applied to real phantom PET images reconstructed with different algorithms. The network yielded better defined shapes for all three phantom inserts. Improved estimates of the maximum and mean RC were measured using the images predicted by the CNN. This indicates that the proposed approach can be applied to PET images reconstructed with different algorithms.

In this chapter, the simulation of the PET images was somewhat simplified. We only simulated partial volume effect and noise, without considering many other effects that degrade PET image quality. As the ultimate goal would be to apply the proposed deep learning approach to real patient data, the 3D CNN needs to be trained on more realistic images that better represent the complexity of original PET scans. In the next chapter, the proposed approach is improved by training the network on more realistic PET images generated with an analytical simulator.

Chapter 3
Improving Uptake Quantification on PET Images Generated with an Analytical Simulator

The work described in the previous chapter is improved here by using a more complex and much more realistic simulated dataset to train and test the 3D CNN. This improved dataset is obtained by acting on two separate aspects. More realistic ground truth images are created using the 4D extended cardiac-torso (XCAT) phantom [150], which is composed of multiple organs and structures each assigned with a different radioactivity concentration. Secondly, an analytical simulator allows the simulation of PET raw data with more realistic noise and resolution properties. The pipeline of the method used in this chapter is presented in Fig. 3.1. At first, tumour shapes with different sizes and activity patterns are simulated. Subsequently, these tumours are placed in random positions within the left lung of the XCAT phantom to generate the ground truth images. An attenuation map is created for each ground truth image. The ground truth images and the corresponding attenuation maps are given as an input to the analytical simulator [151], which generates realistic PET raw data. The simulated PET images are reconstructed from the PET raw data using CASToR, an open source fully quantitative reconstruction

platform. The dataset composed of ground truth images and simulated PET images is given as an input to a 3D CNN, which is trained with the aim to recover the ground truth activity concentration and shape both for the tumour and for the background. The CNN trained on simulated images is tested on a set of unseen simulated images and then applied to real phantom data.

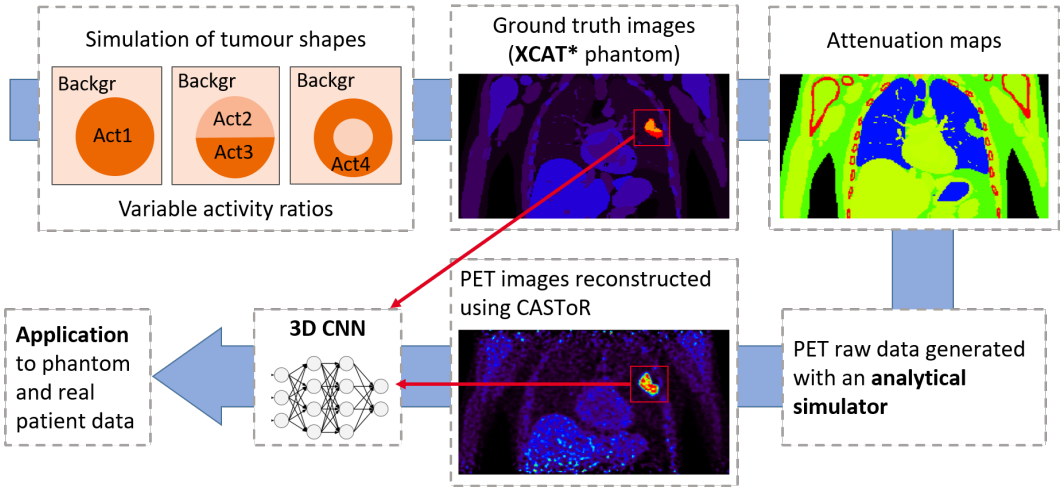


Figure 3.1: Overview of the deep learning approach. At first, tumour shapes with different shapes, sizes and activity patterns are simulated. These simulated tumours are placed in random positions in the left lung of an XCAT phantom to generate the ground truth images. The ground truth images and the corresponding attenuation maps are given as an input to an analytical simulator which generates the PET raw data. These data are subsequently reconstructed using the free software CASToR. The ground truth images and reconstructed PET images are used to train a 3D CNN, which is then applied to unseen simulated and phantom data.

Although different network architectures are tested in this chapter, they all have some common features. The basic CNN structure is shown in Fig. 3.2. The input images are processed by a convolutional layer with 32 filters of size 3×3×3 and assigned with ReLU activation. This is followed by a batch normalisation layer, forming a convolutional block. The first convolutional block is followed by N additional convolutional blocks, and the number N defines the depth of the CNN. Finally, a convolutional layer assigned with 1

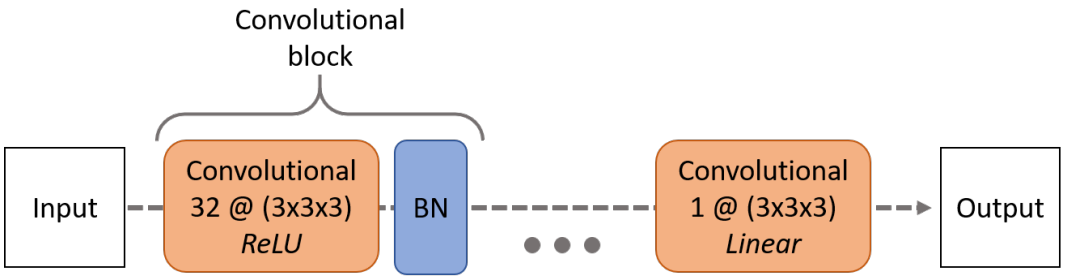


Figure 3.2: Illustration of the CNN architecture. The CNN is composed of a series of convolutional blocks, formed of a convolutional layer followed by batch normalisation layers. These convolutional layers are characterised by 32 filters with size 3×3×3 and by a ReLU activation function. A final convolutional layer assigned with 1 filter and with linear activation yields the output images.

filter of size 3×3×3 and with linear activation yields the output images. For simplicity, we henceforth express the depth of a CNN in terms of the number of convolutional blocks that compose its architecture. The next sections explain more in detail each step of the proposed method. In our experiments 65% of the simulated images are used for training, 15% for validation and the remaining 20% for testing. The loss function used to train the CNNs is always mean squared error.

3.1 Simulation of Images

The first step of our method is the simulation of tumour-like 3D shapes characterised by three different activity patterns. In this experiment, the simulated dataset is composed of 2100 tumour shapes with volumes spanning 0.01 to 200 ml. More specifically, there are 800 tumours assigned with a uniform activity distribution, 589 hollow tumours assigned with background activity concentration in the inner core and 721 tumours split in half where each half is assigned with a different radioactivity concentration. Activity concentrations ranging from 6 kBq/ml to 35 kBq/ml are assigned to the simulated tumour shapes. This range is chosen so that the tumours in the final simulated PET images

would have maximum activities comparable to the ones measured in a real patient ^{18}F-FDG PET NSCLC dataset. To generate the ground truth images, the simulated tumours are placed in random positions within the left lung of the 4D extended cardiac-torso (XCAT) phantom [150]. This is a digital anthropomorphic phantom which provides a model of human anatomy and physiology. In this experiment we use a prone male anatomy XCAT model composed of 800×800×1750 voxels with voxel size 1.01×1.01×1.01 mm^3, as shown in Fig. 3.3. This model includes various tissues and organs, each assigned with a different activity concentration based on average activity measurements obtained from a real non-small cell lung cancer (NSCLC) ^{18}F-FDG patient dataset. Attenuation maps are generated for each ground truth image, using the XCAT phantom attenuation values. The simulated tumours are assigned with an attenuation coefficient that belongs to the soft tissue attenuation range. In order to match the voxel size required by the analytical simulator, the images are rebinned using spline interpolation to a voxel size 2.09×2.09×2.03 mm^3 before being processed by the simulator. Furthermore, the rebinned images are cropped around the lungs to a final image size of 344×344×127 voxels to reduce the time needed for the simulation of PET raw data. An example of a cropped ground truth image together with the corresponding attenuation map is shown in Fig. 3.4. The PET raw data are generated using the analytical simulator developed by Stute *et al.* [151]. This simulator is programmed to replicate the geometry of Siemens mMR scanner (Siemens Biograph mMR, Erlangen, Germany). In this work, only the PET component of the simulator was modelled and the anatomical information from the MR was not used. Each PET image is generated with 100 million total prompt counts, including scatter and random events. The PET raw data are generated with an anisotropic, spatially-invariant PSF with FWHM = [4.5, 4.5, 4.0] mm^3. The PET images are reconstructed using the open source software CASToR, using the OSEM

algorithm with 6 iterations and with PSF modelling. The PSF used to reconstruct the PET images always matches the PSF used to generate the PET raw data. One of the aims of this work was to test if the CNN could generalise well to data acquired with different scanners. One way to simulate the differences that characterise different scanners is to generate and reconstruct PET images using a range of PSFs. As a result, a new dataset composed of 100 images (of which 33 uniform tumours, 34 tumours in halves and 33 hollow tumours) was generated using a range of anisotropic spatially-invariant PSFs. A value randomly drawn from a Gaussian distribution (mean μ= 4.5 mm, variance σ= 0.2 mm) was assigned to the transaxial components of the PSF, both in the simulation and in the reconstruction, with a perfect match. The axial component was calculated by dividing this value by 1.125 and a constant ratio between transaxial and axial PSF components was maintained.

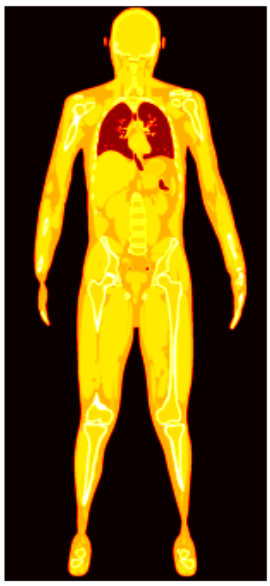
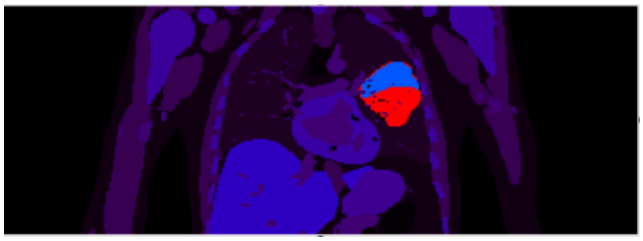
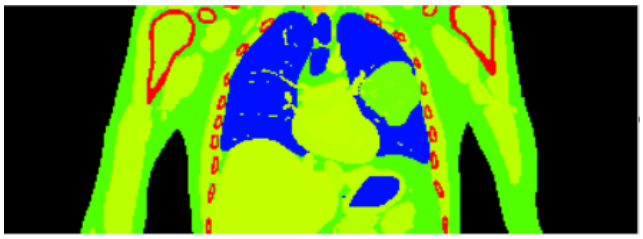


Figure 3.3: Coronal view of the XCAT phantom model. This image shows how different organs are modelled in the XCAT phantom.



(a) Ground truth image



(b) Attenuation map

Figure 3.4: This figure shows a representative ground truth image in (a) and its corresponding attenuation map in (b). These images are given as input to the analytical simulator, to generate realistic PET raw data.

3.2 Results

The dataset composed of 2100 pairs of ground truth images and reconstructed PET images is used to train and test a 3D CNN. Although the training dataset differs from the one used in the previous experiments, some of the results obtained in the previous chapter can be used as a starting point for the new simulations. We have learnt in previous experiments that training the network with Adam optimiser is more efficient than training the network with RMSprop and that including batch normalisation layers in the network architecture leads to a better performance. For this reason, in this chapter BN layers are incorporated in the CNN architecture and the CNNs are trained using Adam as optimiser. Other parameters such as network depth, number of epochs used for training and data normalisation methods are investigated more in detail and optimised specifically for this new dataset. In section 3.2.1, networks with

different depths are tested and the results are qualitatively and quantitatively assessed. The best performing network is trained for increasingly high numbers of epochs in section 3.2.3. Different data normalisation methods are tested in section 3.2.4. The results obtained with the best performing CNN are then compared to commonly used denoising and deconvolution methods in section 3.2.5. In section 3.2.6, the 3D CNN trained on images generated with a single PSF is tested on a set of PET images generated using a range of PSFs. Finally, the 3D CNN is applied to phantom data in section 3.2.7.

3.2.1 Comparison between CNNs composed of 5 to 7 convolutional blocks

The optimisation of the CNN starts from the choice of the network architecture. CNNs with different depths were analysed and a preliminary visual assessment indicated that a depth of 5 to 10 convolutional blocks would be sufficient to observed improved tumour shapes. In this section, we focus on three CNNs composed respectively of 5, 6 and 7 convolutional blocks. These networks are trained and tested on simulated PET data generated with a single PSF. The CNNs are trained for 200 epochs. Visual results obtained testing the three networks on a representative uniform tumour are presented in Fig. 3.5. The predicted image obtained using a CNN composed of 5 convolutional blocks, shown in Fig. 3.5 (c), is overall denoised and the tumour is characterised by sharper edges. In this image, as expected, the background is also improved and structures such as ribs are partly recovered. Although the predicted tumour is characterised by better defined edges and more uniform activity distribution, in this example a black area with low activity concentration is predicted by the network right above the tumour. As the dataset used for training and testing is more complex than the one used in previous experi-

ment, we expect that more layers may be needed to improve the performance. The image presented in Fig. 3.5 (e) is predicted by the CNN composed of 7 blocks. The network predicts a denoised image with sharper tumour edges but once again an area with low activity concentration is visible right above the tumour shape. The best results are yielded by the CNN composed of 6 blocks, which provides denoised images with better defined tumours, and a uniform background surrounding the tumour as shown in Fig. 3.5 (d). This network predicts the best images when applied to a heterogeneous tumour and to a hollow tumour, as shown in Fig. 3.6 and Fig. 3.7 respectively.

A quantitative analysis is performed after the visual assessment. An average MSSIM value of 0.46±0.06 is measured, over the total image area, between the ground truth images and the images predicted by the CNN composed of 5 blocks. Two identical MSSIM values equal to 0.50±0.05 are measured using the images predicted by the CNNs composed of 6 and 7 blocks. The analysis of the recovery coefficients is subsequently performed. The maximum RCs measured using the data predicted by the CNN composed of 6 blocks are plotted against each tumour volume in Fig. 3.8 (a). The recovery curves increase rapidly for tumours with volumes smaller than around 5 ml. An analogous behaviour is observed for the RC_{mean} and RC_{median} curves, shown in Fig. 3.8 (b) and (c).

During the network optimisation, we focus on the quantitative results estimated only for volumes larger than 5 ml. A separate analysis of the small tumours RCs predicted by the optimised network is presented later in this chapter. A summary of all the RCs measured for the three networks and averaged over the tumour volumes in the test dataset that exceed 5 ml is presented in table 3.1. As shown in the first row, the CNN made of 5 blocks yields improved estimates of RC_{max} but it does not recover accurately the mean and median RC. The CNN composed of 6 blocks improves the estimation of RC_{mean}

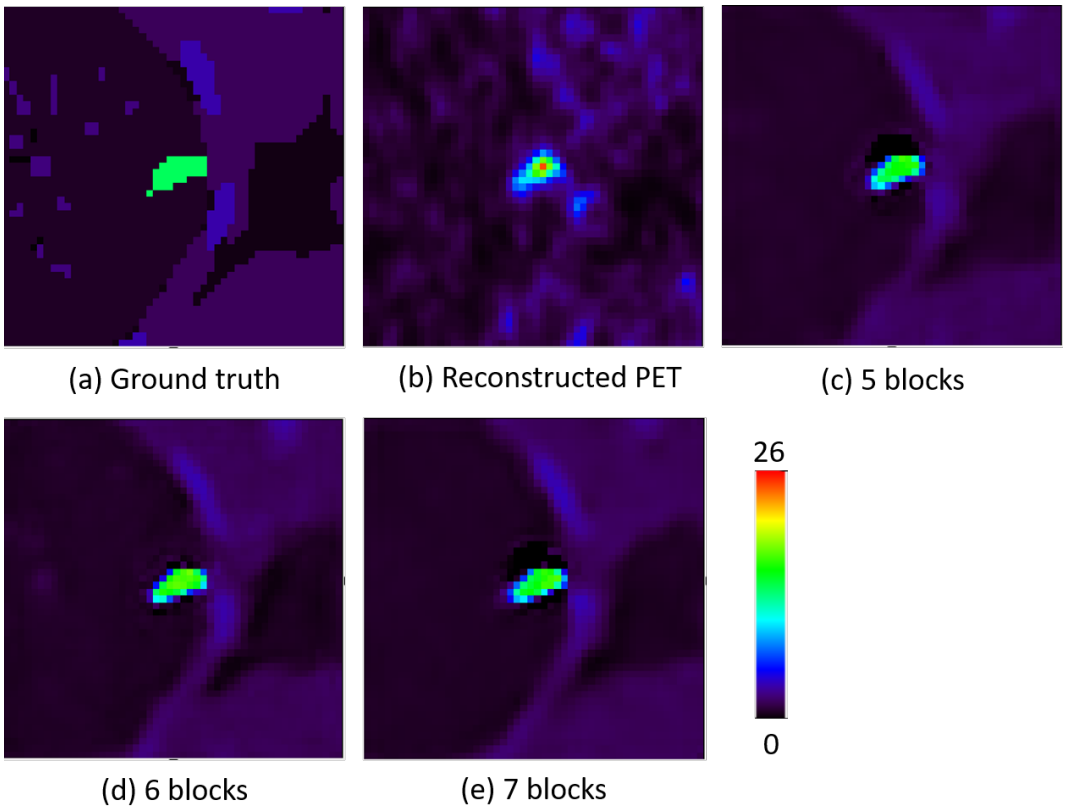


Figure 3.5: Transverse views of a uniform tumour. The ground truth image is shown in (a) and the reconstructed PET image is shown in (b). The image yielded by the CNN composed of 5 convolutional blocks is presented in (c). The images predicted by the CNNs composed of 6 and 7 convolutional blocks are presented in (d) and (e) respectively. The best result is yielded by the CNN composed of 6 blocks, which provides denoised images with better defined tumours, and a uniform background surrounding the tumour. The images are all shown using the same colour scale which is expressed in kBq/ml.

and RC_{median} yielding values close to 1. Further increasing the number of convolutional blocks to 7 does not improve the estimated RC_{mean} and RC_{median}, as shown in the third row of the table. Although all three CNNs yield improved RC_{max} estimates, the best measurement corresponds to the network made of 7 blocks. Taking into account the observations derived from the visual assessment and the evaluations obtained with the quantitative analysis, the CNN composed of 6 convolutional blocks has overall a better performance on this

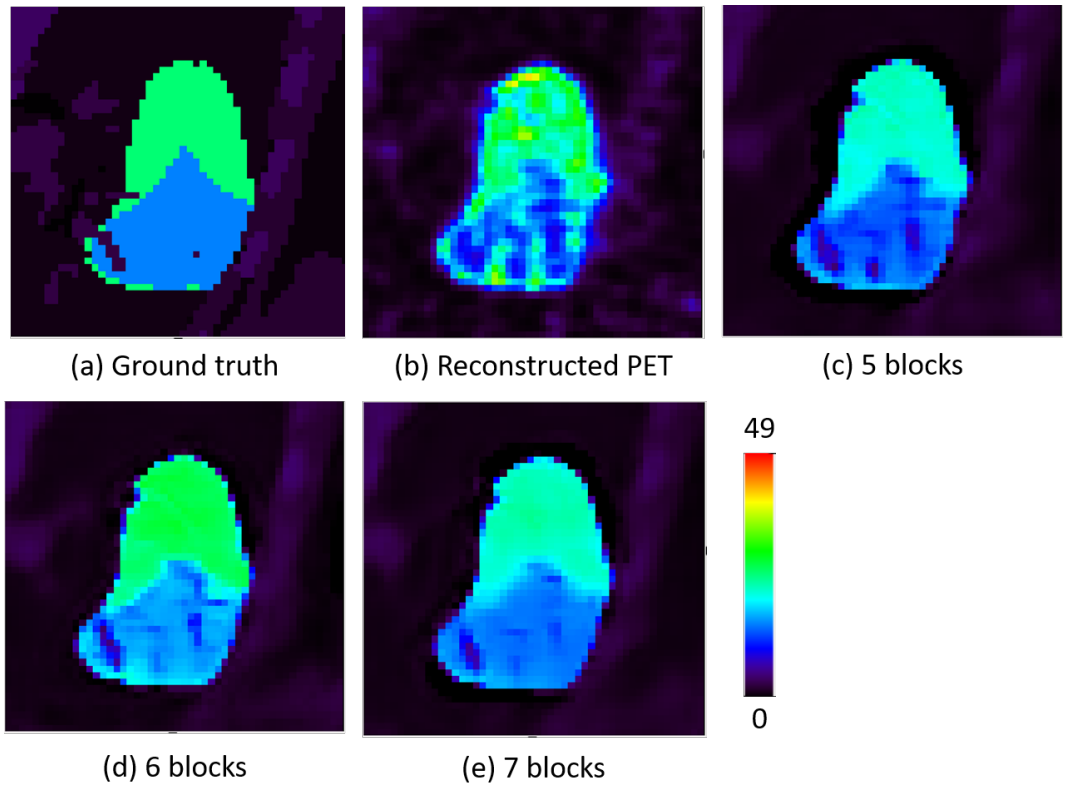


Figure 3.6: Coronal views of a heterogeneous tumour. The ground truth image is shown in (a) and the reconstructed PET image is shown in (b). The image yielded by the CNN composed of 5 convolutional blocks is presented in (c). The images predicted by the CNNs composed of 6 and 7 convolutional blocks are presented in (d) and (e) respectively. The best result is yielded by the CNN composed of 6 blocks, which provides denoised images with better defined tumours, and a uniform background surrounding the tumour. The images are all shown using the same colour scale which is expressed in kBq/ml.

dataset.

3.2.2 Small Tumours Analysis

As shown in Fig. 3.8, the RCs are not well recovered for tumour volumes smaller than 5 ml. In this section, the small tumours predicted by the CNN composed of 6 blocks and trained for 200 epochs are analysed. Again, some simulated small tumours present an elongated shape and thus are more affected

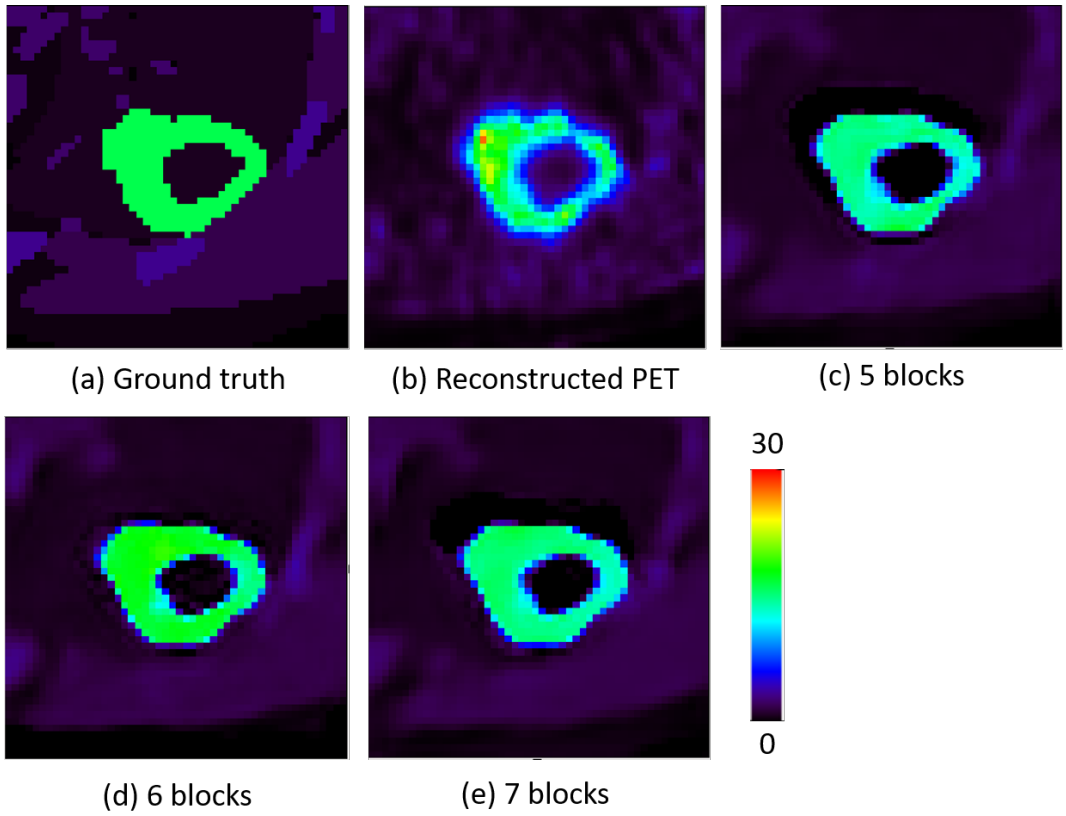


Figure 3.7: Transverse views of a hollow tumour. The ground truth image is shown in (a) and the reconstructed PET image is shown in (b). The image yielded by the CNN composed of 5 convolutional blocks is presented in (c). The images predicted by the CNNs composed of 6 and 7 convolutional blocks are presented in (d) and (e) respectively. The best result is yielded by the CNN composed of 6 blocks, which provides denoised images with better defined tumours, and a uniform background surrounding the tumour. The images are all shown using the same colour scale which is expressed in kBq/ml.

by partial volume effect. The tumour sphericity, defined in Eq. 2.5, is measured to assess if there is any correlation between the tumour shape and the network's performance. Fig. 3.9 shows the RC_{max} values predicted by the 3D CNN, plotted for tumours with volume smaller than 5 ml. In this plot, the RC values are colour-coded based on the tumour sphericity. This plot suggests that worse RC_{max} estimates are correlated with less spherical tumour shapes. As the RC curve is rapidly changing, an average RC would not describe well

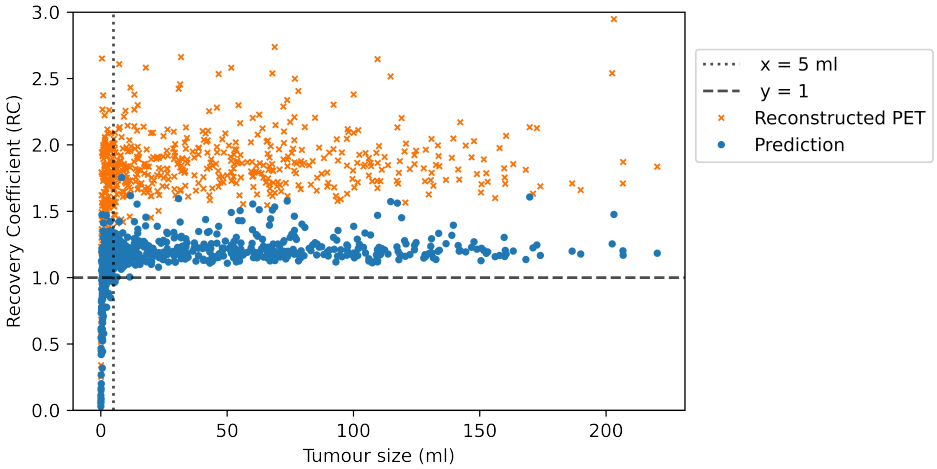
convolutional blocks	RC_{max}	RC_{median}	RC_{mean}
5	1.11±0.09	0.87±0.03	0.82±0.03
6	1.22±0.10	**1.03±0.04**	**0.98±0.04**
7	**1.05±0.10**	0.91±0.05	0.87±0.05
Reconstructed PET	1.87±0.21	0.87±0.07	0.86±0.05

Table 3.1: The RCs measured for CNNs characterised by different depths are compared here. The bottom row shows the RC values measured using the reconstructed PET images and the ground truth images. The best RC_{max} estimate is yielded by the CNN composed of 7 convolutional blocks, whereas the best RC_{median} and RC_{mean} estimates are given by the CNN composed of 6 convolutional blocks. These results are measured for tumour volumes larger than 5 ml.

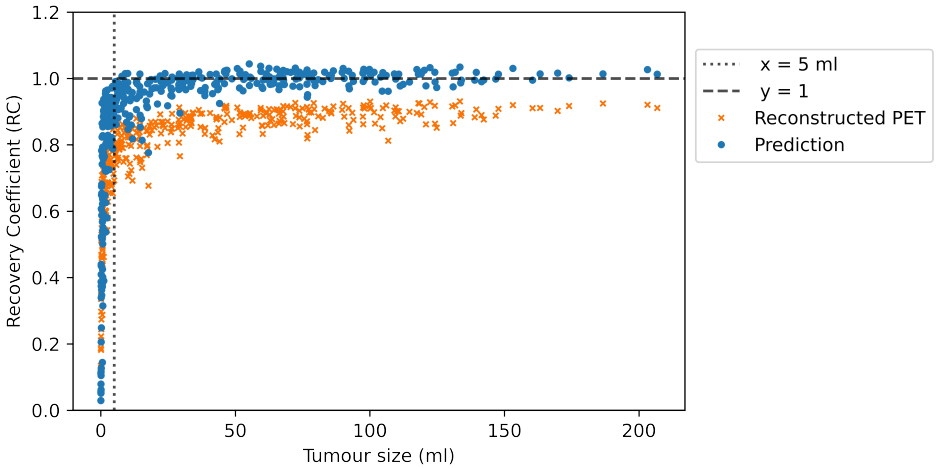
the plots in Fig. 3.9. We provide a quantitative description of a small part of this curve, only taking into consideration the tumour volumes that lay in the interval spanning from 1 ml to 2 ml. The chosen 1 ml interval is split in four smaller segments, then the average RC is measured for each segment. Subsequently, an overall RC is obtained for the 1 ml interval by averaging these four RCs. The maximum, median and mean RCs are measured both using the reconstructed PET images and the CNN predictions, and the results are summarised in table 3.2. The table shows that the CNN trained on non-normalised data yields improved RCs for small tumours, compared to the RCs measured using the reconstructed PET images.

1 ml <V <2 ml	RC_{max}	RC_{median}	RC_{mean}
CNN	**1.11±0.01**	**0.90±0.05**	**0.89±0.20**
Reconstructed PET	1.65±0.05	0.63±0.02	0.78±0.16

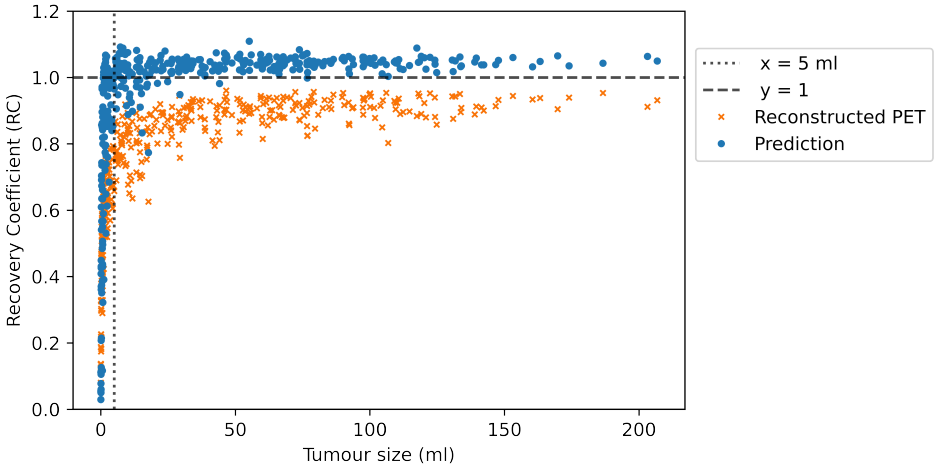
Table 3.2: The average RCs measured for tumours with volumes comprised between 1 ml and 2 ml are presented here. The CNN predicts improved RCs.



(a) RC_{max}



(b) RC_{mean}



(c) RC_{median}

Figure 3.8: RC curves for the CNN made of 6 blocks. The RC_{max} is plotted against the tumour volume in (a). The RC_{mean} and RC_{median} are plotted against the tumour volume in (b) and (c), respectively. All RCs are not well recovered for tumour volumes smaller than around 5 ml. The CNN made of 6 blocks provides improved RC estimates for tumour volumes bigger than 5ml.

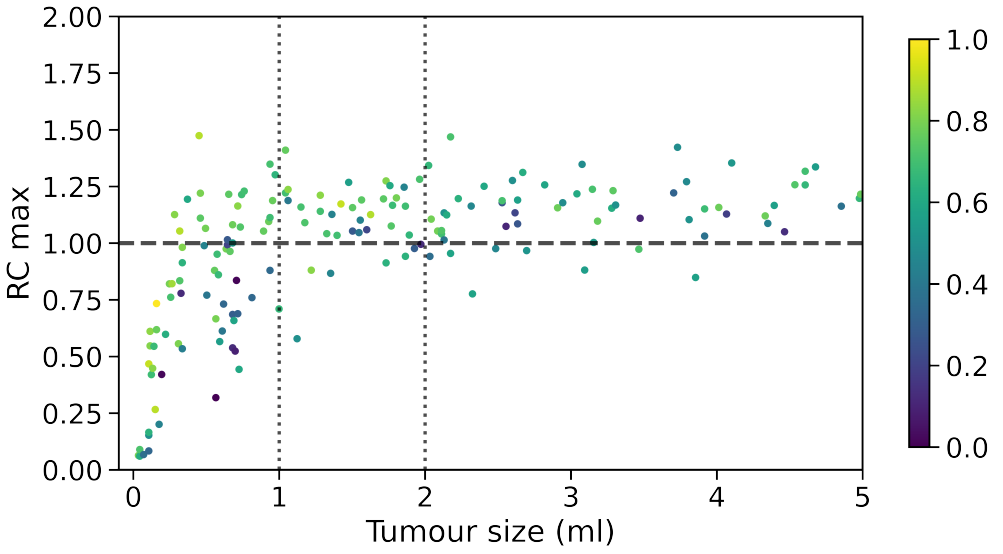


Figure 3.9: The RC_{max} measured using the CNN's predicted images is plotted here against the tumour volume, for tumours smaller than 5 ml. The tumour sphericity is colour-coded so that more spherical tumours are represented in yellow, whereas less spherical tumours are shown in blue. Worse RC_{max} estimates are correlated with lower tumour sphericity.

3.2.3 Impact of the Number of Epochs Used for Training

In this section, the 3D CNN with 6 convolutional blocks is trained with increasingly high numbers of epochs in order to assess the impact of this parameter on the CNN's predictions. Visual results obtained training the same CNN for 100, 200, 300 and 500 epochs are presented in Fig. 3.10. Fig. 3.10 (c) is obtained by training the CNN for 100 epochs. This image is denoised, but the CNN predicts an area of low activity concentration right above the tumour shape, where the ground truth image presents a uniform background. The networks trained for 200 and 300 epochs yield the images presented in Fig. 3.10 (d) and Fig. 3.10 (e) respectively. These images are denoised and in both cases the tumours present better defined edges. Further increasing the number of

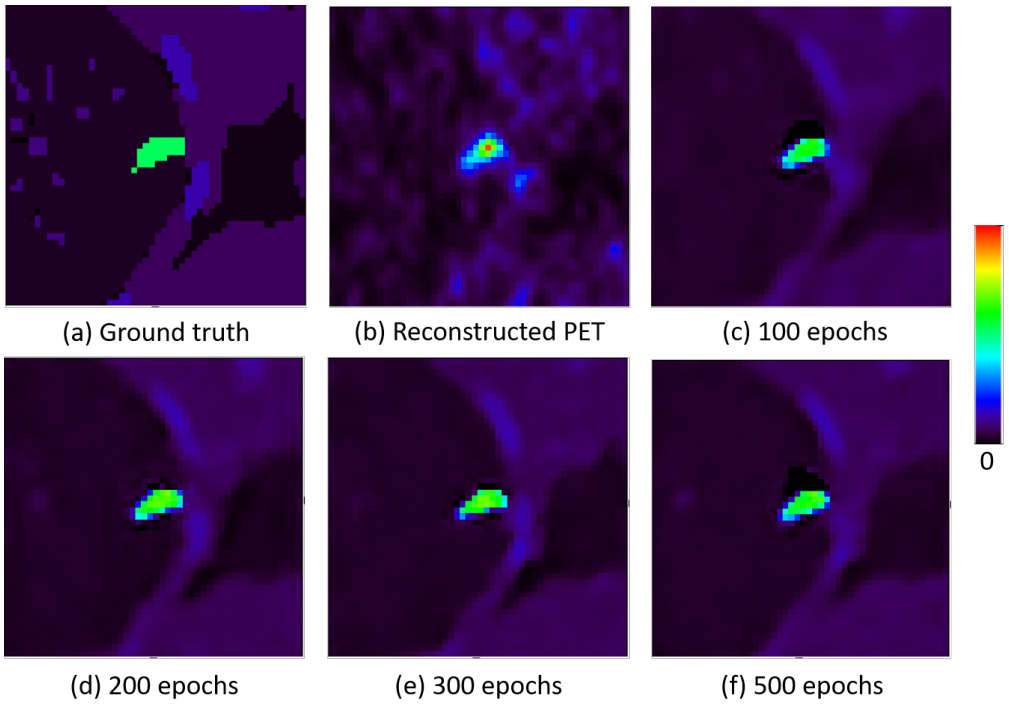


Figure 3.10: Transverse views of a uniform tumour. The ground truth image is shown in (a) and the reconstructed PET image is shown in (b). The images yielded by the CNN trained for 100 epochs are shown in (c). The same CNN is trained for increasingly high number of epochs, and the results are shown in images (d), (e) and (f). The networks trained for 200 and 300 epochs yield the best images. The images are shown using the same colour scale which is expressed in kBq/ml.

epochs up to 500 does not improve the image quality. In fact, in Fig. 3.10 (f) a black area with low activity concentration is predicted right above the tumour shape. The same effects can be observed when the network is applied to a hollow tumour, as shown in Fig 3.11. Table 3.3 summarises the average RC_{max}, RC_{median} and RC_{mean} measured for tumour volumes exceeding 5 ml. All networks yield an improved RC_{max}, the best result is achieved by training the CNN for 100 epochs. The best estimates of RC_{median} and RC_{mean} are yielded by the network trained for 200 epochs. Further increasing the number of epochs used for training results in a decrease of mean and median RCs. Considering both the visual and quantitative assessment, the 3D CNN with

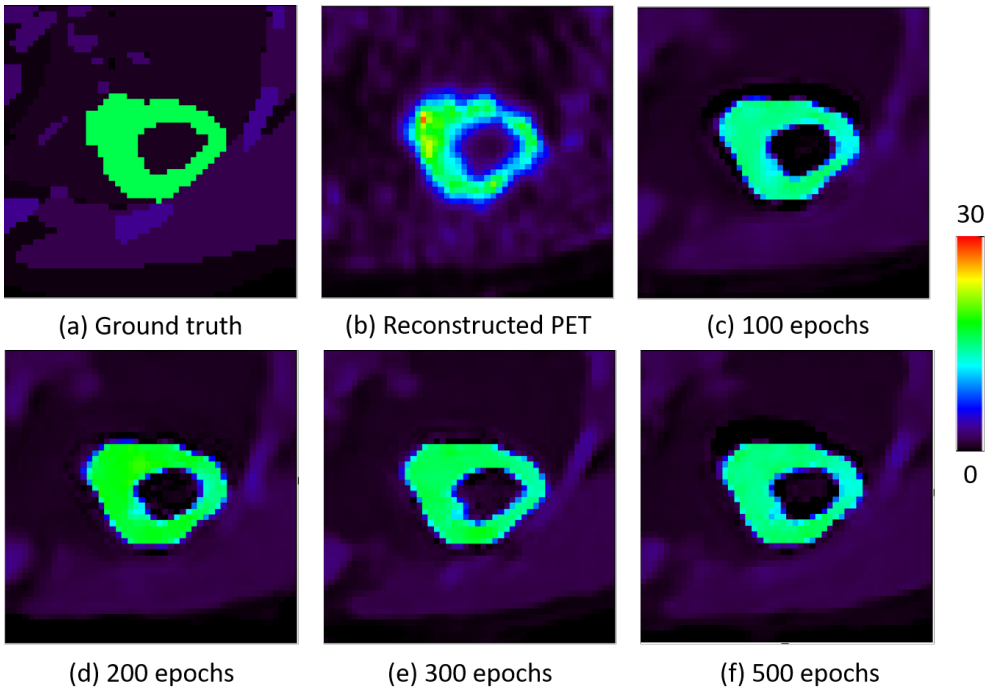


Figure 3.11: Transverse views of a hollow tumour. The ground truth image is shown in (a) and the reconstructed PET image is shown in (b). The images yielded by a CNN trained for 100 epochs are shown in (c). The same CNN is trained for increasingly high number of epochs, and the results are shown in images (d), (e) and (f). The best result in terms of shape definition is yielded by the CNN trained for 200 epochs. The networks trained for 200 and 300 epochs yield the best images. The images are shown using the same colour scale which is expressed in kBq/ml.

6 convolutional blocks trained for 200 epochs performs best on this specific dataset. These settings are used in the next sections for further experiments.

3.2.4 Impact of Different Data Normalisation Methods

In this section, different normalisation methods are applied to the reconstructed PET images before training and the results are compared. The same CNN composed of 6 blocks and trained for 200 epochs was used in all the experiments presented in this section. The normalisation methods used in this chapter are the same as the ones presented in section 2.2.3.

Epochs	RC_{max}	RC_{median}	RC_{mean}
100	**1.06±0.10**	0.86±0.03	0.81±0.03
200	1.22±0.10	**1.03±0.04**	**0.98±0.04**
300	1.11±0.08	0.94±0.03	0.90±0.03
500	1.11±0.09	0.88±0.03	0.84±0.03
Reconstructed PET	1.87±0.21	0.87±0.07	0.86±0.05

Table 3.3: The RCs measured for CNNs trained using increasingly high numbers of epochs are analysed here. The bottom row shows the RC values measured using the reconstructed PET images. The best RC_{max} is obtained training the CNN for 100 epochs. The best RC_{median} and RC_{mean} are predicted by the CNN trained for 200 epochs. These results are measured for tumour volumes larger than 5 ml.

Fig. 3.12 shows a comparison between transverse views of the same tumour assigned with uniform activity distribution, obtained pre-processing the reconstructed PET images with different normalisation methods. In Fig. 3.12 the ground truth image and the reconstructed PET image are shown in (a) and (f). The image predicted by the 3D CNN trained on non-normalised data (b) is denoised and the predicted tumour has sharper edges and a more uniform activity distribution. The image in (c) is predicted by the CNN trained on a dataset normalised using the 'sklearn' normalise function, that scales the input vectors individually to unit norm. This image is denoised and the tumour shape is better defined, but an area characterised by low activity concentration is predicted above the tumour. The image in (d), obtained training the CNN on a dataset where each image is normalised by its maximum value, is characterised by blurrier tumour edges than the ones obtained using the non-normalised data. Furthermore, some low activity voxels are predicted around the tumour edges. Finally, the image in (e) was obtained training and testing the CNN on a dataset normalised by dividing each image by the maximum activity concentration measured over the whole dataset. In this case the predicted tumour has sharper edges and it is less blurry than the tumours in

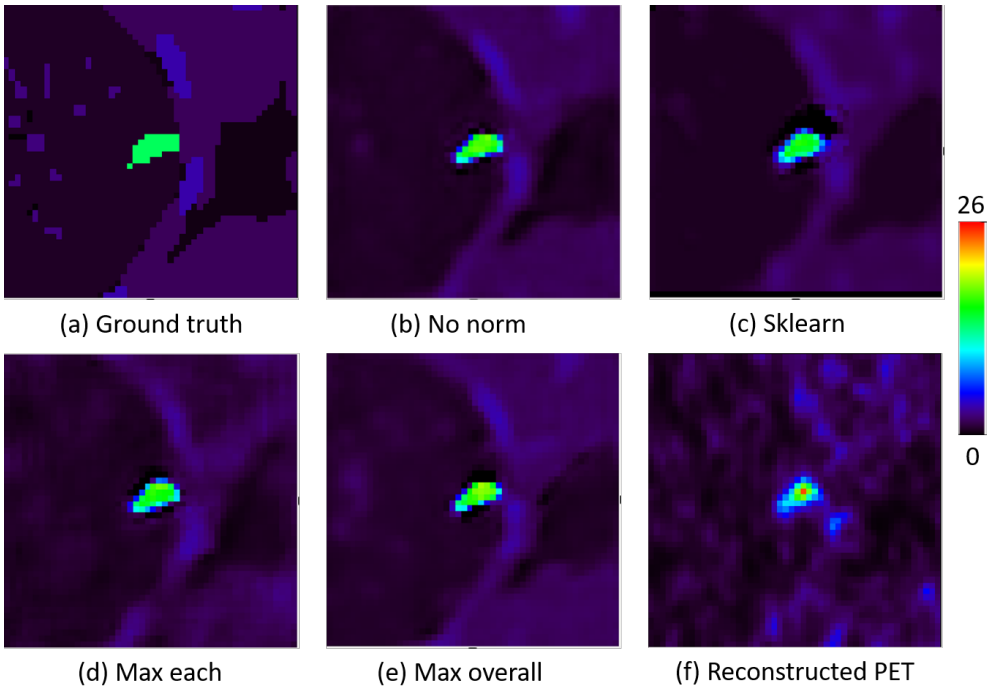


Figure 3.12: Transverse views of a uniform tumour. The ground truth image and the reconstructed PET image are shown in (a) and (f). In (b), the CNN was trained and tested on non-normalised data. The image in (c) was obtained normalising the PET images with the sklearn normalisation function. The image in (d) was obtained normalising each PET image by its maximum value. The image in (e) was obtained normalising the PET images by the maximum activity concentration overall. The best predictions are obtained when using non-normalised data and when using the maximum overall activity concentration as normalisation factor.

(c) and (d). Similar observations can be made looking at Fig. 3.13 and Fig. 3.14, which present the same comparison for a heterogeneous tumour and for a hollow tumour. Low activity areas are predicted when the sklearn function is used and when each image is normalised using its maximum value as normalisation factor. The images that more closely resemble the ground truth are obtained when the CNN is trained on non-normalised data or when using the maximum intensity recorded in the whole training dataset as a normalisation factor. The visual assessment is followed by a quantitative analysis. The same MSSIM value, equal to 0.50 ± 0.06, is obtained both for the non-normalised

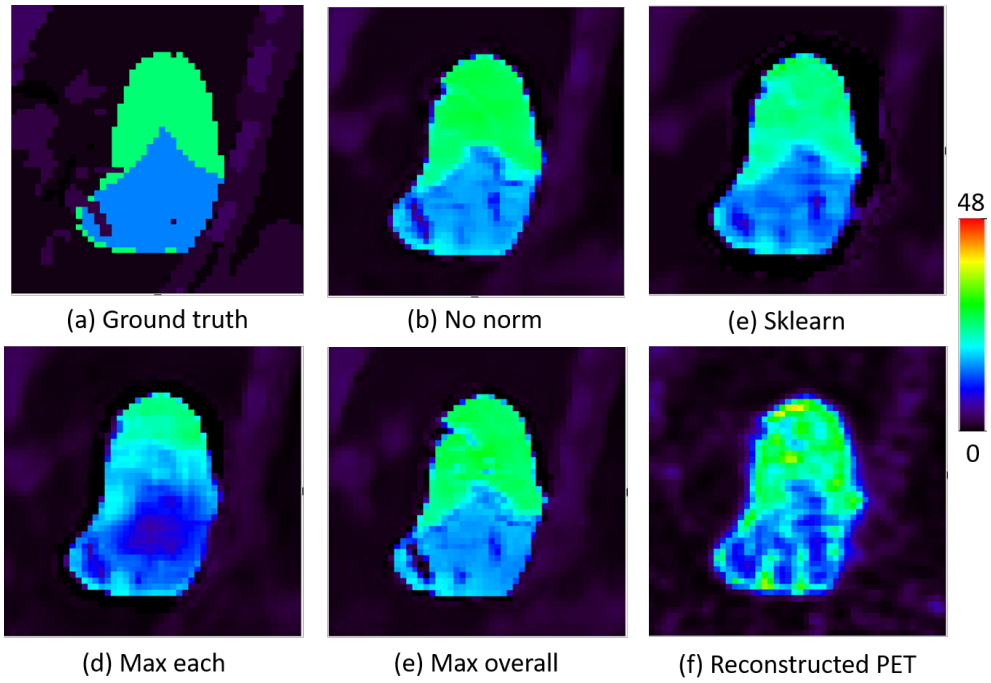


Figure 3.13: Transverse views of a heterogeneous tumour. The ground truth image and the reconstructed PET image are shown in (a) and (f). In (b), the CNN was trained and tested on non-normalised data. The image in (c) was obtained normalising the PET images with the sklearn normalisation function. The image in (d) was obtained normalising each PET image by its maximum value. The image in (e) was obtained normalising the PET images by the maximum activity concentration overall. The best predictions are obtained when using non-normalised data and when using the maximum overall activity concentration as normalisation factor.

dataset and for the data normalised using the overall maximum intensity as normalisation factor. The MSSIM values obtained using other normalisation methods are equal or lower than 0.45. The maximum, median and mean RC averaged over the test dataset are presented in table 3.4. The non-normalised dataset and the dataset normalised using the overall maximum intensity lead to very similar results. The CNN trained on non-normalised data yields the best RC_{max} and RC_{mean} estimates.

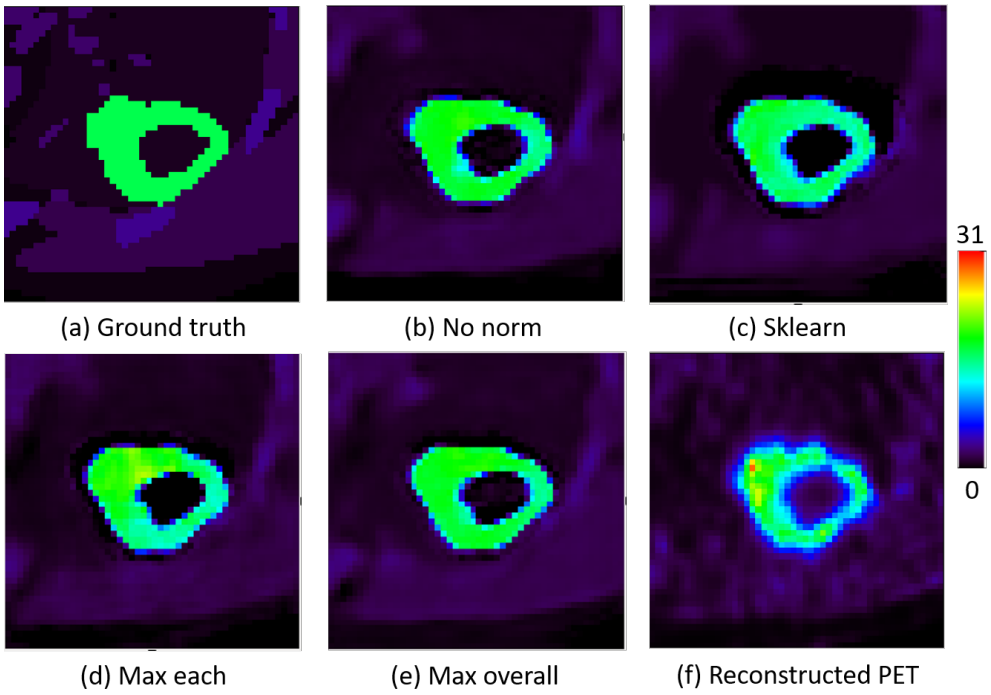


Figure 3.14: Transverse views of a hollow tumour. The ground truth image and the reconstructed PET image are shown in (a) and (f). In (b), the CNN was trained and tested on non-normalised data. The image in (c) was obtained normalising the PET images with the sklearn normalisation function. The image in (d) was obtained normalising each PET image by its maximum value. The image in (e) was obtained normalising the PET images by the maximum activity concentration overall. The best predictions are obtained when using non-normalised data and when using the maximum overall activity concentration as normalisation factor.

3.2.5 Comparison with Standard Deconvolution and Denoising Methods

The results given by the 3D CNN trained on non-normalised data are finally compared to other methods that are standardly used to denoise and deconvolve PET images. Table 3.5 illustrates the RCs measured after applying various post-processing methods to the reconstructed PET images. The results in the first row are measured after applying a Gaussian filter with FWHM = [2.25, 2.25, 2.00] mm^3 to the reconstructed PET images. Two more filters

Normalisation	RC_{max}	RC_{median}	RC_{mean}
None	**1.22±0.10**	1.03±0.04	**0.98±0.04**
Individual max	1.28±0.19	0.91±0.07	0.88±0.07
Overall max	1.23±0.09	**1.02±0.04**	0.96±0.04
Sklearn	1.28±0.10	0.95±0.04	0.90±0.04
Reconstructed	1.87±0.21	0.87±0.07	0.86±0.05

Table 3.4: Average RCs yielded by a CNN trained and tested on data normalised with different methods. These results are relative to tumour volumes larger than 5 ml. The best RC_{median} and RC_{mean} are predicted by the CNN trained on non-normalised data. The best RC_{max} is obtained when normalising simulated PET images by the overall maximum activity concentration.

with FWHM = [4.5, 4.5, 4.0] mm^3 and FWHM = [6.75, 6.75, 6.00] mm^3 are applied to the same PET images for further comparison, and the resulting RCs are shown in the second and third row respectively. The Gaussian filter with FWHM = [6.75, 6.75, 6.00] mm^3 provides the best RC_{max} measurement. Gaussian filtering leads to underestimated RC_{mean} and RC_{median}, and in one case these RCs are worse than the ones measured using the reconstructed PET images. Similarly, the application of a median filter with filter size 3×3×3 only improves the RC_{max}, which is anyway worse than the one provided by the CNN. Lucy-Richardson deconvolution does not improve the RC_{max}, but it provides improved RC_{mean} and RC_{median} compared to the ones obtained using Gaussian filters. Applying a Gaussian filter with FWHM = [4.5, 4.5, 4.0] after Lucy-Richardson deconvolution does not significantly improve the RC measurements. We can conclude that overall the 3D CNN outperforms the standard image post-processing methods of Gaussian smoothing, Lucy-Richardson deconvolution, and a combination of the two.

A visual comparison between the best performing methods is shown in Fig. 3.15, using a hollow tumour as a representative case. The images shown in Fig. 3.15 (a) and (b) correspond to the ground truth image and to the

	RC_{max}	RC_{median}	RC_{mean}
Gauss 2.25 mm	1.70±0.18	0.87±0.07	0.85±0.05
Gauss 4.5 mm	1.29±0.11	0.81±0.06	0.84±0.09
Gauss 6.75 mm	**1.18±0.08**	0.82±0.10	0.79±0.07
Median	1.29±0.09	0.85±0.09	0.80±0.05
LR	2.51±0.29	0.90±0.07	0.89±0.04
LR + Gauss 4.5 mm	1.48±0.33	0.88±0.08	0.84±0.06
CNN	1.22±0.10	**1.03±0.04**	**0.98±0.04**
Reconstructed	1.87±0.21	0.87±0.07	0.86±0.05

Table 3.5: This table shows average RCs obtained by applying standard post-processing methods to the reconstructed PET images. The methods used for post-processing include three different Gaussian filters (rows 1, 2 and 3), a median filter (row 4) and LR deconvolution with 5 iterations (row 5). The results in row 6 are obtained by applying LR deconvolution to the reconstructed PET images, and subsequently a Gaussian filter. The RCs measured using the CNN predictions are shown in row 7 and the measurements extracted from the reconstructed PET images are shown in the bottom row. The CNN provides the best RC_{mean} and RC_{median}. The best RC_{max} is obtained when applying the 6.75 Gaussian filter and the RC_{max} estimate provided by the CNN is second best. These results are measured for tumour volumes larger than 5 ml.

reconstructed PET image. The image predicted by the CNN trained on non-normalised data, shown in (c), has sharper edges than all the images presented in the bottom row, which are obtained with standard post-processing methods. Compared to the standard techniques, the CNN can predict a more uniform tumour activity distribution and it improves the definition of the tumour and of some background structures.

3.2.6 Application to Data Generated with Different PSFs

The CNN trained on non-normalised data is applied to a dataset composed of reconstructed PET images generated with an analytical simulator using a

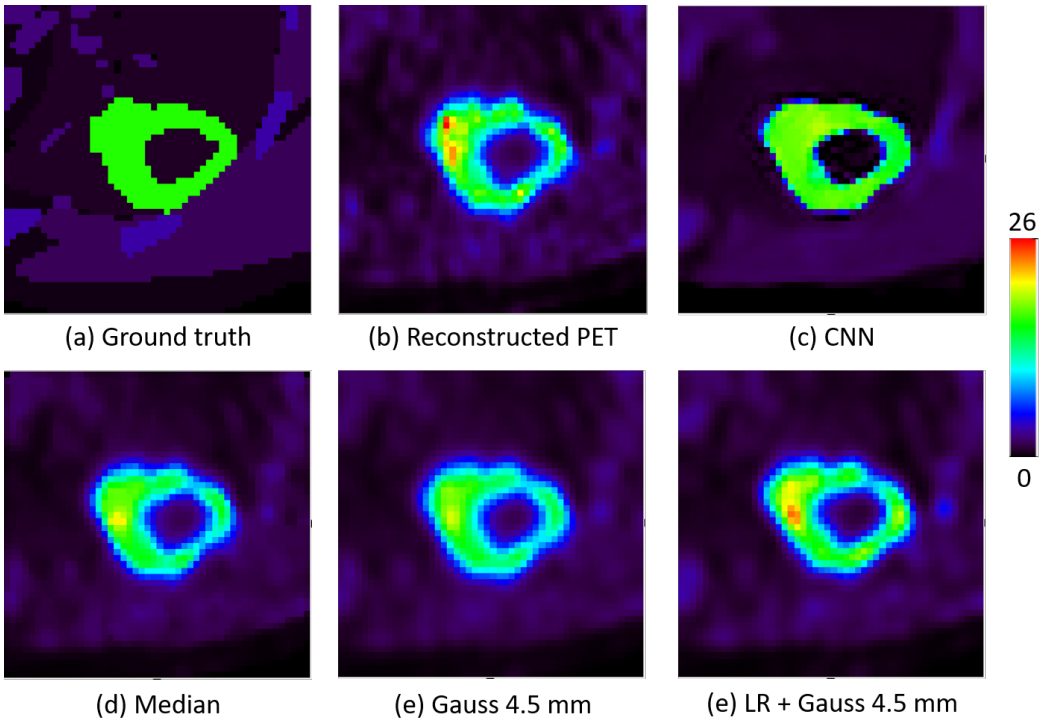


Figure 3.15: Comparison between images of a hollow tumour obtained with the proposed deep learning approach and with other post-processing methods. The ground truth image and reconstructed PET image are shown in (a) and (b). The image in (c) is predicted by a 3D CNN trained on non-normalised data. In (d) and (e) a median filter and a Gaussian filter with FWHM = [4.5, 4.5, 4.0] mm^3 are respectively applied to the reconstructed PET image. The image in (e) is obtained by applying LR deconvolution to the reconstructed PET image and by subsequently applying a Gaussian filter with FWHM = [4.5, 4.5, 4.0] mm^3. The images are shown with the same colour scale expressed in kBq/ml.

range of PSFs. Three representative tumours belonging to this dataset are shown in Fig. 3.16. The reconstructed PET image of the hollow tumour in the top row is generated and reconstructed using a PSF with FWHM = [4.30, 4.30, 3.82] mm^3. A PSF with FWHM = [4.00, 4.00, 3.55] mm^3 is used to generate the PET image in the second row and a PSF with FWHM = [4.70, 4.70, 4.18] mm^3 is used to generate the PET image of the uniform tumour in the bottom row. In all three cases the images predicted by the CNN are denoised

and the tumours are characterised by sharper edges and by a more uniform activity distribution. The average MSSIM measured using the reconstructed PET images is 0.34±0.05, whereas the average MSSIM measured using the CNN predictions is 0.47±0.07. The RCs presented in table 3.6 show that the CNN improves the estimates of maximum, mean and median RCs when applied to data generated with a range of PSFs that were not learnt during training.

V ≥ 5 ml	RC_{max}	RC_{median}	RC_{mean}
Reconstructed PET	1.95±0.25	0.84±0.08	0.84±0.06
PSF data	**1.23±0.10**	**1.01±0.05**	**0.95±0.06**

Table 3.6: The first row of this table summarises the average RCs extracted from the dataset composed of PET images generated with a range of PSFs. The second row shows the corresponding values measured using the images predicted by the CNN. Tumour volumes smaller than 5 ml are excluded from this analysis.

Tumours with volumes smaller than 5 ml are analysed separately. In Fig. 3.17 the RC_{max} values predicted by the CNN are plotted against tumour volume for tumour volumes smaller than 5 ml. The tumour sphericity is colour-coded in this plot, so that more spherical tumour shapes are shown in yellow and less spherical tumours are shown in dark blue. This plot suggests that tumours with low sphericity are associated to lower RC_{max} values, especially when tumours are smaller than 2 ml. We perform a quantitative evaluation of the RC curves for tumours with volumes comprised between 1 ml and 2 ml. This 1 ml interval is split in four smaller segments, then the average RC is measured for each segment. Subsequently, an overall RC value is obtained for the 1 ml interval by averaging these four RCs. The maximum, median and mean RCs are measured both using the reconstructed PET images and the CNN predictions, and the average results are summarised in table 3.7. The table shows that the CNN trained on non-normalised data, generated

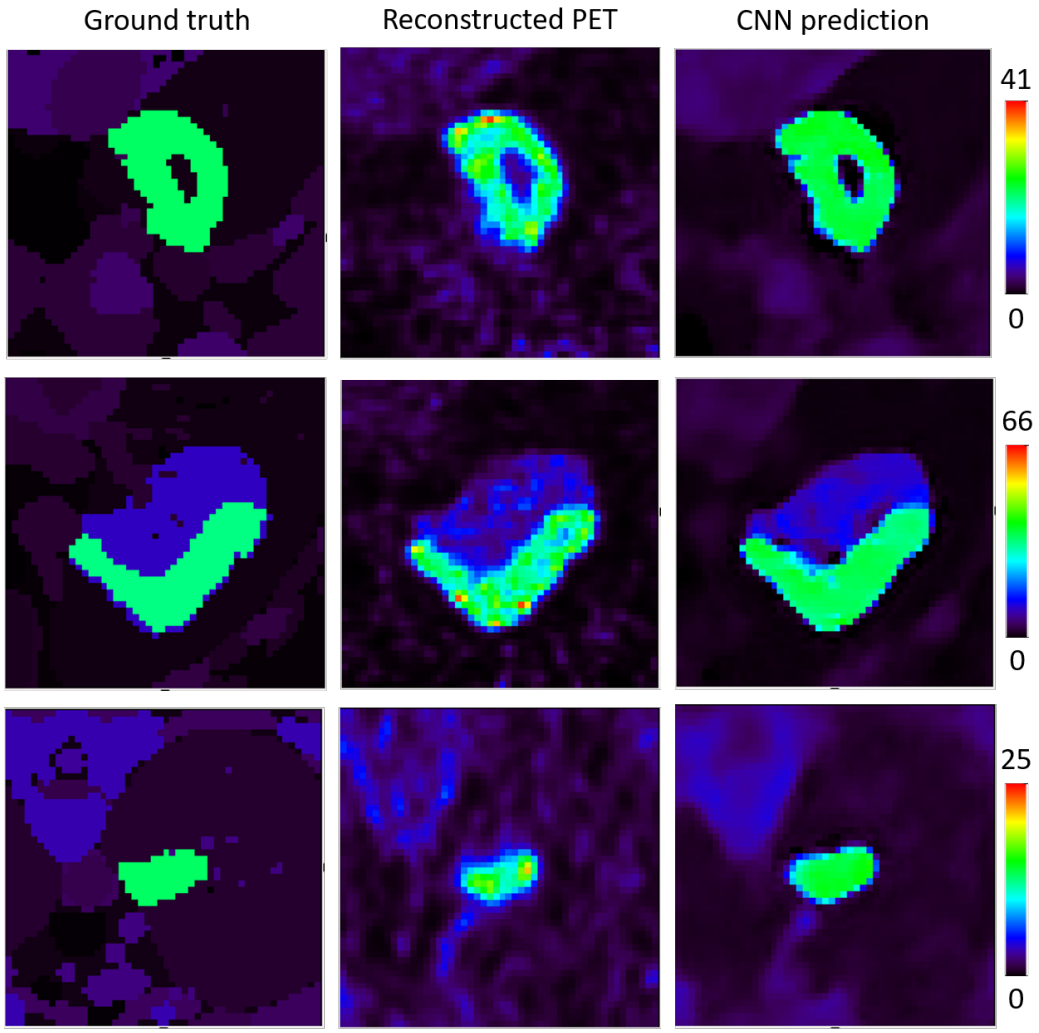


Figure 3.16: Views of three representative volumes belonging to the dataset generated with multiple PSFs. The PET image in the first row is generated and reconstructed using a PSF with FWHM = [4.30, 4.30, 3.82] mm^3. A PSF with FWHM = [4.00, 4.00, 3.55] mm^3 is used to generate and reconstruct the PET image in the second row. A PSF with FWHM = [4.70, 4.70, 4.18] mm^3 is used to generate and reconstruct the PET image in the bottom row. The CNN, trained on PET images generated and reconstructed using a PSF with FWHM = [4.5, 4.5, 4,0] mm^3, predicts the results shown in column 3. All images belonging to the same row are shown with the same colour scale expressed in kBq/ml.

and reconstructed using one PSF, yields improved RCs when applied to small tumours in PET images generated and reconstructed using a range of PSFs.

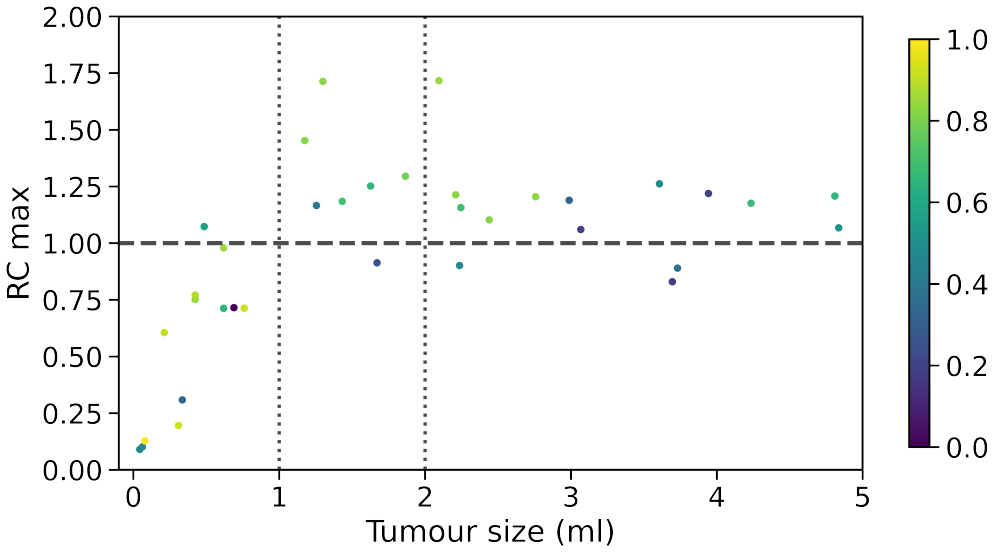


Figure 3.17: The RC_{max} measured using the CNN's predicted images is plotted here against the tumour volume, for tumours smaller than 5 ml. The tumour sphericity is colour-coded so that more spherical tumours are represented in yellow, whereas less spherical tumours are shown in blue. Worse RC_{max} estimates are correlated with lower tumour sphericity.

1 ml $\leq$ V $\leq$ 2 ml	RC_{max}	RC_{median}	RC_{mean}
Reconstructed PET	2.03±0.21	0.59±0.05	0.69±0.08
PSF data	**1.30±0.14**	**0.92±0.14**	**0.82±0.09**

Table 3.7: The first row of this table summarises the average RCs extracted from the dataset composed of PET images generated with a range of PSFs, for tumours with volumes comprised between 1 and 2 ml. The second row shows the corresponding values measured using the images predicted by the CNN. Improved RCs are predicted by the CNN.

3.2.7 Application to Phantom Data

The CNN trained on non-normalised data is finally applied to 50×50×50 voxels regions cropped from phantom PET scans reconstructed with different algorithms. Each of these regions includes one of the three tumour-like phantom inserts described in section 2.3. The three PET images reconstructed with OSEM algorithm are shown in Fig. 3.18, together with the corresponding

CNN predictions. The CNN provides denoised images and better defined tumours. The CNN also predicts a more uniform activity distribution within the tumour shapes. Similar improvements can be observed when the CNN is applied to images reconstructed with PSF modelling, as shown in Fig. 3.19. This result suggests that the proposed method can improve the tumour shapes and activity distributions when applied to phantom PET images reconstructed with different algorithms. The RC_{mean} and RC_{median} are not measured for the phantom images as a tumour segmentation mask was not available. Table 3.8 summarises the RC_{max} values averaged over the three tumour inserts. The CNN predicts improved RC_{max} for all the different reconstruction algorithms. The best results are obtained when applying the CNN on PET images reconstructed using ToF algorithm and OSEM. We expect the CNN to have a good performance on OSEM reconstructed images as this is the algorithm that we used to reconstruct the simulated PET images used for training.

RC_{max}	CNN prediction	PET images
OSEM	1.03±0.11	1.78±0.04
PSF + ToF	1.09±0.16	1.32±0.08
PSF	1.10±0.17	1.37±0.04
ToF	**0.98±0.11**	1.65±0.15

Table 3.8: RC_{max} values measured using PET images reconstructed with different algorithms and averaged over the three phantom inserts. The best results are obtained when applying the CNN on images reconstructed with ToF algorithm and with OSEM.

3.3 Summary

In this chapter we refined our proposed approach by training a 3D CNN on more realistic PET data generated with an analytical simulator. We first trained and tested the 3D CNN on PET data generated and reconstructed us-

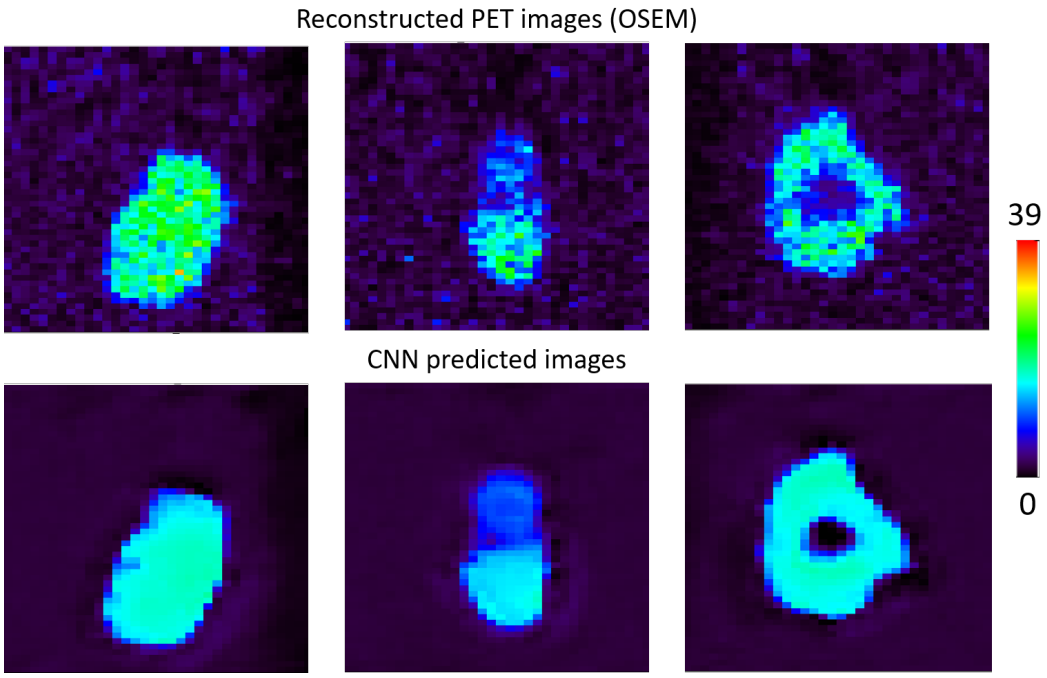


Figure 3.18: The top row presents coronal views extracted from the PET images of the three phantom inserts, reconstructed with OSEM algorithm. The bottom row shows the corresponding predictions yielded by the 3D CNN. The CNN yields denoised images and better defined tumours edges.

ing one spatially-invariant PSF. We tested CNNs with different depths and the 3D CNN composed of 6 convolutional blocks yielded the best results in terms of tumour shape definition without underestimating the background activity concentration. As the RCs are not accurately recovered when the tumours have smaller volumes, a more detailed analysis was performed. The tumour sphericity was measured for tumours smaller than 5 ml and our analysis suggested that the CNN recovers worse recovery coefficients for tumours characterised by a lower sphericity (i.e. a high aspect ratio). We provided average RC values for tumours with volumes comprised between 1 ml and 2 ml, and the RC estimates obtained using the images predicted by the CNN were closer to one than the RCs measured using the reconstructed PET images. The 3D CNN composed of 6 blocks was trained for increasingly high numbers of epochs and the re-

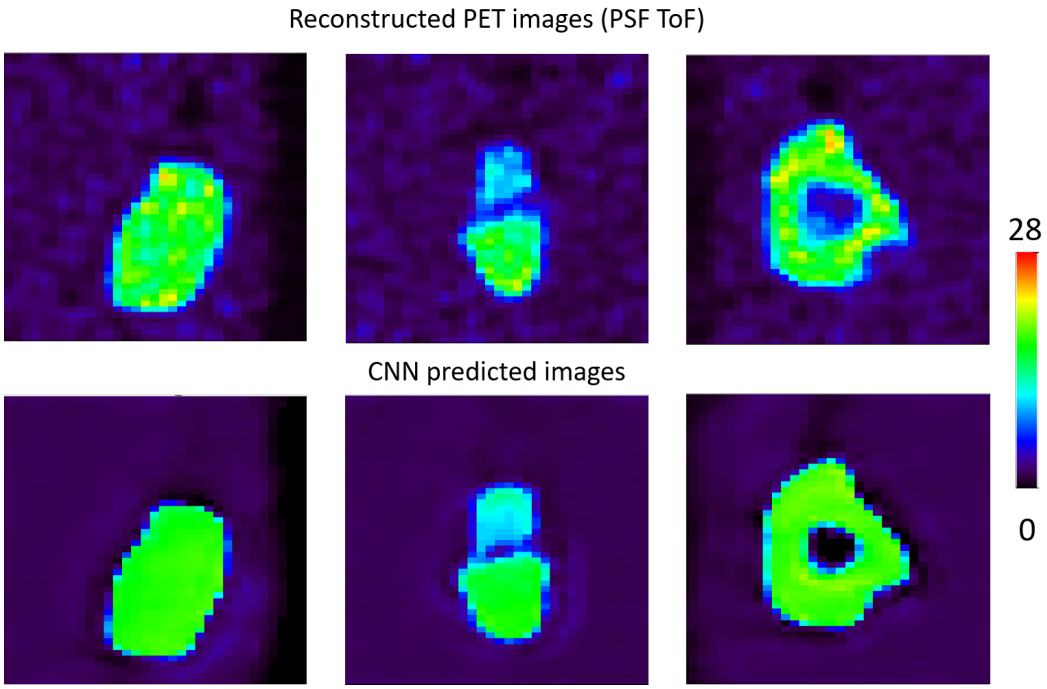


Figure 3.19: The top row presents coronal views extracted from the PET images of the three phantom inserts, reconstructed with ToF algorithm and PSF modelling. The bottom row shows the corresponding predictions yielded by the 3D CNN. The CNN yields denoised images and better defined tumours edges.

sults of these experiments were visually and quantitatively assessed. The CNN trained for 200 epochs predicted better defined tumour shapes and it yielded the best RC_{mean} and RC_{median} estimates. Further increasing the number of epochs led to a worse performance. Different normalisation methods were used to pre-process the simulated PET data before training, and the results were compared. Similar performances were achieved when training the CNN on the non-normalised dataset and on the dataset normalised by dividing each image voxel by the maximum overall activity concentration. The proposed DL approach was qualitatively and quantitatively compared to standard methods that are used to reduce the partial volume effect and image noise. A visual assessment showed that the CNN was able to outperform all the standard methods and to recover better defined tumour shapes and more uniform tu-

mour activity distributions. The 3D CNN also provided better RC_{mean} and RC_{median} estimates than all the other standard methods. The CNN trained on non-normalised data was applied to a dataset composed of reconstructed PET images generated with an analytical simulator using a range of PSFs. The images predicted by the CNN were denoised and the predicted tumours were characterised by sharper edges and by more uniform activity distributions. The CNN also improved the estimates of maximum, mean and median RCs. This suggests that the network can successfully recover the tumour shape and radioactivity distribution when applied to data generated with a range of PSFs that were not learnt during training. Finally, we applied the CNN on real phantom data. For this final test we used PET images reconstructed with four different algorithms. When applied to phantom PET images, the CNN yields denoised images and better defined tumours. The CNN also predicts a more uniform activity distribution within the tumour shapes. The best quantitative results were obtained when applying the CNN on PET images reconstructed with OSEM algorithm and ToF algorithm. Overall, the qualitative and quantitative analysis on phantom data suggests that the proposed method is robust to phantom PET images reconstructed with different algorithms. In the next chapter, we apply the 3D CNN trained on simulated data to real patient data.

Chapter 4
Clinical Applications

This chapter illustrates three applications of the proposed deep learning method to real patient data. In this chapter, few representative examples are presented and commented in detail, more visual examples are shown in the Appendix. In clinical datasets, PET data are characterised by a range of PSFs, which may be different that the one simulated in the experiments presented in previous chapters. Furthermore, clinical PET images are characterised by spatially variant PSFs, whereas we simulated spatially invariant PSFs. Although the results obtained in the previous chapter are encouraging, applying the network trained on simulated PET data to clinical PET images may be problematic. The network may not accurately recover the ground truth as during training it has not learnt how to map an image characterised by a spatially-variant PSF to its ground truth. In this chapter, some tests are performed on clinical datasets and the network's predictions are compared visually and quantitatively to the original PET images. This helps to better understand which particular scans may be best suited for the application of the proposed algorithm, and to identify particular image features that may cause the network to fail.

The CNN trained on PET images generated with an analytical simulator (see chapter 3) is applied to ^{18}F-FDG PET images of patients affected by non-small cell lung cancer. The same CNN is then applied to a Hodgkin lym-

phoma ^{18}F-FDG PET dataset. Finally, the deep learning approach is applied to a PET-MR prostate cancer dataset where the PET scans were acquired using ^{68}Ga-PSMA or ^{11}C-Choline as radiotracers. A new network performing prostate segmentation is introduced as a pre-processing step, to automatically segment the volume of interest from the MR images. The 3D CNN is subsequently applied to small regions cropped around the prostate from the PET images, to restore the original radioactivity distribution and tumour shape.

4.1 Non-Small Cell Lung Cancer PET Dataset

In this section, the deep learning method is applied to a dataset acquired from 47 non-small cell lung cancer (NSCLC) patients who received a ^{18}F-FDG PET/CT scan before chemo-radiotherapy. This dataset was acquired between 2007 and 2009 at King's College London, and it has been previously used by Bashir *et al.* to assess the effects of segmentation algorithms on the measurement of ^{18}F-FDG PET texture parameters in NSCLC [152]. All the patients involved in this study have single tumours, most patients have locally advanced NSCLC and are inoperable. The PET/CT scans were all acquired in the same institution and using the same protocol, at a median of 45 days before treatment. The two scanners used in this study (Discovery VCT or DST, GE Healthcare, Chicago, USA) were cross-calibrated to within 3% as reported in [152]. Patients are injected with a 350-400 MBq ^{18}F-FDG solution intravenously and the PET images are acquired approximately 90 mins after injection. The PET images are acquired from the base of the skull to the upper thighs.

Discovery ST Scanner Settings For the Discovery ST scanner, the average (radial and tangential) transverse spatial resolution (FWHM) at 1, 10 and 20 cm off axis was 6.28 mm, 7.09 mm and 7.45 mm in 2D mode [153]. The reconstruction algorithm used to generate the PET images is OSEM with 1 iteration and 30 subsets. The reconstructed images have voxel size $5.47 \times 5.47 \times 3.27$ mm^3, and they are smoothed post-reconstruction using a Gaussian filter with FWHM = 5 mm^3.

Discovery VCT Scanner Settings The Discovery VCT scanner is characterised by an average radial and tangential spatial resolution of 4.78 mm (FWHM, 2D mode) at 1 cm offset from the centre. Radial and tangential resolutions are characterised by FWHM = 5.70 mm and FWHM = 5.84 mm at 10 cm offset from the centre [154]. The reconstruction algorithm used to generate the PET images is OSEM with 1 iteration and 28 subsets. The reconstructed images have voxel size $4.69 \times 4.69 \times 3.27$ mm^3, and they are smoothed post-reconstruction using a Gaussian filter with FWHM = 6 mm^3.

4.1.1 Neural network

In this experiment, the CNN trained on simulated data is applied to the NSCLC patient dataset. The 3D CNN is composed of 6 blocks and trained for 200 epochs using Adam optimiser as explained in chapter 3. Mean squared error is used as the loss function and the simulated PET images used for training are not normalised. The trained network is applied to VOIs of size $50 \times 50 \times 50$ voxels, cropped around the bulkiest lesion visible in the patient PET scans. The voxel size may be $5.47 \times 5.47 \times 3.27$ mm^3 or $4.69 \times 4.69 \times 3.27$ mm^3 depending on the scanner used to acquire the PET data.

4.1.2 Results

The results obtained applying the CNN trained on simulated data to the NSCLC patient dataset are presented here. The analysis is focused on the metabolically active primary lung tumours. Fig. 4.1 shows transverse views of two representative tumours. The tumour shown in the top row is heterogeneous and two areas characterised by different activities can be distinguished. The corresponding image provided by the CNN shows a heterogeneous tumour characterised by sharper edges. The predicted image is denoised and the high activity region in the original PET image corresponds to a region of high uptake in the predicted image. Another example is presented in the bottom row of Fig. 4.1. This tumour has a higher activity concentration in its inner core, and a lower activity concentration towards the edges. The CNN predicts a denoised image, where the tumour has a better defined shape and a more uniform activity distribution. In this case some background structures are also improved by the CNN. Two further examples of hollow tumours with different shapes, sizes and activities are shown in Fig. 4.2. The CNN recovers denoised images where the hollow tumours have sharper edges. Again, the predicted tumours are characterised by a more uniform activity distribution.

To quantitatively assess our results the maximum, mean and peak SUVs are measured and subsequently compared to the same measurements performed directly on the reconstructed PET images. The automated estimation of SUV_{peak} is computationally intensive, so it could not be performed previously when the test dataset was composed of hundreds of images. The peak value is presented here instead of the median value, as this metric is more commonly used in clinical routine. Fig. 4.3 shows the maximum tumour activity plotted in order of increasing volume. The tumour volume was estimated based on the number of tumour voxels exceeding SUV = 2.5 in a 3D box drawn around the tumour. As expected from the visual assessment, the CNN predicts lower SUV_{max} than

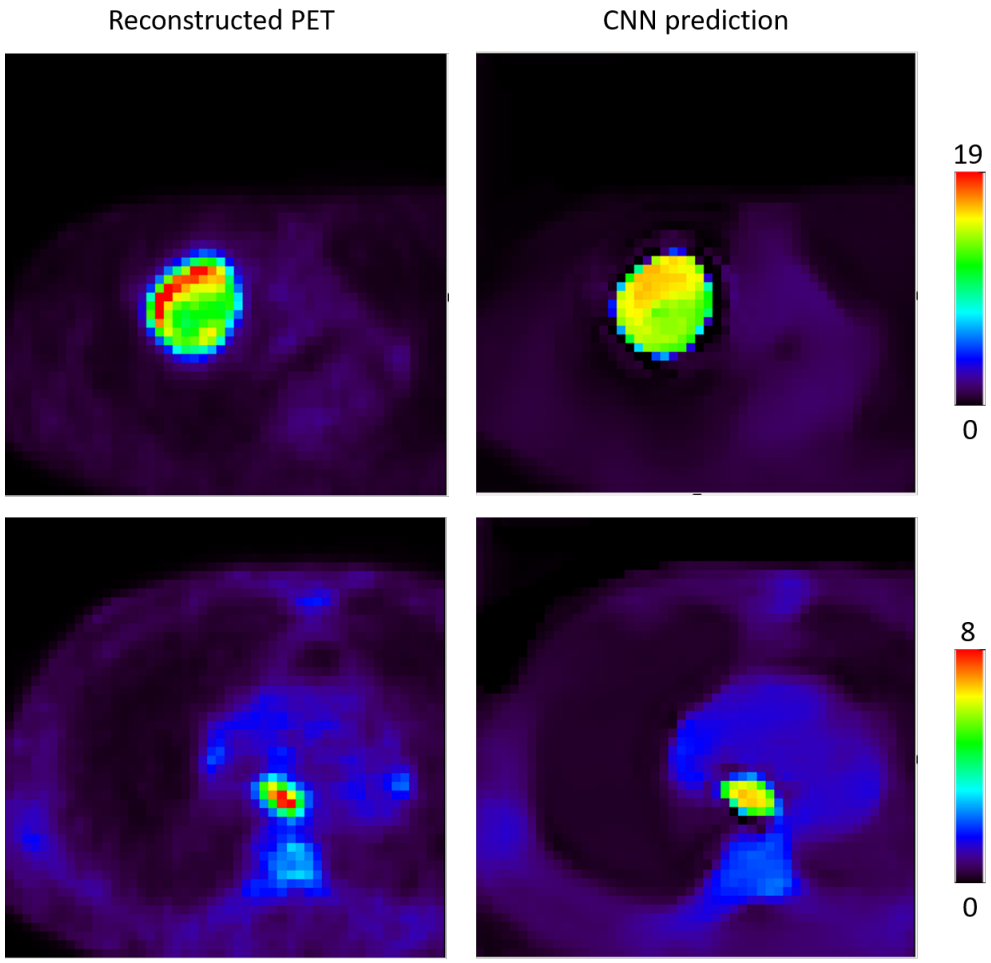


Figure 4.1: Transverse views of two NSCLC lesions. In the top row a heterogeneous tumour is presented, the original PET image is shown in the first column and the corresponding CNN prediction is shown in the second column. In the bottom row, the PET image of a small tumour is presented, together with the CNN prediction. The CNN predicts denoised images, where the tumours have better defined shapes and more uniform activity distributions. In each row the images are shown with the same SUV colour scale.

the ones measured in the original reconstructed PET images. The absolute percentage differences between the SUV_{max} measurements obtained using the original reconstructed PET images and the ones predicted by the CNN range from ± 12 % to ± 37 %. In order to measure the SUV_{mean}, a 3D box is drawn around the tumour and only the voxels exceeding SUV = 2.5 are included in

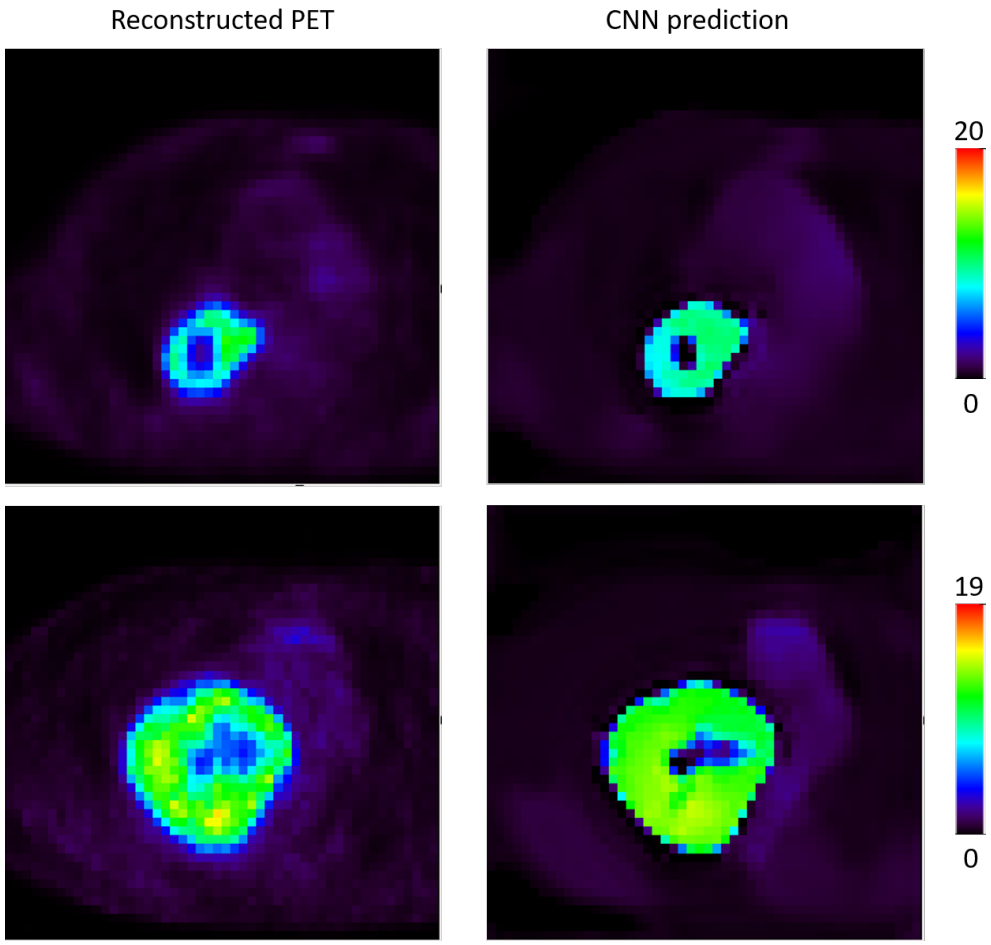


Figure 4.2: Transverse views of two hollow tumours. In the top row a smaller tumour is presented, the original PET image is shown in the first column and the corresponding CNN prediction is shown in the second column. In the bottom row, the PET image of a larger hollow tumour is presented, together with the CNN prediction. The CNN predicts denoised images, where the tumours have better defined shapes and more uniform activity distributions. In each row the images are shown with the same SUV colour scale.

the measurement of the SUV_{mean}. This threshold is often used as cut-off to perform SUV measurements in clinical settings [155]. The SUV_{mean} values measured using this method are plotted in Fig. 4.4. The CNN predicts overall higher mean SUVs than the ones measured using the original reconstructed PET images. The absolute percentage differences between the SUV_{mean} extracted from original reconstructed PET images and the ones predicted by the

CNN range from ±8 % to ±43 %. Finally, the SUV_{peak} measurements are presented in Fig. 4.5. The CNN yields lower estimates for the peak SUVs. This is different than what was observed when testing the CNN on simulated data (see Fig. 2.12). This is due to the fact that the PET images belonging to the NSCLC dataset are characterised by larger voxels than the ones relative to the simulated datasets used in Chapter 2 and Chapter 3. As a result, the 1 cm^3 sphere used to measure the peak values contains over 100 voxels when applied to the simulated PET images, but it only includes around 10 voxels when applied to the NSCLC PET images. The absolute percentage differences between the peak SUVs measured in the original reconstructed PET images and the ones predicted by the CNN range from ±4 % to ±40 %.

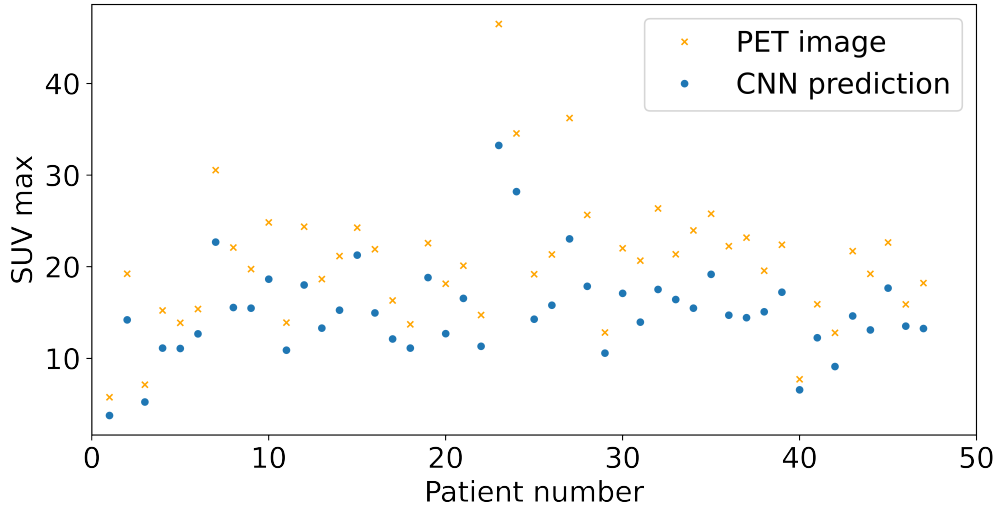


Figure 4.3: Plot of SUV_{max} measurements relative to the NSCLC dataset. The measurements performed using the PET scans are presented in orange, the ones obtained using the CNN predictions are shown in blue. As expected, the estimates of maximum activity yielded by the CNN are lower than the SUV_{max} measured in the original reconstructed PET images.

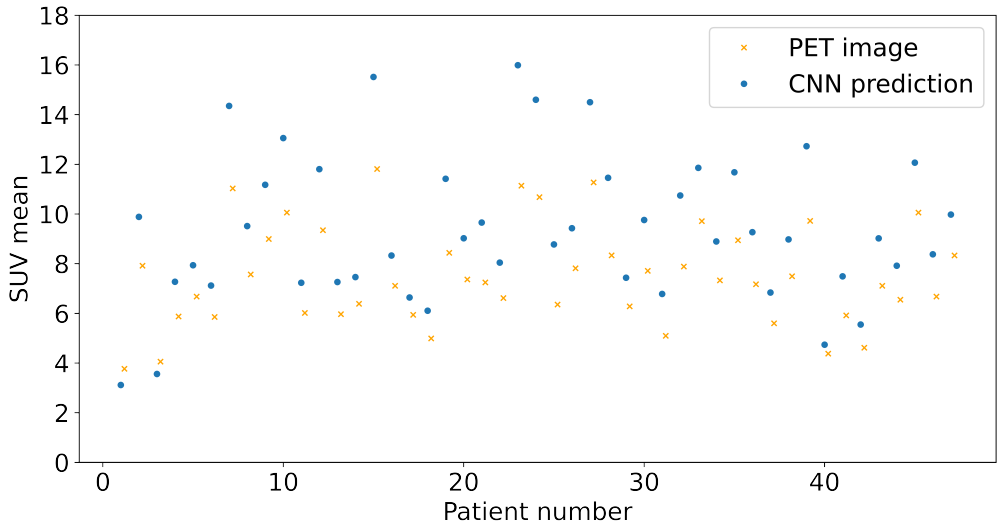


Figure 4.4: Plot of SUV_{mean} measurements relative to the NSCLC dataset. The measurements performed using the PET scans are presented in orange, the ones obtained using the CNN predictions are shown in blue. The SUV_{mean} estimates yielded by the CNN are in most cases higher than the SUV_{mean} measured in the original reconstructed PET images.

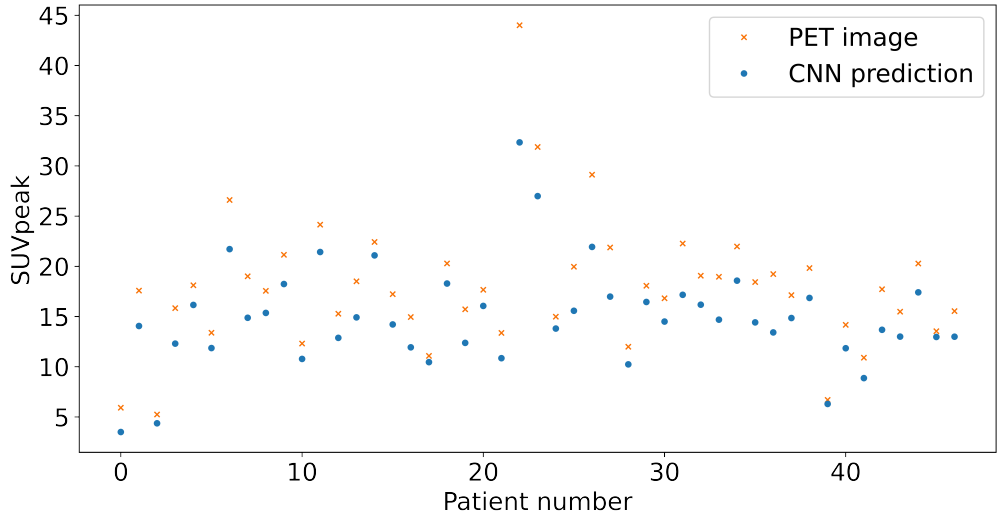


Figure 4.5: Plot of SUV_{peak} measurements relative to the NSCLC dataset. The measurements performed using the PET scans are presented in orange, the ones obtained using the CNN predictions are shown in blue. The CNN yields lower SUV_{peak} estimates.

4.2 Hodgkin Lymphoma PET dataset

The second application of our deep learning approach is performed on a Hodgkin lymphoma dataset. Hodgkin lymphoma (HL) is a type of cancer that predominantly affects young adults, occurring at a median age of 35 years. In HL cancer originates from lymphocytes, which are a type of white blood cell. HL lesions can have different morphology and composition. HL typically presents as painless lymph nodes that are abnormal in size, frequently located in the cervical or supra-clavicular area. In more advanced stages, additional lesions can be located in more distant lymph nodes i.e. below the diaphragm. In the last stages lesions are disseminated and may involve extra-lymphatic organs such as lungs, spleen or liver. Even in advanced stage disease, HL is highly curable with combination chemotherapy, radiation or combined modality treatment.

RATHL dataset The RATHL (response-adapted therapy for advanced Hodgkin lymphoma) dataset is a multicentre randomised trial that involves patients affected by advanced Hodgkin's lymphoma. Recruitment for RATHL was open in hospitals across the UK, Italy, Norway, Sweden, Denmark, Australia, and New Zealand from 2008 to 2012. In this study, after initial staging each patient received a baseline ^{18}F-FDG PET/CT scan before treatment. After two cycles of chemotherapy a new PET/CT scan was acquired to monitor response to treatment. Participants who had negative scans were randomly assigned to receive a further four cycles of chemotherapy. Participants with positive scans were escalated to different chemotherapy regimens [156]. A subset of 13 cases was used in this thesis. For these patients, the original reconstructed PET images are composed of $168 \times 168 \times N$ voxels, where N ranges from 248 to 364 depending on the patient. The original reconstructed PET images were reconstructed using OSEM algorithm with 4 iterations and 8 subsets, with voxel

size $4.06 \times 4.06 \times 5.00$ mm^3, and subsequently smoothed post-reconstruction using a Gaussian filter with FWHM = 5 mm^3.

4.2.1 Neural network

In this experiment, the CNN trained on PET images simulated using the analytical simulator is applied to the RATHL dataset. The 3D CNN is composed of 6 blocks and trained for 200 epochs using Adam optimiser as described in chapter 3. Mean squared error is used as loss function to train the network, and the simulated PET images used for training are not normalised. The trained network is applied to VOIs of size $50 \times 50 \times 50$ voxels cropped around the bulkiest lesion visible in each patient PET scans.

4.2.2 Results

The qualitative and quantitative results obtained applying the 3D CNN trained on simulated data on the RATHL dataset are presented here. The analysis is focused on the bulkiest lesion present in each scan, as clinical measurements were performed only for these lesions. The bulkiest lesion could be identified unambiguously from the image coordinates provided by clinicians. The volumes of the bulkiest lesions ranged from 30 ml to 300 ml. Fig. 4.6 shows a comparison between original reconstructed PET images and CNN predictions for a representative patient. The baseline and follow-up scans are presented in separate columns. The follow-up scans are presented with two different colour scales, to facilitate a direct comparison with the baseline scans. When the CNN is applied to the baseline scan it yields a denoised image where the tumour is characterised by sharper edges. The inner core of the tumour has a cluster of high activity voxels in the original PET image and a region with higher activity can also be distinguished in the image predicted by the CNN.

The second column shows the prediction obtained applying the CNN on the follow-up scan. Again, the CNN yields a denoised image with sharper tumour edges. The recovered tumour is characterised by inhomogeneous activity, and its activity distribution differs from the one in the original PET image. Another example is shown in Fig. 4.7, where analogous images are shown for a different patient. When applied to the baseline scan, the CNN predicts a denoised image with sharper tumour edges. When applied to the follow-up scan, the CNN predicts a large area with homogenous activity. One possible reason why this happens is that the tumour activity in the follow-up scan is low and the lesion is not localised in a well defined area as was the case in the baseline scan. This scattered activity distribution is different from the activity distributions that the CNN learnt during training, so it may be more difficult for the CNN to distinguish tumour voxels from background voxels.

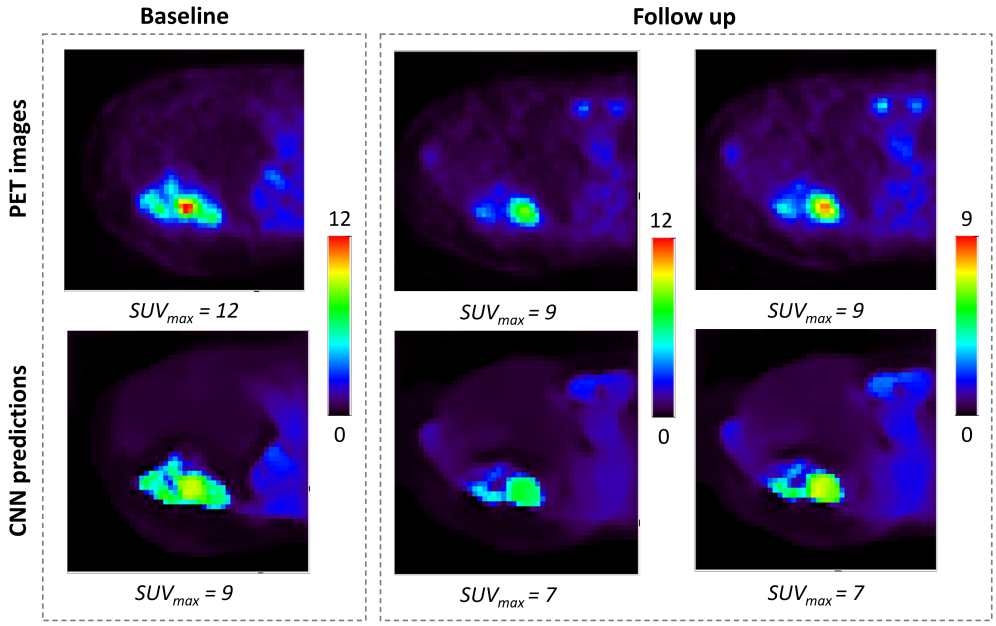


Figure 4.6: Transverse views of PET images of a lymphoma patient. The baseline scan and the follow-up scan are shown in the top row. The corresponding images predicted by the CNN are presented in the bottom row. The CNN predicts denoised images. The images in the same column are shown using the same SUV colour scale.

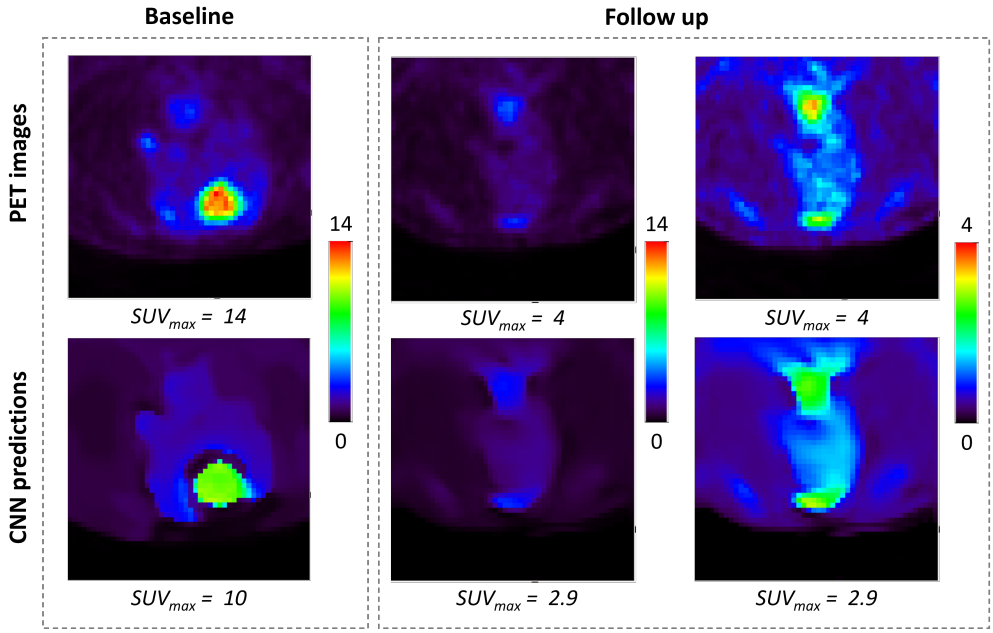
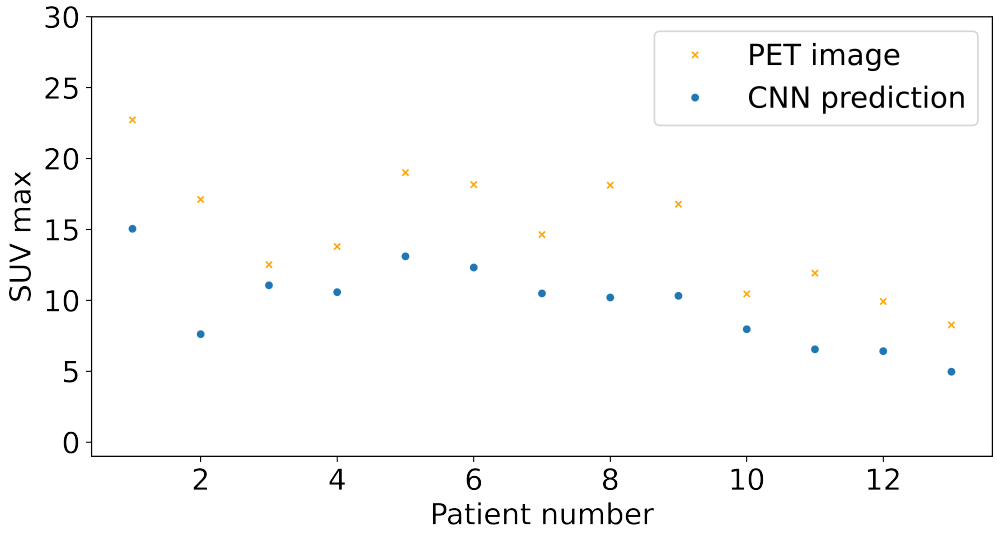


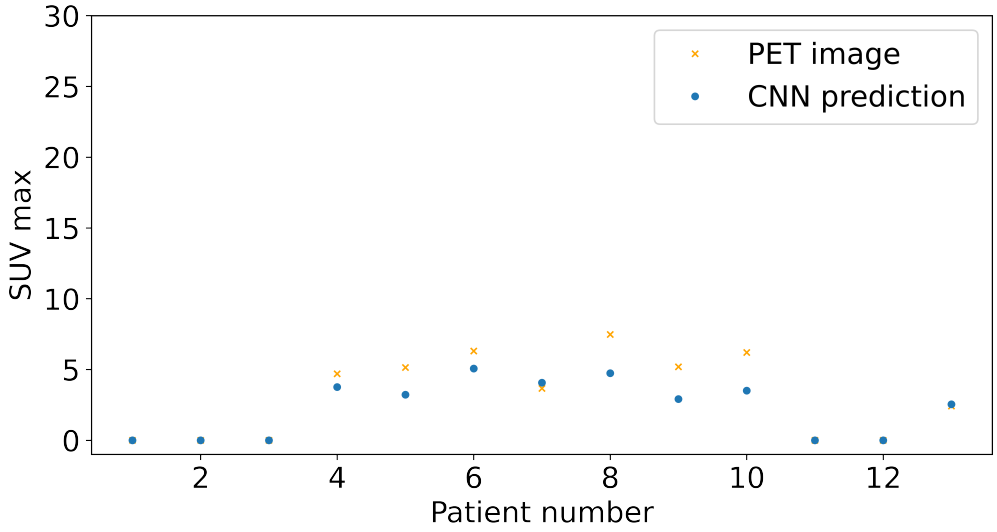
Figure 4.7: Transverse views of PET images of a lymphoma patient. The baseline scan and the follow-up scan are shown in the top row. The corresponding images predicted by the CNN are presented in the bottom row. The CNN predicts denoised images where tumours have sharper edges. The images in the same column are shown using the same SUV colour scale.

A quantitative analysis is performed for the bulkiest lesion in each scan. The maximum, mean and peak SUVs are measured. In order to perform the quantitative analysis, a 3D box is drawn around the bulkiest tumour and only the voxels within the box exceeding SUV = 2.5 are included in the calculation of the SUV_{mean}. In Fig. 4.8 the SUV_{max} measurements are plotted both for the baseline scans and for the follow-up scans. The CNN yields lower SUV_{max} estimates when applied both to the baseline scan and to the follow-up scan. Compared to the SUV_{max} clinical measurements performed on the baseline scans, the SUV_{max} extracted from the images predicted by the CNN are 12% to 55% lower. Similarly, in the follow up scans there is a decrease in SUV_{max}, ranging from 19% to 43%. In one case the CNN yielded a slightly higher maximum value for the follow-up image. In five cases the SUV_{max} is 0, meaning that in those images there are no voxels exceeding the SUV threshold of 2.5

within the given VOI. This only happens in the follow-up scans, when the patients have received therapy and thus some tumours might have disappeared. The SUV_{mean} measurements are plotted in Fig. 4.9. When considering the baseline scans, the CNN predicts slightly higher SUV_{mean} estimates than the ones measured in the original reconstructed PET images. For the four biggest tumours, plotted on the right side, the SUV_{mean} predicted by the network is comparable to the one measured in the original reconstructed PET images. The SUV_{mean} extracted from the images predicted by the CNN are 1% to 53% higher than the corresponding measure extracted from the original PET images. The mean activity concentrations relative to the follow-up scans have similar values in the original PET scans and in the CNN predicted images. This is an encouraging result as it means that the CNN does not generate regions with high activity concentration when the PET scans only contains lesions with low activity concentration or just background. Finally, the SUV_{peak} measurements are plotted in Fig. 4.10. In contrast to what was observed in Chapter 2, when applied to the baseline scans the CNN tends to predict lower SUV_{peak} than the ones measured in the original reconstructed PET images. The values predicted by the CNN are 3% to 11% lower than the ones obtained from the original PET images. This is due to the fact that the PET images belonging to the RATHL dataset are characterised by a large voxel size, so less than 10 voxels are included in the 1 cm^3 sphere used to measure the peak value. In the simulated PET images, which are composed of smaller voxels, more than 100 voxels contribute to the measurement of the peak value. When applied to the follow-up scans the CNN predicts SUV_{peak} that are similar to the ones measured in the original PET images. In this case, the SUV_{peak} obtained using the CNN predictions are 5% to 43% lower than the corresponding measurements extracted from the original PET images.

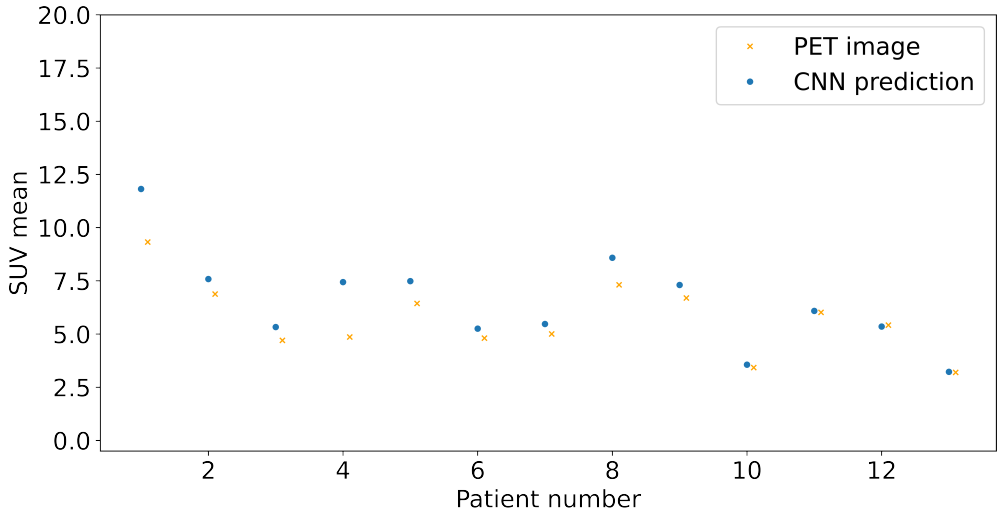


(a) Baseline scan

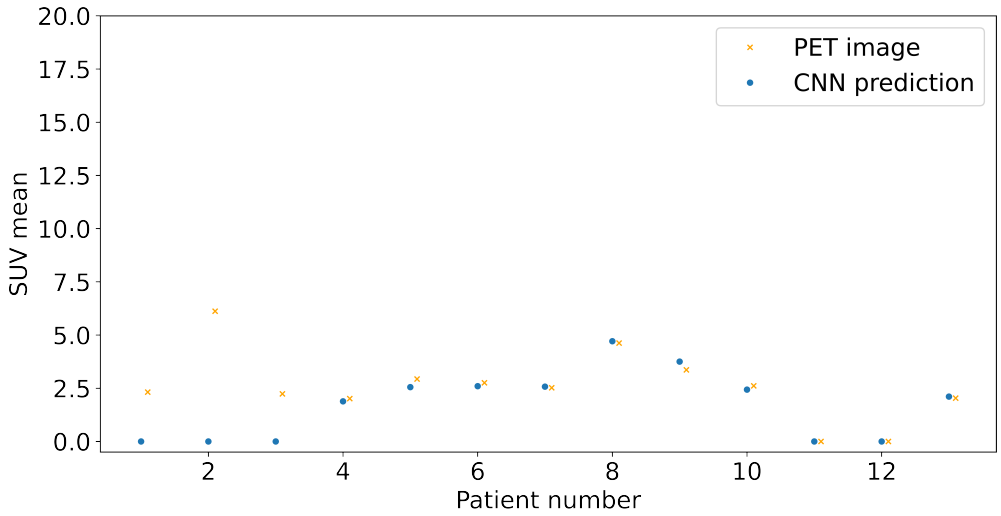


(b) Follow-up scan

Figure 4.8: Plot of the SUV_{max} relative to the RATHL dataset. The values in (a) are measured using the baseline PET scans, whereas the values in (b) are extracted from the follow-up scans. The measurements relative to the original reconstructed PET images are shown in orange, the ones obtained using the images predicted by the CNN are shown in blue. The CNN yields lower SUV_{max}.

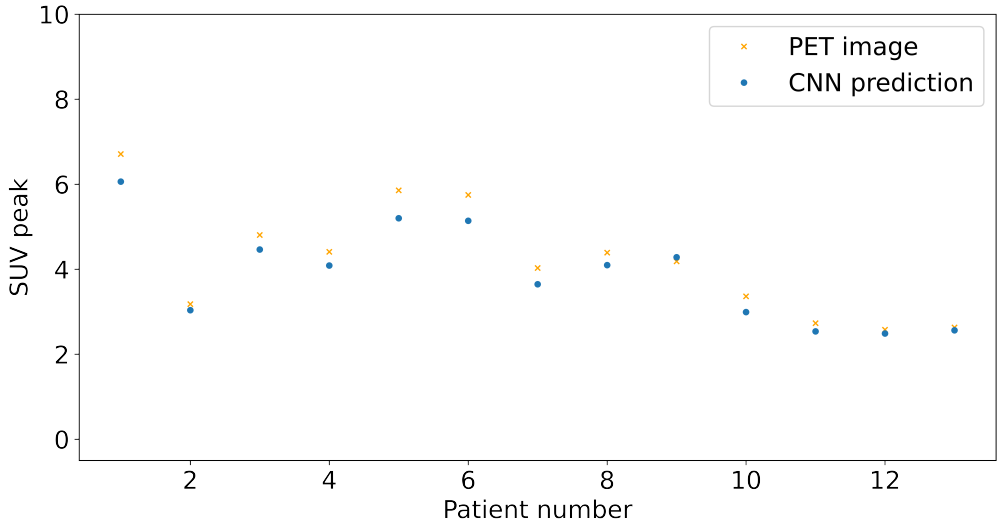


(a) Baseline scan

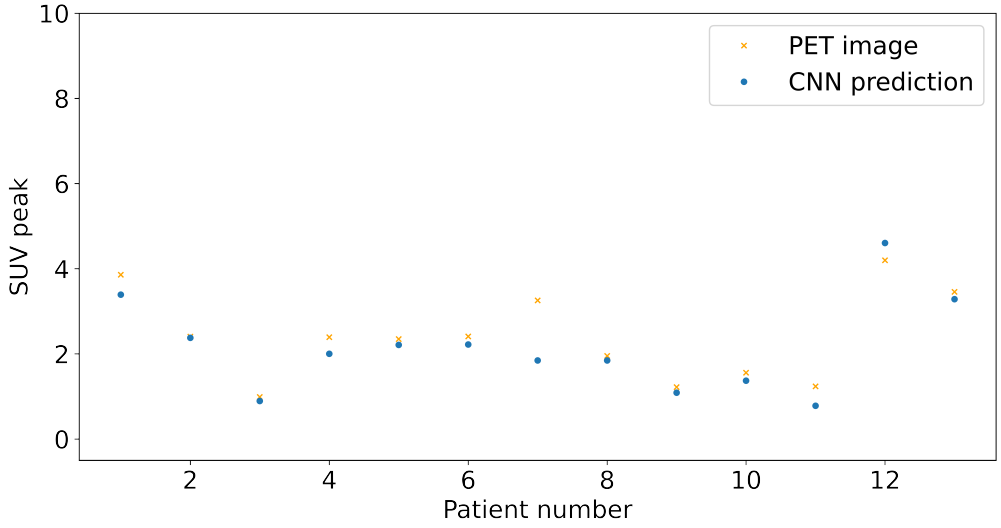


(b) Follow-up scan

Figure 4.9: Plot of the SUV_{mean} relative to the RATHL dataset. The values in (a) are measured using the baseline PET scans, whereas the values in (b) are extracted from the follow-up scans. The measurements relative to the original reconstructed PET images are shown in orange, the ones obtained using the images predicted by the CNN are shown in blue. A small offset was added along the x axis to the PET image values to allow for a better visualisation. The CNN yields overall higher SUV_{mean} estimates when applied to the baseline scans. When applied to the follow-up scans, the CNN predicts SUV_{mean} that are comparable to ones measured in the original reconstructed PET images.



(a) Baseline scan



(b) Follow-up scan

Figure 4.10: Plot of the SUV_{peak} relative to the RATHL dataset. The values in (a) are measured using the baseline PET scans, whereas the values in (b) are extracted from the follow-up scans. The measurements relative to the original reconstructed PET images are shown in orange, the ones obtained using the images predicted by the CNN are shown in blue. The CNN yields overall lower SUV_{peak} estimates when applied to the baseline scans. When applied to the follow-up scans, the CNN predicts SUV_{peak} that are comparable to ones measured in the original reconstructed PET images.

4.3 Prostate Cancer PET-MR Dataset

In this section the deep learning method is applied to a PET-MR prostate cancer dataset. Prostate cancer is one of the most common malignancies affecting men worldwide. It forms in the tissues of the prostate, which is a gland located below the bladder and in front of the rectum. The initial assessment of prostate cancer is usually performed through physical exams and blood tests, which can detect increased levels prostate-specific antigen (PSA) that are usually associated with prostate cancer. Prostate cancer can only be diagnosed with a biopsy, which entails taking samples of cells from different regions of the prostate. Using the biopsy results, the cancer aggressiveness is measured through the Gleason score. This method assigns a score to each tumour area probed during the biopsy based on its cell pattern. Higher values correspond to more aggressive tumours. The two regions with highest score contribute to the final Gleason score, which can range from 6 to 10. Based on the cancer stage different treatments including surgery and/or chemo-radiotheraphy are recommended. Accurate staging of lesions is critical for treatment planning and prognosis in patients with prostate cancer. Biopsy is an invasive procedure which only provides limited information about the tumour composition. Some aggressive parts of the tumour may be missed during the collection of samples for the biopsy, thus resulting in an underestimation of the tumour grade. Imaging techniques can be used to complement the information obtained form the biopsy, to visualise prostate cancer lesions and to assess the extent of the disease. PET imaging enables the detection of small metastatic prostate cancer lesions, giving more information about tumour radiotracer uptake and tumour burden. Prostate-specific radiotracers such as ^{68}Ga-PSMA and ^{11}C-Choline have been developed to target more effectively prostate cancer cells. Anatomic imaging techniques such as MR or CT help to identify and

further characterise lesions. In the context of MRI, T1 and T2 sequences are commonly used to evaluate tumour infiltrations in structures located near the prostate, such as the seminal vesicles and lymph nodes [157, 158]. MRI offers a high spatial resolution and high soft-tissue contrast, thus it is the imaging modality of choice for tumour staging. Measurements such as prostate volume, shape and location can be extracted from MRI and subsequently used in the surgical planning for prostatectomy or radiotherapy. The measurement of prostate volume can help, together with other relevant parameters, to assess the stage of prostate cancer and to monitor response to treatment, however, the limited accuracy and the variability of prostate volume measurements hamper the usefulness of this metric in clinical practice. In recent years, a lot of work has been done to develop algorithms that automatically perform prostate MR segmentation and thus measure the tumour volume [159–161]. In this chapter, prostate segmentation performed using MR images is used to automatically define a region of interest that can subsequently be used for PET image analysis.

PROMISE dataset The Prostate MR Image Segmentation (PROMISE12) challenge was setup to enable standardised evaluation and objective comparison between different segmentation methods. 100 prostate MR scans from four centres were included in the challenge, with differences in scanner manufacturer, field strength and acquisition protocol [162]. A subset of the PROMISE dataset including 50 transverse T2-weighted MR images, each paired with a prostate segmentation performed by an experienced reader, is used in this thesis. The prostate segmentations are performed on a slice-by-slice basis using a contouring tool. The dataset includes both benign disease and prostate cancer and the scans present a range of prostate sizes and appearance.

PROSPER dataset PROSPER is a randomised trial, performed at 254 sites worldwide with the aim of assessing the efficacy and safety of hormone therapy using enzalutamide in adult patients. The trial involves around 1400 patients with non-metastatic castration-resistant prostate cancer. A subset of 28 patients was made available for the experiments in this thesis. For each patient, one or two PET-MR scans are acquired before radiotherapy or chemotherapy, using ^{11}C-Choline or ^{68}Ga-PSMA as radiotracers. The original reconstructed PET images used in this thesis are acquired using the Siemens Biograph mMR (Siemens, Erlangen, Germany). The axial ans transverse spatial resolution (FWHM) of this scanner are respectively 4.1 mm and 4.0 mm at 1 cm offset from the centre. At 10 cm offset from the centre the axial and transverse spatial resolutions (FWHM) are 6.4 mm and 5.0 mm. The images used in this section are reconstructed using OSEM algorithm with 3 iterations and 21 subsets. The original reconstructed PET images are smoothed post-reconstruction using an isotropic Gaussian filter with FWHM = 4 mm^3. The size of these PET images is $172 \times 172 \times 127$ voxels, with voxel size $4.17 \times 4.17 \times 2.03$ mm^3. Additionally, in this section we use MR T2-weighted images acquired with a magnetic field of strength 3 T. These images are made of $448 \times 336 \times 88$ voxels, with voxel size $0.85 \times 0.85 \times 2.0$ mm^3.

4.3.1 Method

The deep learning method proposed in previous experiments is here expanded by incorporating MR images, thus enabling automated extraction of the prostate and surrounding regions. An overview of the method is illustrated in Fig. 4.11. A UNet is initially trained using a set of MR T2-weighted images taken from the PROMISE dataset, together with the corresponding prostate segmentations. The UNet is then tested on 5 unseen sets of 2D MR slices taken from the PROMISE dataset to obtain a prediction of the prostate segmentation. The re-

sulting segmentations are qualitatively and quantitatively assessed against the gold standard segmentations performed by experienced readers. As part of the data processing pipeline (Fig 4.11), the UNet is applied to the PROSPER T2-weighted MR images to obtain a prediction of the prostate segmentation. Each MR scan in the PROSPER dataset is paired with a simultaneously acquired PET scan. The prostate segmentations provided by the UNet are resized to the dimensions of the corresponding original reconstructed PET images. A $50 \times 50 \times 50$ voxels VOI is automatically cropped around the prostate for each PET scan, using the coordinates or the predicted MR segmentations as a reference. A 3D CNN trained on simulated data is finally applied to the VOIs cropped from the original reconstructed PET images, with the aim to obtain a prediction of the ground truth tumour shape and activity distribution.

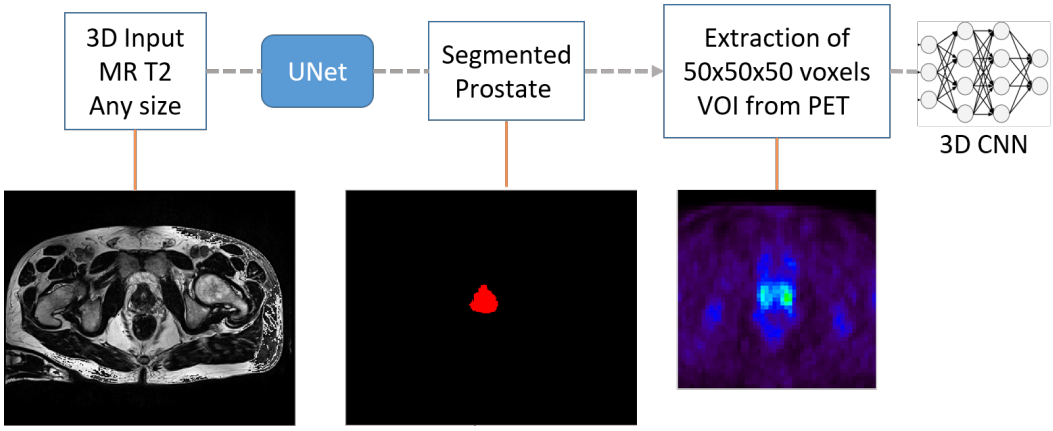


Figure 4.11: Illustration of the proposed method, applied to a PET-MR dataset. T2-weighted MR images and their corresponding prostate segmentations are given as an input to a UNet. When tested on an unseen set of MR T2-weighted images, the UNet automatically segments the prostate. The segmentations predicted by the UNet are resized to the dimensions of the corresponding PET images, and the coordinates of the segmentation masks are used as a reference to automatically crop a $50 \times 50 \times 50$ voxels VOI containing the prostate from the PET scans. Finally, the 3D CNN is applied to the cropped VOIs with the aim to restore the 'ground truth' activity distributions and tumour shapes.

Segmentation network The first part of the deep learning method involves the use of a UNet to automatically perform prostate segmentation from MR T2-weighted images. The UNet architecture used to segment the prostate is adapted from the work of Mirzaev [163], who participated to the PROMISE segmentation challenge and subsequently published his results. The UNet architecture is shown in Fig. 4.12. It accepts 2D images as input and it is composed of two symmetric paths which respectively perform downsampling and upsampling operations. In the downsampling path, convolutional blocks are followed by max pooling layers, which have the effect of reducing the size of the input image. As shown in Fig. 4.13 a convolutional block is composed of two convolutional layers with filter size 3×3 and with ReLU activation. Each convolutional layer is followed by a batch normalisation (BN) layer. A dropout layer with dropout rate 50% is introduced after the first BN to reduce overfitting. The input image is concatenated to the output of the second BN layer to obtain the output image. This is done because short skip connections within the convolutional block should lead to faster convergence when training [164]. The downsampling path is composed of seven convolutional blocks in total, and the number of kernels employed in each block is half the number of kernels used in the previous block. The image dimensions at the bottom of the UNet are 2×2 pixels. The upsampling path is composed of convolutional blocks followed by deconvolution layers that upsample the input images by projecting each input voxel to a bigger region through the kernel. The output of each upsampling layer is concatenated to the output of the corresponding convolutional block located in the downsampling path. These connections provide the details that are needed to reconstruct accurate segmentation boundaries. Finally, a convolutional layer with sigmoid activation yields the probability of each voxel belonging to the prostate or to the background.

This network is trained on 2D transverse slices composing the MR T2-

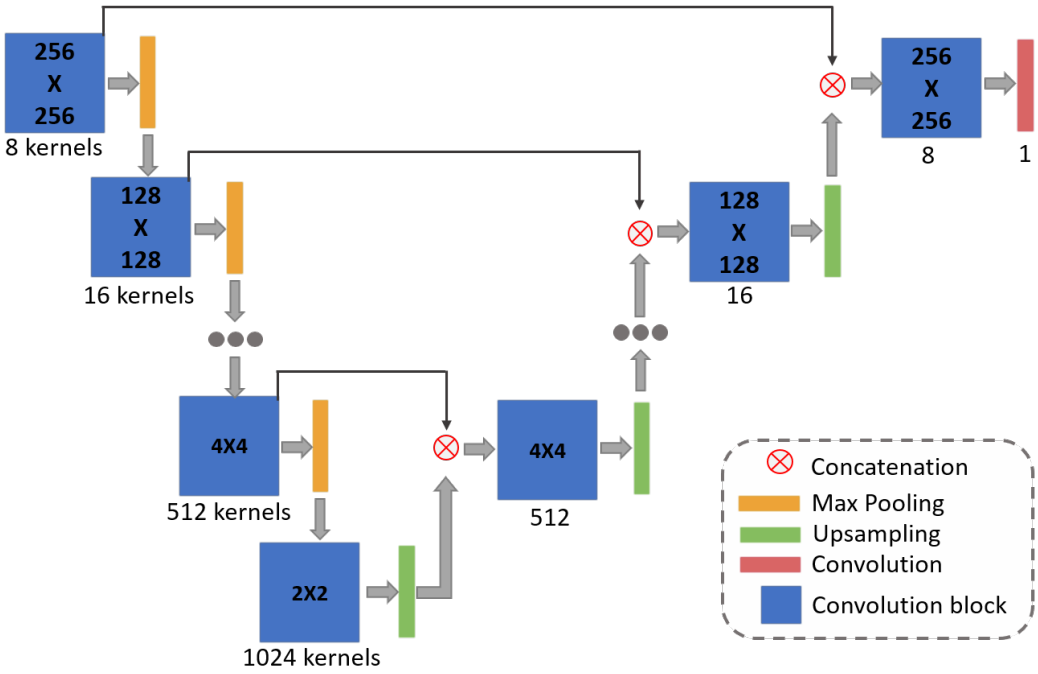


Figure 4.12: Schematic illustration of the UNet architecture. The UNet is composed of two symmetric paths which perform downsampling and upsampling operations. In the downsampling path, convolutional blocks are followed by max pooling layers which have the effect of reducing the size of the input image. The downsampling path is composed of seven convolutional blocks in total, and the number of kernels employed in each block is half the number of kernels used previous layer. The upsampling path is composed of convolution blocks followed by deconvolution layers that upsample the input images. The output of each upsampling layer is concatenated to the output of the corresponding convolutional block located in the downsampling path. Finally a convolutional layer outputs the segmented images. The input image is concatenated to the output of the second batch normalisation layer to obtain the output image.

weighted images. As the MR images in the training dataset are acquired with different scanners and protocols, they may have different sizes. To account for this, all MR images are resized to 256×256 pixels using a nearest-neighbour interpolation before being processed by the UNet. This technique replaces every pixel in the input image with the nearest pixel in the output. In this experiment 50 MR images are used for training, 5 for validation and 5 for

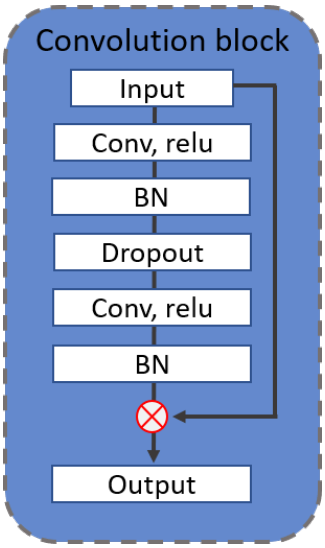


Figure 4.13: Illustration of a convolution block. Each convolution block is composed of two convolutional layers with filter size 3x3, assigned with ReLU activation. Each convolutional layer is followed by a batch normalisation layer. A dropout layer with dropout rate 50% is introduced after the first batch normalisation.

testing. The images are augmented with random rotations, shifts and flips to increase the size of the training dataset. Training is performed for 20 epochs using Adam optimiser and Dice loss function. This function is based on the maximisation of the Dice coefficient, which measures the overlap of two sets of binary data A and B as:

$$Dice = \frac{2\,|A \cap B|}{|A| + |B|} \tag{4.1}$$

where $|A|$ and $|B|$ are the cardinalities of the two sets (i.e. the number of elements in each set). When A and B are segmentation masks, $|A \cap B|$ is measured as the element-wise multiplication between the two masks followed by the sum of the elements in the resulting matrix. So the dice loss function between a predicted value a and the corresponding estimate b predicted by the network is defined as:

$$Dice\,(a,b) = \frac{2\,\sum_i^N a_i\,b_i}{\sum_i^N a_i + \sum_i^N b_i} \tag{4.2}$$

where N is the number of pixels composing the images [165].

CNN The optimised CNN trained on PET images simulated using the analytical simulator is applied to the original reconstructed PET images belonging to the PROSPER dataset. The 3D CNN is composed of 6 blocks and trained for 200 epochs using Adam optimiser as explained in chapter 3. Mean squared error loss function is used to train the network, and the simulated PET images used for training are not normalised. The trained network is applied to VOIs of size $50 \times 50 \times 50$ voxels, cropped around the prostate.

4.3.2 Results

The results obtained performing prostate segmentation on the PROMISE dataset are presented in Fig. 4.14. This figure shows two transverse views of T2-weighted MR images superimposed with the prostate segmentation generated by the UNet. It was possible to quantitatively assess the results as a ground truth segmentation was provided, as described in section 4.3. The mean Dice coefficient measured using the segmentations predicted by the UNet is 0.88 ± 0.04, similar to what is reported by Mirzaev in [163]. After checking that the UNet could accurately perform prostate segmentation on the PROMISE dataset, the network was applied to the PROSPER T2-weighted MR images. Two representative images are shown in Fig. 4.15. The UNet can correctly identify and segment the prostate in both cases. It was not possible to measure the Dice coefficient for the PROSPER data as the ground truth segmentation was not available. For the purpose of this thesis, a visual assessment performed on all patients was sufficient to confirm that the UNet could correctly locate and segment the prostate. As the final goal was to apply the 3D CNN on a

larger VOI around the prostate, an accurate delineation of the prostate was not essential.

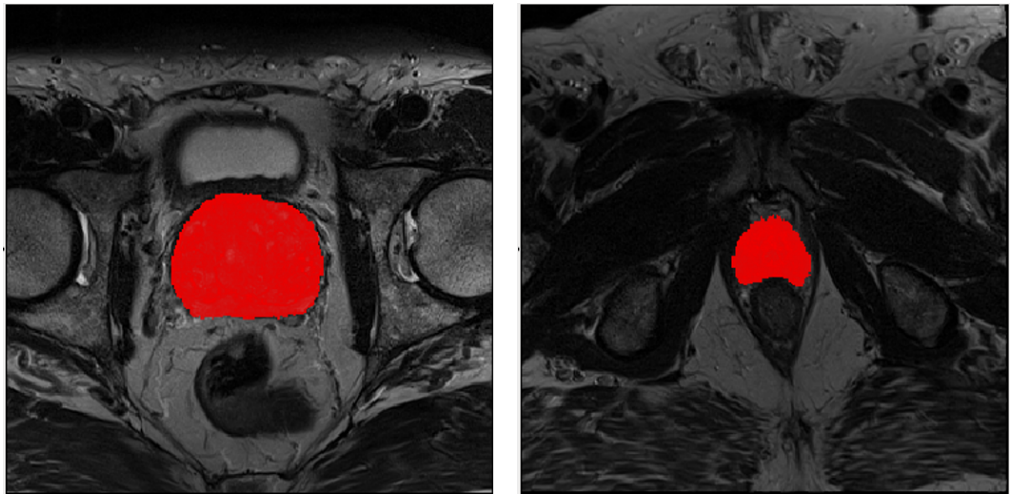


Figure 4.14: Transverse views of T2-weighted MR images belonging to the PROMISE dataset. The prostate segmentations generated by the UNet are shown in red, superimposed to the MR images.

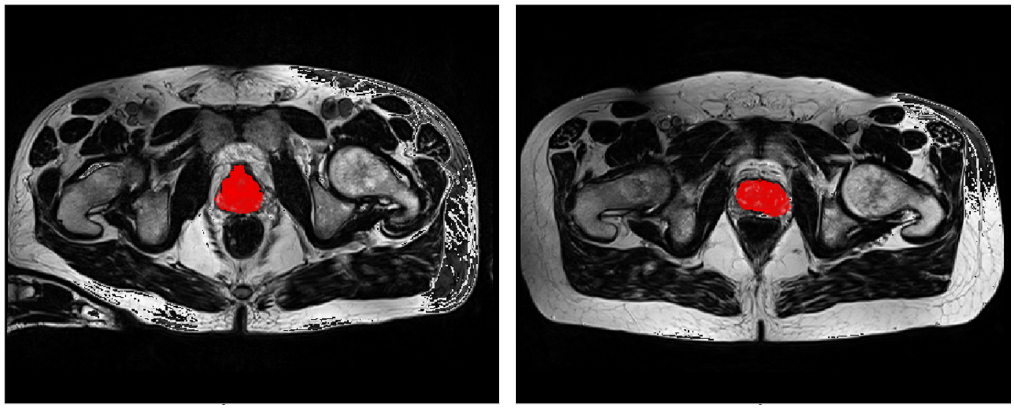


Figure 4.15: Transverse views of T2-weighted MR images belonging to the PROSPER dataset. The prostate segmentations generated by the UNet are shown in red, superimposed to the MR images.

After resizing the MR images to the size of the corresponding original reconstructed PET images, $50 \times 50 \times 50$ regions are cropped from the original reconstructed PET images around the prostate, using the MR segmentation as a reference. The results obtained applying the 3D CNN trained on simulated

data to the PROSPER PET images are shown in Fig. 4.16 and Fig. 4.17. Fig. 4.16 contains two examples where the PET scans were acquired using ^{68}Ga-PSMA as radiotracer. In all figures, the bladder is represented as a bright red area. In the first patient PET scan, presented in the first row, a region with increased activity is present next to the bladder. The corresponding CNN prediction also shows a region with higher activity concentration, although its distribution is more uniform and less localised. In the second row, the original PET scan does not show any uptake in the prostate region and the same can be observed in the image predicted by the network. Two examples of PET scans acquired using ^{11}C-Choline are shown in Fig. 4.17. The PET scan in the first row shows a small area with increased activity which is not recovered in the corresponding image predicted by the CNN. In the second row, the tumour in the original PET scan is bigger and more localised. The CNN predicts a lesion characterised by lower activity, but it also generates some high activity voxels in the area between the bladder and the lesion. One possible reason why the CNN cannot accurately recover prostate cancer lesions is that these images are quite different from the ones used for training. In the training dataset generated with the analytical simulator the background always has lower activity concentration than the tumours. On the contrary, in the PROSPER PET images the bladder always has higher uptake than the surrounding areas. Finally, in the PROSPER PET scans (especially the ones acquired with ^{68}Ga-PSMA) the increased activity concentration associated to the tumours is often spread over many voxels which do not form a well defined 3D structure. These lesions are different from the ones simulated in the training dataset, which are characterised by a more compact 3D shape.

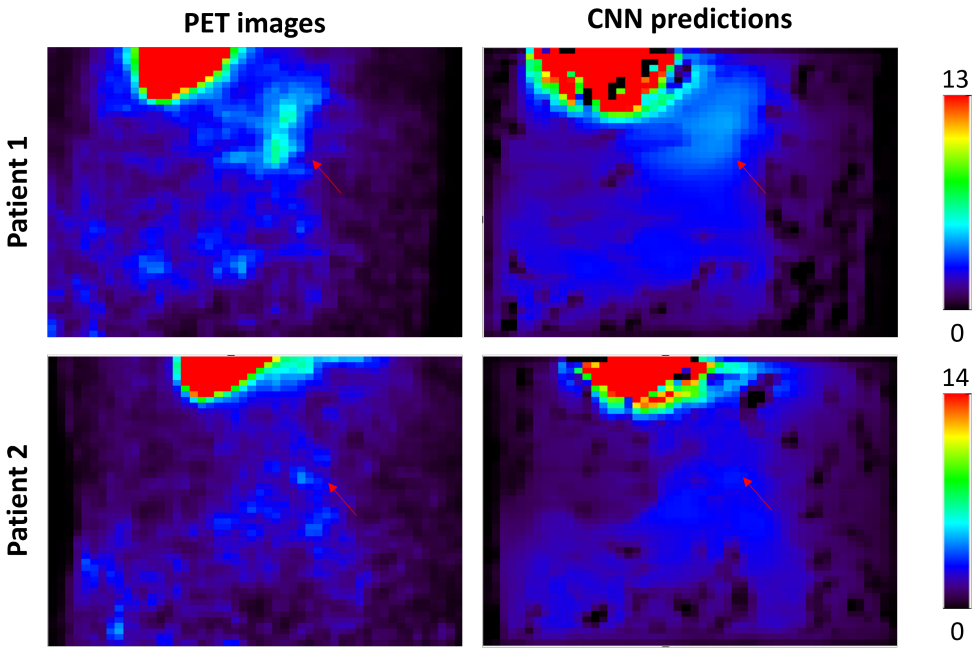


Figure 4.16: Sagittal views of PET images belonging to the PROSPER dataset, acquired using ^{68}Ga-PSMA as radiotracer. Images relative to different patients are presented in each row. The original reconstructed PET images are shown in the first column, whereas the CNN predictions are presented in the second column. The CNN predicts denoised images but it does not recover better defined tumour shapes. The images in the same row are shown using the same colour scale, which is expressed in terms of SUV.

4.4 Summary

In this chapter, we presented three applications of the proposed deep learning method on real data. The first experiment was performed on a dataset composed of 47 NSCLC patients, who received a ^{18}F-FDG PET/CT scan before chemo-radiotherapy. The CNN trained on simulated PET data was applied to $50 \times 50 \times 50$ voxels regions cropped around the metabolically active primary lung tumours. The visual assessment showed that CNN could predict denoised images where tumours were characterised by better defined shapes. The predicted tumours were characterised by more uniform activity distributions. To

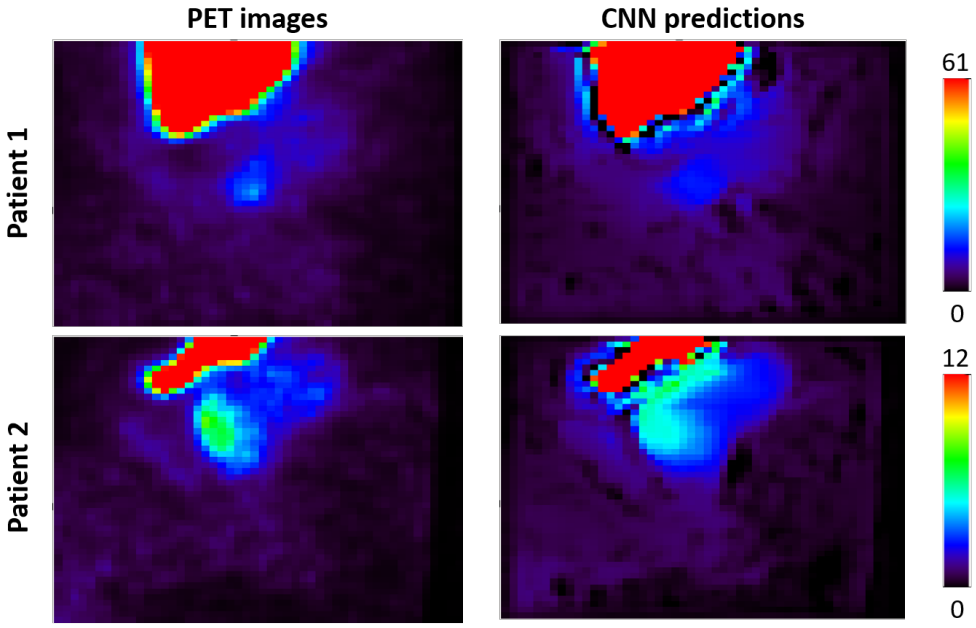


Figure 4.17: Sagittal views of PET images belonging to the PROSPER dataset, acquired using ^{11}C-Choline as radiotracer. Images relative to different patients are presented in each row. The original reconstructed PET images are shown in the first column, whereas the CNN predictions are presented in the second column. The CNN predicts denoised images but it does not recover better defined tumour shapes. The images in the same row are shown using the same SUV colour scale.

quantitatively assess our results the maximum, mean and peak SUVs were measured and subsequently compared to the same measurements performed on the original reconstructed PET images. The CNN predicted lower maximum activities within the tumours. The CNN also yielded higher estimates for the mean activities. This means that the network can denoise the images without "smoothing" the activity within the tumour as a Gaussian filter would. The peak activities predicted by the CNN were overall lower than the ones measured in the original reconstructed PET images.

The second application of the deep learning approach was performed on a Hodgkin lymphoma dataset. In this study, each patient received a baseline ^{18}F-FDG PET/CT scan before treatment and a follow-up scan after treatment.

The CNN trained on simulated PET images was applied to VOIs cropped around the bulkiest lesion in each of the patient PET scans. When the CNN was applied to the baseline scans, it predicted denoised images with better defined tumours. When applied to the follow-up scans, the CNN yielded denoised images but it was not always able to accurately recover the lesion shapes. One possible explanation for this result is that in the follow-up scans the lesions have a low activity concentration and they may not be characterised by a well defined shape, thus they are different from the activity distributions that the CNN learnt during training. The network yields lower maximum and peak activity concentration estimates when applied both to the baseline scans and to the follow-up scans. The CNN predicts overall higher mean activities than the ones measured in the original reconstructed PET images. The mean activities relative to the follow-up scans have similar values in the original PET scans and in the CNN predicted images.

Finally, the deep learning approach was applied to a prostate cancer dataset. A UNet was trained on MR T2-weighted images and on the corresponding prostate segmentations belonging to the PROMISE dataset. Once tested on an unseen set of MR T2-weighted images, the network could successfully segment the prostate. The UNet was subsequently applied to the PROSPER T2-weighted images to automatically predict the prostate segmentation masks. These masks were used to locate the prostate in the PET scans and to automatically crop a VOI around the prostate from the original reconstructed PET images. The CNN trained on simulated data was then applied to the cropped original reconstructed PET images to recover the ground truth activity distribution and tumour shape. In this case, the CNN could not accurately recover the tumour shapes and activity concentrations. This may be due to the fact that the lesions in the PROSPER dataset are different from the lesions in the training dataset, both in terms of shape and activity distribution. The pres-

ence of the bladder, which is characterised by a high uptake, may also hamper the recovery of tumours in the prostate. One possible solution to improve the CNN performance could be threshold the original reconstructed PET images so that a low activity is assigned to the bladder. Another possibility would be to train the CNN on a dataset that better represents the PROSPER PET data.

The CNN trained on simulated data performed differently when applied to different clinical datasets, depending on the radioactivity distribution and image features that characterised the datasets. The experiments performed in this chapter allowed the identification of some factors that hamper the prediction of accurate images. These observations could be considered as a first step towards the identification of suitable clinical datasets that can be matched to our training data to obtain realistic predictions once the CNN is applied. Further tests are needed to assess the reliability of the proposed approach when applied to clinical data. Once tested on large datasets, in order to be used in clinical routine, the network should be able to recognise input PET images which are too different from the training dataset and may thus lead to incorrect predictions. When a suitable dataset is identified, a comparison between the SUV values and tumour volumes yielded by the CNN and the corresponding clinical measurements can be performed. Although we expect the results to be different, the comparison between these two sets of measurements can provide information of the network's performance.

Chapter 5
Discussion and Future Work

5.1 Discussion

This thesis has presented a deep learning method to improve radiotracer uptake and tumour shape definition on simulated PET data, phantom and real patient data. In this chapter, a general discussion about the deep learning method is provided, including limitations and future work perspectives.

5.1.1 Simulation of Tumour Shapes

The method used to simulate tumour-like shapes has some limitations. Real tumours are often characterised by heterogeneous composition and texture and by complex spatial structures. Mathematical models have been developed to describe specific aspects of cancer, but they do not capture all the characteristics of tumour biology. For this reason, it is very difficult to accurately simulate the complex biology of tumours. In general, there appears to be very little data in the literature as to what is the true 'ground truth radioactivity distribution' for different tumour types. Limited information can be obtained about tumour histology through biopsy, but it is difficult to correlate this information to the radiotracer distribution in the images due to the limited number of samples. A high-resolution map of the whole tumour radioactivity distribu-

tion can be obtained if the tumour is removed soon after injection, sliced up, and imaged with autoradiography, but this process is very difficult in every respect. In this thesis three radioactivity distribution patterns were assigned to the simulated tumour shapes. These are simplified representations of what can be observed in real patients PET scans. More heterogeneous radioactivity distributions would better represent the complex tumour activity patterns that are typically observed in PET scans. Future works includes the investigation of more complex tumour shapes, associated with more heterogeneous radioactivity distributions.

5.1.2 Simulation of PET Images

In this thesis, an analytical simulator was used to generate the PET images that were then given as an input to a 3D CNN. Although this simulation method provided more realistic PET data, some limitations need to be discussed. First of all, the PET images were created using the same spatially invariant PSF both for the generation of the PET raw data and for the PET image reconstruction. On the contrary, in a clinical setting PET images would be characterised by spatially variant PSFs, and there may be a mismatch between the system's PSF and the one modelled in the reconstruction algorithm at different areas of the field of view (i.e. notably at large radial offset) [89]. Further tests are needed to assess the performance of the deep learning approach when applied to PET images generated with PSFs that may not perfectly match the PSFs used to generate the PET raw data. Furthermore, in this thesis the PET images were generated using PSFs characterised by the same anisotropy. The robustness of the proposed method to data generated with PSFs with different anisotropies remains to be evaluated. When PET images were simulated using a range of PSFs, the Gaussian distribution from which the PSF FWHM values were drawn had a mean value of 4.5 mm, with

a range of ± 0.5 mm. Using this range of PSF values resulted in more variability in the PET images, but did not introduce a systematic bias. Further experiments are needed to assess whether drawing the PSF FHWM values from a distribution with an offset mean would negatively affect the network's performance. In this thesis the CNN was only trained on PET data generated and reconstructed with a single PSF, future work also includes training the network on PET images generated with a range of PSFs. Future experiments could investigate the impact of using a more diverse dataset for training, including PET images generated with different noise levels. Furthermore, having a larger training dataset could further improve the CNN performance. This could be achieved by generating more images with the analytical simulator or by augmenting the training dataset. The quantitative assessment on simulated data showed that small tumours were not accurately recovered by the CNN. One possible solution to address this problem could be to include more small tumour samples in the training dataset.

5.1.3 Application on Phantom and Real Patient Data

Our experiments showed that improved RCs could be obtained when applying the CNN to phantom PET data. The CNN improved the tumour shape definition and quantification in phantom PET images reconstructed with different algorithms. This is an encouraging results as it indicated that the network could potentially be applied to PET data that do not exactly match the data used to train the CNN. The network trained on simulated data could denoise PET images belonging to a lymphoma dataset (RATHL) and to a NSCLC dataset. When the network was applied to the prostate cancer dataset it could not accurately recover the tumour shapes and radioactivity concentrations. This may be due to the fact that these tumours and other nearby features in the patient data had significantly different characteristics to the

ones used for training. We conclude that it is important to carefully choose the datasets on which to apply the CNN, so that the real PET images are not too different from the ones used for training. Given the difficulty of obtaining radiotracer distribution ground truth data, a possible way to assess the network performance is by seeing if it can contribute directly to predicting patient outcome. Such approaches in general require very large datasets that include clinical information in addition to imaging. Datasets available for PET are on the whole quite small, though future 'big-data' approaches that can access very large numbers of routine clinical scans might make this possible.

5.1.4 Network Optimisation

In each neural network, many different parameters need to be tuned in order to achieve the best results. Network optimisation is still not well understood and modern techniques are usually heuristic and based on user experience. The approach chosen in this thesis was to tune a set of network parameters manually, by only changing one parameter at a time and subsequently selecting the best performing network at every step. The analysis was focused on few key aspects such as the network depth, the number of epochs and the optimiser. Preliminary tests were performed changing other parameters such as learning rate, batch size and number of filters assigned to each layer. A visual assessment suggested that fine tuning these parameters did not hugely affect the results. The performance of the CNN may be further refined by optimising these parameters as well. Grid search could also be used to improve the optimisation, thus building and evaluating a CNN model for each possible combination of all the parameters. As a final result the grid search algorithm would output the combination that achieved the best score on the validation dataset. The disadvantage of using grid search is that there is no guarantee that this method finds the perfect solution for the test dataset. Furthermore,

this method is time consuming and computationally intensive when applied to large datasets.

5.1.5 Deep Learning Limitations

Supervised deep learning approaches are affected by some intrinsic limitations. It is difficult to fully evaluate supervised deep learning approaches applied to oncologic medical images, as the 'ground truth' tumour structure, texture and radioactivity distribution are unknown (see Sec. 5.1.3). This limitation is not unique to deep learning approaches and also applies to, for example, image reconstruction algorithms which are usually evaluated using phantom data at end. In some cases biopsy may give more information on the tumour composition and aggressiveness but it is extremely difficult to correlate this information to what is observed in medical images as the location of the limited number of biopsy samples cannot be accurately mapped to image coordinates. Another limitation of deep learning methods is that they work as a 'black box' and it is difficult to understand the underlying decision-making process. Although advances have been made in the development of explainable AI models, this limitation still hampers their applicability in the clinical routine. Deep learning models can fail when applied to rare cases or abnormal lesions that differ from the examples learnt during training. It is thus necessary to isolate these instances and to report these results separately. Although large labelled datasets are often not available in medical imaging, improved performances can be obtained leveraging the great amount of heterogeneous medical data that is generated every day in hospitals. The integration of this information is hampered by the fact that results obtained through a range of exams may be acquired in different hospitals, stored separately and thus may be difficult to access. Aggregating information obtained from radiology, pathology and genomic exams may lead to a better patient stratification and disease man-

agement. A lot of work has been done in recent years to produce publicly available large multi- modal datasets acquired in multiple institutions (i.e UK Biobank), that can be used to train neural networks. Deep learning methods applied to large datasets composed of multi-modal data correlated with patient outcome and pathology may provide valuable information on the underlying mechanisms of disease onset and progression.

5.2 Conclusions

To conclude this thesis a summary of the main contributions is given, together with an overview of possible future perspectives. A 3D deep learning approach was developed with the aim to improve the quantification of tumour radiotracer activity in PET images. In this work:

- A CNN was trained and optimised on a dataset composed of PET images simulated by adding noise and by applying a Gaussian filter to the ground truth images. The CNN, tested on unseen simulated PET images, predicted denoised images with better defined tumour shapes. The CNN provided improved mean and median RCs. The proposed deep learning method outperformed standardly used image post-processing techniques such as Gaussian filtering and LR deconvolution.

- The 3D network was trained and optimised on a PET dataset created using an analytical simulator. In this dataset the PET images were generated and reconstructed using a single PSF. The optimised CNN predicted better defined tumour shapes and improved RC estimates. The CNN outperformed standardly used image post-processing techniques such as Gaussian filtering or LR deconvolution.

- The deep learning approach was successfully applied to PET images gen-

erated with multiple PSFs, yielding denoised images with better defined tumours. The results obtained on PET data generated with multiple PSFs suggest that this method can potentially improve quantification of tumour radioactivity concentration when applied to data acquired with different scanners.

- The CNN trained on simulated PET images was applied to phantom PET images reconstructed with different algorithms, yielding denoised images with better defined tumours.

- The CNN trained on simulated PET images, was applied to three real patients datasets. The CNN predicted denoised PET images and improved tumour SUV_{max} and SUV_{mean} values when applied to the NSCLC dataset and to the lymphoma (RATHL) datasets.

- The CNN failed to recover prostate lesions when applied to the PROSPER dataset, probably due to the disparity between training and testing data

The proposed deep learning model has the potential to reduce the inter and intra- observer variability that affects quantitative measurements such as SUV_{mean}. Using a deep learning method for the recovery of PET images potentially reduces the time needed for the quantitative assessment of images, as the network only needs to be trained once and it can then recover enhanced clinical PET images in a short time. One advantage of this deep learning method is that it can be applied retrospectively as a post-processing step, thus list-mode data are not required.

5.2.1 Future Prospects

The proposed deep learning method could be integrated in an end-to-end pipeline that allows the automated prediction of tumour radiotracer activity concentration. Furthermore, this method could be applied to larger images, such as whole body PET images. The network could be applied to patches extracted from larger PET images, and the enhanced image patches predicted by the CNN could then be composed to form a bigger image. In this case, PET images only containing healthy tissue would probably need to be included in the training dataset, so that the network could learn how to enhance images that do not contain a tumour. This approach would enable the estimation of tumour burden in whole body PET scans. In a clinical setting, this deep learning method could also be integrated in the image reconstruction process, so that clinicians would be presented with the original PET image together with the enhanced image predicted by the CNN.

The deep learning approach developed in this thesis could be further expanded and applied to larger datasets that integrate multimodal images. The advantage of a more heterogeneous dataset would be to exploit the complementarity of the information extracted from different imaging modalities. For example, MR data simulated using the analytical simulator and paired with the corresponding simulated PET data could be given as an input to a neural network. The network would then be trained to simultaneously output an enhanced PET image together with a tumour segmentation derived from MR. One aspect that needs to be considered in this case is that in clinical practice PET and MR images are affected by motion, so image registration methods may need to be applied as a pre-processing step. Another aspect is that tumour cells may not be characterised by a high radiotracer concentration, so the tumour boundaries in PET and MR may not correspond.

Future work could also include developing a regression network that di-

rectly predicts quantitative values such as SUV instead of predicting PET images. Preliminary experiments were performed to train a 3D CNN that could directly yield SUV values as output, but the network's predictions were not accurate. Further experiments are needed to understand why this approach did not generate accurate SUV predictions.

Publications

Publications

During the course of my PhD, I have contributed to the publication of the following papers:

- L. Dal Toso, E. Pfaehler, R. Boellaard, J. A. Schnabel, P. K. Marsden. Deep Learning Based Approach to Quantification of PET Tracer Uptake in Small Tumors. In: Knoll F., Maier A., Rueckert D., Ye J. (eds) Machine Learning for Medical Image Reconstruction. MLMIR 2019. Lecture Notes in Computer Science, vol 11905. Springer, Cham

- A. Valladares, S. Ahangari, T. Beyer, R. Boellaard, Z. Chalampalakis, C. Comtat, L. Dal Toso, A. E. Hansen, M. Koole, J. Mackewn, P.K Marsden, J. Nuyts, F. Padormo, R. Peeters, S. Poth , E. Solari, I. Rausch. Clinically Valuable Quality Control for PET/MRI Systems: Consensus Recommendation From the HYBRID Consortium. 2019, Frontiers in Physics , vol.7.

- L. Dal Toso, Z. Chalampalakis, C. Comtat, I. Buvat, J. A. Schnabel, P. K. Marsden. Improved 3D Tumour Definition and Quantification of Uptake in Simulated Lung Tumours Using Deep Learning. *Accepted - Physics in Medicine and Biology.*

Conference Presentations

During the course of this PhD, the following works have been presented at international conferences:

- L. Dal Toso, E. Pfaehler, R. Boellaard, J. A. Schnabel, P. K. Marsden. Deep Learning Based Approach to Quantification of PET Tracer Uptake in Small Tumors. *Machine Learning for Medical Image Reconstruction (MLMIR), held in conjunction with MICCAI*. Shenzhen, China, 2019.

- L. Dal Toso, E. Pfaehler, R. Boellaard, G. Cook, V. Goh, A. Schnabel, P. K. Marsden. Deep Learning based approach to quantification of PET tracer uptake in small tumours. *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Manchester, United Kingdom, 2019.

- L. Dal Toso, Z. Chalampalakis, C. Comtat, I. Buvat, J. A. Schnabel, P. K. Marsden, Improved 3D Tumour Definition and Quantification of Radiotracer Uptake Using Deep Learning. *European Association of Nuclear Medicine (EANM) Congress*, Virtual, 2020.

Bibliography

[1] A. Reader, G. Corda-D'Incan, A. Mehranian, C. da Costa-Luis, S. Ellis, and J. Schnabel, "Deep Learning for PET Image Reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences* **PP** (2020) 1–1.

[2] I. Häggström, C. R. Schmidtlein, G. Campanella, and T. J. Fuchs, "DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem," *Medical Image Analysis* **54** (2019) 253–262.

[3] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal* **79** (1974) 745.

[4] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Am.* **62** no. 1, (1972) 55–59.

[5] J. Yang, S. C. Huang, M. Mega, K. P. Lin, A. W. Toga, G. W. Small, and M. E. Phelps, "Investigation of partial volume correction methods for brain FDG PET studies," *IEEE Transactions on Nuclear Science* **43** no. 6, (1996) 3322–3327.

[6] K. Erlandsson, I. Buvat, P. Pretorius, B. Thomas, and B. Hutton, "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology," *Physics in medicine and biology* **57** (2012) R119–59.

[7] F. Hofheinz, J. Maus, B. Beuthien-Baumann, L. Oehme, J. Steinbach, J. Kotzerke, and J. Hoff, "Suitability of bilateral filtering for

edge-preserving noise reduction in PET," *EJNMMI research* **1** (2011) 23.

[8] J. Dutta, R. M. Leahy, and Q. Li, "Non-local means denoising of dynamic PET images," *PLOS ONE* **8** no. 12, (2013) 1–15.

[9] K. Gong, J. Guan, C. C. Liu, and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences* **3** no. 2, (2019) 153–161.

[10] Z. Xu, M. Gao, G. Z. Papadakis, B. Luna, S. Jain, D. J. Mollura, and U. Bagci, "Joint solution for PET image segmentation, denoising, and partial volume correction," *Medical Image Analysis* **46** (2018) 229–243.

[11] W. Lu, J. A. Onofrey, Y. Lu, L. Shi, T. Ma, Y. Liu, and C. Liu, "An investigation of quantitative accuracy for deep learning based denoising in oncological PET," *Physics in Medicine & Biology* **64** no. 16, (2019) 165019.

[12] J. Schaefferkoetter, U. Metser, and P. Veit-Haibach, "Deep learning for improving image quality with noisy PET data," *Journal of Nuclear Medicine* **60** no. supplement 1, (2019) 245–245.

[13] T. A. Song, S. R. Chowdhury, F. Yang, and J. Dutta, "Super-resolution PET imaging using convolutional neural networks," *IEEE Transactions on Computational Imaging* **6** (2020) 518–528.

[14] V. J. Cunningham, V. W. Pike, D. Bailey, C. A. Freemantle, B. C. Page, A. K. Jones, M. J. Kensett, D. Bateman, S. K. Luthra, and T. Jones, "A method of studying pharmacokinetics in man at picomolar drug concentrations," **32** (1991) 167–72.

[15] Y. F. Tai and P. Piccini, "Applications of positron emission tomography (PET) in neurology," *Journal of Neurology, Neurosurgery & Psychiatry* **75** no. 5, (2004) 669–676.

[16] I. Sarikaya, "PET studies in epilepsy," *Am J Nucl Med Mol Imaging* **12** (2015) 416–430.

[17] D. Le Guludec, R. Lautamäki, J. Knuuti, J. Bax, and F. Bengel, "Present and future of clinical cardiovascular PET imaging in Europe - a position statement by the european council of nuclear cardiology (ECNC)," *European Journal of Nuclear Medicine and Molecular Imaging* **35** no. 9, (2008) 1709–1724.

[18] J. Machac, "Cardiac positron emission tomography imaging," *Seminars in Nuclear Medicine* **35** no. 1, (2005) 17–36. Cardiovascular Nuclear Medicine.

[19] R. Driessen, P. Raijmakers, W. Stuijfzand, and P. Knaapen, "Myocardial perfusion imaging with PET," *The International Journal of Cardiovascular Imaging* **33** no. 7, (2017) 1021–1031.

[20] L. Mcloughlin, F. O'Kelly, C. O'Brien, M. Sheikh, J. Feeney, W. Torreggiani, and J. Thornhill, "The improved accuracy of planar bone scintigraphy by adding single photon emission computed tomography (SPECT-CT) to detect skeletal metastases from prostate cancer," *Irish journal of medical science* **185** (2014) 101–105.

[21] G. Cook, G. Azad, and V. Goh, "Imaging bone metastases in breast cancer: Staging and response assessment," *Journal of Nuclear Medicine* **57** (2016) 27S–33S.

[22] S. M. Moriguchi, K. Hiromoto Koga, P. H. Alves Togni, and M. J. Dos Santos, "Clinical applications of nuclear medicine," *Medical Imaging in Clinical Practice* (2013) .

[23] M. T. Madsen, "Recent advances in SPECTimaging," *Journal of Nuclear Medicine* **48** no. 4, (2007) 661–673.

[24] P. N. T. Wells and H.-D. Liang, "Medical ultrasound: imaging of soft tissue strain and elasticity," *Journal of The Royal Society Interface* **8** no. 64, (2011) 1521–1549.

[25] F. Faraji and R. C. Gaba, "Radiologic modalities and response assessment schemes for clinical and preclinical oncology imaging," *Frontiers in Oncology* **9** (2019) 471.

[26] T. Yankeelov, R. Abramson, and C. Quarles, "Quantitative multimodality imaging in cancer research and therapy," *Nature reviews. Clinical oncology* **11** (2014) 670–680.

[27] B. F. Hutton and M. Braun, "Software for image registration: Algorithms, accuracy, efficacy," *Seminars in Nuclear Medicine* **33** no. 3, (2003) 180–192.

[28] T. Beyer, L. Freudenberg, D. Townsend, and J. Czernin, "The future of hybrid imaging-part 1: Hybrid imaging technologies and SPECT/CT," *Insights into imaging* **2** (2011) 161–9.

[29] T. Jones and D. Townsend, "History and future technical innovation in positron emission tomography," *Journal of Medical Imaging* **4** (2017) 011013.

[30] R. Schmitz, A. Alessio, and P. Kinahan, "The physics of PET / CT scanners," 2013.

[31] T. Beyer, D. Townsend, and T. Blodgett, "Dual-modality PET/CT tomography for clinical oncology," *The quarterly journal of nuclear medicine* **46** (2002) 24–34.

[32] F. Ul-Hassan and G. J. Cook, "PET/CT in oncology," *Clinical Medicine* **12** no. 4, (2012) 368–372.

[33] R. F. Muzic and F. P. DiFilippo, "Positron emission tomography-magnetic resonance imaging: Technical review," *Seminars in Roentgenology* **49** no. 3, (2014) 242–254. Combining PET with MRI – Part I.

[34] H. Wood, "PET-MRI scanner opens new frontier in medical imaging," *Medical Technology Business Europe* (2010) .

[35] T. Beyer, N. Schwenzer, S. Bisdas, C. Claussen, and B. Pichler, "MR / PET – hybrid imaging for the next decade," 2010.

[36] J. Martí-Climent, E. Prieto, V. Morán, L. Rodríguez, M. Rodriguez, J. Arbizu, M. Garcia-Velloso, and J. Richter, "Effective dose estimation

for oncological and neurological PET/CT procedures," *EJNMMI Research* **7** (2017) 37.

[37] Y. Li, L. Jiang, H. Wang, H. Cai, Y. Xiang, and L. Li, "Effective radiation dose of 18F-FDG PET/CT: how much does diagnostic CT constribute?," *Radiation Protection Dosimetry* **187** no. 2, (2019) 183–190.

[38] E. C. Ehman, G. B. Johnson, J. E. Villanueva-Meyer, S. Cha, A. P. Leynes, P. E. Z. Larson, and T. A. Hope, "PET/MRI: Where might it replace PET/CT?," *Journal of Magnetic Resonance Imaging* **46** no. 5, (2017) 1247–1262.

[39] M. Al-Tamimi and G. Sulong, "Tumor brain detection through MR images: A review of literature," *Journal of Theoretical and Applied Information Technology* **62** (2014) .

[40] A. J. Taylor, M. Salerno, R. Dharmakumar, and M. Jerosch-Herold, "T1 mapping: Basic techniques and clinical applications," *JACC: Cardiovascular Imaging* **9** no. 1, (2016) 67–81.

[41] J. Zwanenburg, J. Hendrikse, F. Visser, T. Takahara, and P. Luijten, "Fluid attenuated inversion recovery (FLAIR) MRI at 7.0 Tesla: Comparison with 1.5 and 3.0 Tesla," *European radiology* **20** (2009) 915–922.

[42] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts, "MR pulse sequences: What every radiologist wants to know but is afraid to ask," *RadioGraphics* **26** no. 2, (2006) 513–537.

[43] U. A. van der Heide, M. Frantzen-Steneker, E. Astreinidou, M. E. Nowee, and P. J. van Houdt, "MRI basics for radiation oncologists," *Clinical and Translational Radiation Oncology* **18** (2019) 74–79.

[44] M. Potter, E. Newport, and K. J. Morten, "The Warburg effect: 80 years on," *Biochemical Society Transactions* **44** no. 5, (2016) 1499–1505.

[45] M. Unterrainer, C. Eze, H. Ilhan, S. Marschner, O. Roengvoraphoj, N.-S. Schmidt-Hegemann, F. Walter, W. Kunz, P. Munck af Rosenschold, R. Jeraj, N. Albert, A.-L. Grosu, M. Niyazi, P. Bartenstein, and C. Belka, "Recent advances of PET imaging in clinical radiation oncology," *Radiation oncology (London, England)* **15** (2020) 88.

[46] S. S. Anand, H. Singh, and A. K. Dash, "Clinical applications of PET and PET-CT," *Med J Armed Forces India* **65** (2009) 353–358.

[47] A. Zhu, D. Lee, and H. Shim, "Metabolic positron emission tomography imaging in cancer detection and therapy response," *Semin Oncol.* **38** (2011) 55–69.

[48] M. M. Abouzied, E. S. Crawford, and H. A. Nabi, "18F-FDG imaging: Pitfalls and artifacts," *Journal of Nuclear Medicine Technology* **33** no. 3, (2005) 145–155.

[49] E. Safaie, R. Matthews, and R. Bergamaschi, "PET scan findings can be false positive," *Tech Coloproctol* **19** (2015) 329–330.

[50] H. Zhuang and A. Alavi, "18-Fluorodeoxyglucose positron emission tomographic imaging in the detection and monitoring of infection and inflammation," *Seminars in Nuclear Medicine* **32** no. 1, (2002) 47–59.

[51] S. Asad, S. Aquino, N. Piyavisetpat, and A. Fischman, "False-positive FDG positron emission tomography uptake in nonmalignant chest abnormalities," *AJR. American journal of roentgenology* **182** (2004) 983–989.

[52] M. Truong, C. Viswanathan, B. Carter, O. Mawlawi, and E. Marom, "PET/CT in the thorax: pitfalls," *Radiologic clinics of North America* **52** (2014) 17–25.

[53] N. Long and C. Smith, "Causes and imaging features of false positives and false negatives on F-PET/CT in oncologic imaging," *Insights into imaging* **2** (2011) 679–698.

[54] G. Giovacchini, M. Picchio, E. Coradeschi, V. Scattoni, V. Bettinardi, C. Cozzarini, M. Freschi, F. Fazio, and C. Messa, "[(11)C]choline uptake with PET/CT for the initial diagnosis of prostate cancer: relation to PSA levels, tumour stage and anti-androgenic therapy," *Eur J Nucl Med Mol Imaging* **35** (2008) 1065–1073.

[55] B. Scher, M. Seitz, W. Albinger, R. Tiling, M. Scherr, H.-C. Becker, M. Souvatzogluou, F. J. Gildehaus, H.-J. Wester, and S. Dresel, "Value of 11C-choline PET and PET/CT in patients with suspected prostate cancer," *European journal of nuclear medicine and molecular imaging* **34** (2007) 45–53.

[56] J. Young, M. Jauregui-Osoro, W. Wong, M. Cooper, G. Cook, S. Barrington, M. Ma, P. Blower, and E. Aboagye, "An overview of nuclear medicine research in the UK and the landscape for clinical adoption," *Nuclear Medicine Communications* (2021) .

[57] M. P. Yadav, S. Ballal, R. K. Sahoo, S. N. Dwivedi, and C. Bal, "Radioligand therapy with 177Lu-PSMA for metastatic castration-resistant prostate cancer: A systematic review and meta-analysis," *American Journal of Roentgenology* **213** no. 2, (2019) 275–285.

[58] A. G. Harris, A. I. Vinik, T. M. O'Dorisio, and M. S. O'Dorisio, "Radioligand theranostics in the management of neuroendocrine tumors," *Pancreas* **49** (2020) 599–603.

[59] J. E. Maxwell and J. R. Howe, "Imaging in neuroendocrine tumors: an update for the clinician," *International Journal of Endocrine Oncology* **2** no. 2, (2015) 159–168.

[60] Z. Keidar, R. Gill, E. Goshen, O. Israel, T. Davidson, M. Morgulis, N. Pirmisashvili, and S. Ben-Haim, "68Ga-PSMA PET/CT in prostate cancer patients – patterns of disease, benign findings and pitfalls," *Cancer Imaging* **18** (39, 2018) .

[61] W. Jones, K. Griffiths, P. C. Barata, and C. J. Paller, "PSMA theranostics: Review of the current status of PSMA-targeted imaging

and radioligand therapy," *Cancers* **12** no. 6, (2020) 1367. http://dx.doi.org/10.3390/cancers12061367.

[62] M. Conti and L. Eriksson, "Physics of pure and non-pure positron emitters for PET: A review and a discussion," *EJNMMI Physics* **3** (2016) 8.

[63] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, "Chapter 18 - Positron Emission Tomography," in *Physics in Nuclear Medicine (Fourth Edition)*, S. R. Cherry, J. A. Sorenson, and M. E. Phelps, eds., pp. 307–343. W.B. Saunders, Philadelphia, fourth edition ed., 2012.

[64] S. Vandenberghe, E. Mikhaylova, E. D'Hoe, P. Mollet, and J. S. Karp, "Recent developments in time-of-flight PET," *EJNMMI Phys* **3** (2016) 3.

[65] P. Razifar, M. Sandström, H. Schnieder, B. Langstrom, E. Maripuu, E. Bengtsson, and M. Bergström, "Noise correlation in PET, CT, SPECT and PET/CT data evaluated using autocorrelation function: a phantom study on data, reconstructed using FBP and OSEM," *BMC medical imaging* **5** (2005) 5.

[66] M. S. Rzeszotarski, "The AAPM/RSNA physics tutorial for residents," *RadioGraphics* **19** no. 3, (1999) 765–782.

[67] S. Meikle and R. Badawi, "Quantitative techniques in PET," in *Positron Emission Tomography*, D. Bailey, P. Townsend, D.W.and Valk, and M. Maisey, eds., pp. 93–126. Springer, London, 2005.

[68] S. Cherry, M. Dahlbom, and E. Hoffman, "3D PET using a conventional multislice tomograph without septa," *Journal of Computer Assisted Tomography* **15** no. 4, (1991) 655–668.

[69] M. A. Lodge, R. D. Badawi, R. Gilbert, P. E. Dibos, and B. R. Line, "Comparison of 2-dimensional and 3-dimensional acquisition for 18F-FDG PET oncology studies performed on an LSO-based scanner," *Journal of Nuclear Medicine* **47** no. 1, (2006) 23–31.

[70] R. D. Badawi, P. K. Marsden, B. F. Cronin, J. L. Sutcliffe, and M. N. Maisey, "Optimization of noise-equivalent count rates in 3D PET," *Physics in Medicine and Biology* **41** no. 9, (1996) 1755–1776.

[71] D. Visvikis, D. Griffiths, D. Costa, J. Bomanji, and P. Ell, "Clinical evaluation of 2D versus 3D whole-body PET image quality using a dedicated BGO PET scanner," *European journal of nuclear medicine and molecular imaging* **32** no. 9, (2005) 1050—1056.

[72] National Electrical Manufacturers Association (NEMA), "Performance measurements of positron emission tomographs (PET)," *NEMA Standards Publication, NEMA NU 2-2018* (2018) 1–41.

[73] T. R. DeGrado, T. G. Turkington, J. J. Williams, C. W. Stearns, J. M. Hoffman, and R. E. Coleman, "Performance characteristics of a whole-body PET scanner," *Journal of Nuclear Medicine* **35** no. 8, (1994) 1398–1406.

[74] K. Gong, S. R. Cherry, and J. Qi, "On the assessment of spatial resolution of PET systems with iterative image reconstruction," *Physics in Medicine and Biology* **61** no. 5, (2016) N193–N202.

[75] W. W. Moses, "Fundamental limits of spatial resolution in PET," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **648** (2011) S236–S240.

[76] C. S. Levin and E. J. Hoffman, "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution," *Phys. Med. Biol.* **44** (1999) 781–99.

[77] C. Burger, G. Goerres, S. Schoenes, A. Buck, A. Lonn, and G. Schulthess, "PET attenuation coefficients from CT images: Experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients," *European journal of nuclear medicine and molecular imaging* **29** (2002) 922–7.

[78] C. J. Thompson, A. Dagher, D. N. Lunney, S. C. Strother, and A. C. Evans, "A technique to reject scattered radiation in PET transmission scans," in *Physics and Engineering of Computerized Multidimensional Imaging and Processing*, T. F. Budinger, Z.-H. Cho, and O. Nalcioglu, eds., vol. 0671, pp. 244 – 253, International Society for Optics and Photonics. SPIE, 1986.

[79] C. J. Thompson, R. Lecomte, and J. Cadorette, "Feasibility of using beta-gamma coincidence for 3D PET attenuation correction," *IEEE Transactions on Nuclear Science* **47** no. 3, (2000) 1176–1181.

[80] M. L. Camborde, C. J. Thompson, D. Togane, Nan Zhang, and A. Reader, "A positron-decay triggered transmission source for positron emission tomography," *IEEE Transactions on Nuclear Science* **51** no. 1, (2004) 53–57.

[81] H. Zaidi and B. Hasegawa, "Determination of the attenuation map in emission tomography," *Journal of Nuclear Medicine* **44** no. 2, (2003) 291–315.

[82] H. Zaidi and M.-L. Montandon, "Scatter compensation techniques in PET," *Pet Clinics* **2** (2007) 219–234.

[83] E. Hoffman, S. Huang, M. Phelps, and D. Kuhl, "Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences," *Journal of computer assisted tomography* **5** no. 3, (1981) 391—400.

[84] D. G. Politte and D. L. Snyder, "Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography," *IEEE Transactions on Medical Imaging* **10** no. 1, (1991) 82–89.

[85] D. Brasse, P. Kinahan, C. Lartizien, C. Comtat, M. Casey, and C. Michel, "Correction methods for random coincidences in fully 3D whole-body PET: Impact on data and image quality," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **46** (2005) 859–67.

[86] F. H. Fahey, "Data acquisition in PET imaging," *Journal of Nuclear Medicine Technology* **30** no. 2, (2002) 39–49.

[87] S. Cherry, M. Dahlbom, and M. Phelps, "PET: Physics, instrumentation, and scanners," in *PET*. Springer, New York, NY, 2004.

[88] S. Tong, A. M. Alessio, and P. E. Kinahan, "Image reconstruction for PET/CT scanners: past achievements and future challenges," *Imaging in Medicine* **2** no. 5, (2010) 529–545.

[89] A. Rahmim, J. Qi, and V. Sossi, "Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls," *Medical Physics* **40** no. 6Part1, (2013) 064301, `https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.4800806`.

[90] D. W. Wilson and B. M. W. Tsui, "Noise properties of filtered-backprojection and ML-EM reconstructed emission tomographic images," *IEEE Transactions on Nuclear Science* **40** no. 4, (1993) 1198–1203.

[91] A. Teymurazyan, T. Riauka, H.-S. Jans, and D. Robinson, "Properties of noise in positron emission tomography images reconstructed with filtered-backprojection and row-action maximum likelihood algorithm," *Journal of digital imaging* **26** (2012) 447–456.

[92] M. Soret, S. Bacharach, and I. Buvat, "Partial-volume effect in PET tumor imaging," *Journal of Nuclear Medicine* **48** (2007) 932–945.

[93] E. Hoffman, S. Huang, and M. Phelps, "Quantitation in positron emission computed tomography: 1. Effect of object size," *Journal of computer assisted tomography* **3** no. 3, (1979) 299—308.

[94] T. Meechai, S. Tepmongkol, and C. Pluempitiwiriyawej, "Partial-volume effect correction in positron emission tomography brain scan image using super-resolution image reconstruction," *The British Journal of Radiology* **88** no. 1046, (2015) 20140119.

[95] H. Zaidi and N. Karakatsanis, "Towards enhanced PET quantification in clinical oncology," *The British Journal of Radiology* **91** no. 1081, (2018) 20170508.

[96] R. Wahl, H. Jacene, Y. Kasamon, and M. Lodge, "From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors," *Journal of Nuclear Medicine* **50 Suppl 1** (2009) 122S–50S.

[97] M. Adams, T. Turkington, J. Wilson, and Z. Wong, "A systematic review of the factors affecting accuracy of SUV measurements," *American journal of roentgenology* **195** (2010) 310–20.

[98] P. Ziai, M. R. Hayeri, A. Salei, A. Salavati, S. Houshmand, A. Alavi, and O. M. Teytelboym, "Role of optimal quantification of FDG PET imaging in the clinical practice of radiology," *RadioGraphics* **36** no. 2, (2016) 481–496.

[99] H. Im, T. Bradshaw, M. Solaiyappan, and S. Cho, "Current methods to define metabolic tumor volume in positron emission tomography: Which one is better?," *Nuclear Medicine and Molecular Imaging* **52** no. 1, (2018) 5–15.

[100] S. Chakraborty, R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, T. D. Kelley, D. Braines, M. Sensoy, C. J. Willis, and P. Gurram, "Interpretability of deep learning models: A survey of results," in *2017 IEEE (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 1–6. 2017.

[101] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences* **116** no. 44, (2019) 22071–22080.

[102] S. S. Kunapuli, P. Bh, and U. Singh, "Enhanced medical image de-noising using auto encoders and MLP," in *Intelligent Technologies and Applications*, I. Bajwa, F. Kamareddine, and A. Costa, eds.,

vol. 932, pp. 3–15, INTAP 2018. Communications in Computer and Information Science. Springer, Singapore, 2019.

[103] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 241–246. 2016.

[104] J. Cui, K. Gong, N. Guo, C. Wu, X. Meng, K. Kim, K. Zheng, Z. Wu, L. Fu, B. Xu, Z. Zhu, J. Tian, H. Liu, and Q. Li, "PET image denoising using unsupervised deep learning," *European Journal of Nuclear Medicine and Molecular Imaging* **46** (12, 2019) 2780–2789.

[105] C. T. Sari and C. Gunduz-Demir, "Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images," *IEEE Transactions on Medical Imaging* **38** no. 5, (2019) 1139–1149.

[106] C. Vununu, S.-H. Lee, and K.-R. Kwon, "A strictly unsupervised deep learning method for HEp-2 cell image classification," *Sensors* **20** no. 9, (2020) 2717.

[107] S. Kazeminia, C. Baur, A. Kuijper, B. van Ginneken, N. Navab, S. Albarqouni, and A. Mukhopadhyay, "GANs for medical image analysis," *Artificial Intelligence in Medicine* **109** (2020) 101938.

[108] C. Yu, J. Liu, and S. Nemati, "Reinforcement learning in healthcare: A survey," *arXiv e-prints* (2019) arXiv:1908.08796.

[109] W. Zhao, J. Yang, Y. Sun, C. Li, W. Wu, L. Jin, Z. Yang, B. Ni, P. Gao, P. Wang, Y. Hua, and M. Li, "3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas," *Cancer Research* **78** (2018) 6881–6889.

[110] G. Kang, K. Liu, B. Hou, and N. Zhang, "3D multi-view convolutional neural networks for lung nodule classification," *PLOS ONE* **12** (2017) e0188290.

[111] M. A. Heuvelmans, P. M. van Ooijen, S. Ather, C. F. Silva, D. Han, C. P. Heussel, W. Hickes, H.-U. Kauczor, P. Novotny, H. Peschl,

M. Rook, R. Rubtsov, O. von Stackelberg, M. T. Tsakok, C. Arteta, J. Declerck, T. Kadir, L. Pickup, F. Gleeson, and M. Oudkerk, "Lung cancer prediction by deep learning to identify benign lung nodules," *Lung Cancer* **154** (2021) 1–4.

[112] D. Wallis, C. Nioche, M. Soussan, M. Lacroix, P. Akl, and I. Buvat, "Automated detection of mediastinal cancer nodes from 18F-FDG PET/CT scans of NSCLC patients using radiomic and CNN features," *Journal of Nuclear Medicine* **61** no. supplement 1, (2020) 507–507.

[113] D. Ardila, A. Kiraly, S. Bharadwaj, B. Choi, J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine* **25** (2019) 954–961.

[114] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific Reports* **8** (2018) 4165.

[115] L. Shen, L. Margolies, J. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," *Scientific Reports* **9** (2019) 1–12.

[116] P. Herent, B. Schmauch, P. Jehanno, O. Dehaene, C. Saillard, C. Balleyguier, J. Arfi-Rouche, and S. Jégou, "Detection and characterization of MRI breast lesions using deep learning," *Diagnostic and Interventional Imaging* **100** no. 4, (2019) 219–225.

[117] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks.," *Nature* **542** no. 7639, (2017) 115–118.

[118] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis* **35** (2017) 18–31.

[119] A. Adegun and S. Viriri, *Deep Learning Model for Skin Lesion Segmentation: Fully Convolutional Network*, vol. 11663, pp. 232–242. Springer, Cham, 2019.

[120] Y. Chen, L. Xing, H. Bagshaw, M. Buyyounouski, and B. Han, "Deep learning-based intraprostatic lesion segmentation using multi-parametric MRI for prostate radiation therapy," *International Journal of Radiation Oncology*Biology*Physics* **108** (2020) S100.

[121] E. Arvaniti, K. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. Wild, J. Rüschoff, and M. Claassen, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Scientific Reports* **8** (2018) 12054.

[122] K. Nagpal, D. Foote, Y. Liu, C. Po-Hsuan, E. Wulczyn, F. Tan, N. Olson, J. Smith, A. Mohtashamian, J. Wren, G. Corrado, R. MacDonald, L. Peng, M. Amin, A. Evans, A. Sangoi, C. Mermel, J. Hipp, and M. Stumpe, "Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer," *npj Digit. Med* **2** no. 48, (2018) .

[123] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[124] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[125] R. Yamashita, M. Nishio, R. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging* **9** (2018) .

[126] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* **86** no. 11, (1998) 2278–2324.

[127] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Information Processing Systems* **25** (2012) .

[128] P. Dutta, P. Upadhyay, M. De, and R. G. Khalkar, "Medical image analysis using deep convolutional neural networks: CNN architectures and transfer learning," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 175–180. 2020.

[129] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 117–122. 2018.

[130] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array* **3-4** (2019) 100004.

[131] A. Meyer, G. Chlebus, M. Rak, D. Schindele, M. Schostak, B. van Ginneken, A. Schenk, H. Meine, H. K. Hahn, A. Schreiber, and C. Hansen, "Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI," *Computer Methods and Programs in Biomedicine* **200** (2021) 105821.

[132] Q. Ke, J. Liu, M. Bennamoun, S. An, F. Sohel, and F. Boussaid, "Chapter 5 - computer vision for human–machine interaction," in *Computer Vision for Assistive Healthcare*, M. Leo and G. M. Farinella, eds., Computer Vision and Pattern Recognition, pp. 127–145. Academic Press, 2018.

[133] Z. Qian, T. L. Hayes, K. Kafle, and C. Kanan, "Do we need fully connected output layers in convolutional networks?," 2020.

[134] 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. Lecture Notes in Computer Science*, N. Navab, J. Hornegger, W. M. , and A. F. Frangi, eds., vol. 9351, pp. 234–241, Cham. Springer, 2015.

[135] U. Baid, S. Talbar, S. Rane, S. Gupta, M. H. Thakur, A. Moiyadi, N. Sable, M. Akolkar, and A. Mahajan, "A novel approach for fully automatic intra-tumor segmentation with 3D U-Net architecture for gliomas," *Frontiers in Computational Neuroscience* **14** (2020) 10.

[136] S. Baek, Y. He, B. Allen, J. Buatti, B. Smith, L. Tong, Z. Sun, J. Wu, M. Diehn, B. Loo, K. Plichta, S. Seyedin, M. Gannon, K. Cabel, Y. Kim, and X. Wu, "Deep segmentation networks predict survival of non-small cell lung cancer," *Scientific Reports* **9** (2019) 17286.

[137] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, and M. Andre, "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," *Biomedical Signal Processing and Control* **61** (2020) 102027.

[138] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[139] S. Guo and Z. Yang, "Multi-Channel-ResNet: An integration framework towards skin lesion analysis," *Informatics in Medicine Unlocked* **12** (2018) 67–74.

[140] Q. Wang, F. Shen, L. Shen, J. Huang, and W. Sheng, "Lung nodule detection in CT images using a raw patch-based convolutional neural network," *Journal of Digital Imaging* **32** (2019) 1–9.

[141] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[142] J. Hai, K. Qiao, J. Chen, H. Tan, J. Xu, L. Zeng, D. Shi, and B. Yan, "Fully convolutional DenseNet with multiscale context for automated breast tumor segmentation," *Journal of Healthcare Engineering* **2019** (2019) 1–11.

[143] B. Liu, Y. Zhao, B. Yang, S. Zhao, R. Gu, and M. Gahegan, "A gastric cancer recognition algorithm on gastric pathological sections based on multistage attention-DenseNet," *Concurrency and Computation: Practice and Experience* (2021) e6188.

[144] B. He, D. Dong, Y. She, C. Zhou, M. Fang, Y. Zhu, H. Zhang, Z. Huang, T. Jiang, J. Tian, and C. Chen, "Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker," *Journal for ImmunoTherapy of Cancer* **8** no. 2, (2020) e000550.

[145] S. Bekisz and L. Geris, "Cancer modeling: From mechanistic to data-driven approaches, and from fundamental insights to clinical applications," *Journal of Computational Science* **46** (2020) 101198.

[146] E. Pfaehler, R. Beukinga, J. Jong, R. Slart, C. Slump, R. Dierckx, and R. Boellaard, "Repeatability of 18 F-FDG PET radiomic features: a phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method," *Medical Physics* **46** (2018) 665–678.

[147] A. M. Loening and S. S. Gambhir, "AMIDE: A free software tool for multimodality medical image analysis," *Molecular Imaging* **2** no. 3, (2003) 131–137.

[148] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on* **13** (2004) 600 – 612.

[149] M. Vanderhoek, S. B. Perlman, and R. Jeraj, "Impact of the definition of peak standardized uptake value on quantification of treatment response," *Journal of Nuclear Medicine* **53** no. 1, (2012) 4–11.

[150] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical Physics* **37** no. 9, (2010) 4902–4915.

[151] S. Stute, C. Tauber, C. Leroy, M. Bottlaender, V. Brulon, and C. Comtat, "Analytical simulations of dynamic PET scans with realistic count rates properties," pp. 1–3. 2015.

[152] U. Bashir, M. Siddique, S. Dhillon, N. Patel, P. Bassett, D. Landau, V. Goh, and G. Cook, "The effects of segmentation algorithms on the measurement of 18F-FDG PET texture parameters in non-small cell

lung cancer," *European Journal of Nuclear Medicine and Molecular Imaging Research* **7** no. 60, (2017) .

[153] S. Morzenti, E. De Ponti, L. Guerra, A. Zorz, C. Landoni, C. Crivellaro, F. Elisei, L. Baratto, M. Arosio, and A. Crespi, "Performance evaluation of the Discovery IQ - GE PET/CT scanner according to NEMA NU2-2012 standard," *Journal of Nuclear Medicine* **56** no. supplement 3, (2015) 1846–1846.

[154] M. Teras, T. Tolvanen, J. J. Johansson, J. Williams, and J. Knuuti, "Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT," *Eur J. Nuc. Med. Mol. Imaging* **34** no. 10, (2007) 1683–1692.

[155] D. Hellwig, T. P. Graeter, D. Ukena, A. Groeschel, G. W. Sybrecht, H.-J. Schaefers, and C.-M. Kirsch, "18F-FDG PET for mediastinal staging of lung cancer: Which SUV threshold makes sense?," *Journal of Nuclear Medicine* **48** no. 11, (2007) 1761–1766.

[156] J. Russell, A. Collins, A. Fowler, M. Karanth, C. Saha, S. Docherty, J. Padayatty, K. Maw, I. Lentell, L. Cooke, A. Hodson, N. Shah, S. Sadullah, N. Grigoropoulos, W. Qian, A. Kirkwood, B. Uttenthal, P. Johnson, and G. Follows, "Advanced-stage hodgkin lymphoma in the east of england cancer network: A 10-year comparative analysis of outcomes for patients treated with ABVD or escalated BEACOPP in the non-trial setting compared with age-matched patients treated in the multinational RATHL trial," *Annals of Hematology* **100** (2017) .

[157] S. Meissner, J.-C. Janssen, V. Prasad, G. Diederichs, B. Hamm, W. Brenner, and M. Makowski, "Accuracy of standard clinical 3T prostate MRI for pelvic lymph node staging: Comparison to 68Ga-PSMA PET-CT," *Scientific Reports* **9** (2019) .

[158] I. Caglic, V. Kovac, and T. Barrett, "Multiparametric MRI - local staging of prostate cancer and beyond," *Radiology and Oncology* **53** (2019) 159 – 170.

[159] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "PSNet: prostate segmentation on MRI based on a convolutional neural network," *Journal of Medical Imaging* **5** no. 2, (2018) 1 – 6.

[160] B. Turkbey, S. V. Fotin, R. J. Huang, Y. Yin, D. Daar, O. Aras, M. Bernardo, B. E. Garvey, J. Weaver, H. Haldankar, N. Muradyan, M. J. Merino, P. A. Pinto, S. Periaswamy, and P. L. Choyke, "Fully automated prostate segmentation on MRI: Comparison with manual segmentation methods and specimen volumes," *American Journal of Roentgenology* **201** no. 5, (2013) W720–W729.

[161] D. K. Lee, D. J. Sung, C.-S. Kim, Y. Heo, J. Y. Lee, B. J. Park, and M. J. Kim, "Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula," *American Journal of Roentgenology* **214** no. 6, (2020) 1229–1238.

[162] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman, and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge," *Medical Image Analysis* **18** no. 2, (2014) 359–373.

[163] I. Mirzaev, "Fully convolutional neural network with residual connections for automatic segmentation of prostate structures from MR images," 2017.

[164] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," *CoRR* **abs/1608.04117** (2016) .

[165] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR* **abs/1606.04797** (2016) .

Appendix

This section contains further examples relative to the clinical datasets presented in chapter 4. These visual result support the evaluations that have been presented in the main body of the thesis.

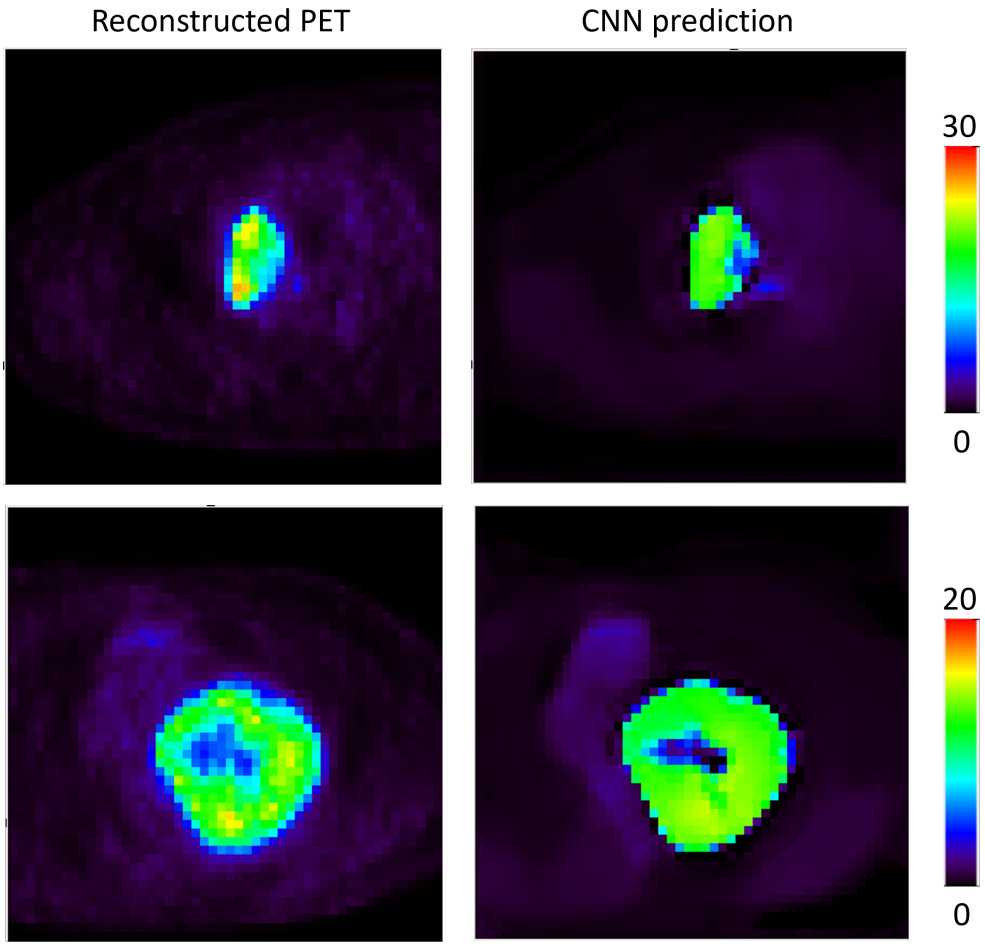


Figure 5.1: Transverse views of two NSCLC lesions. The CNN predicts denoised images, where the tumours have better defined shapes and more uniform activity distributions. In each row the images are shown with the same SUV colour scale.

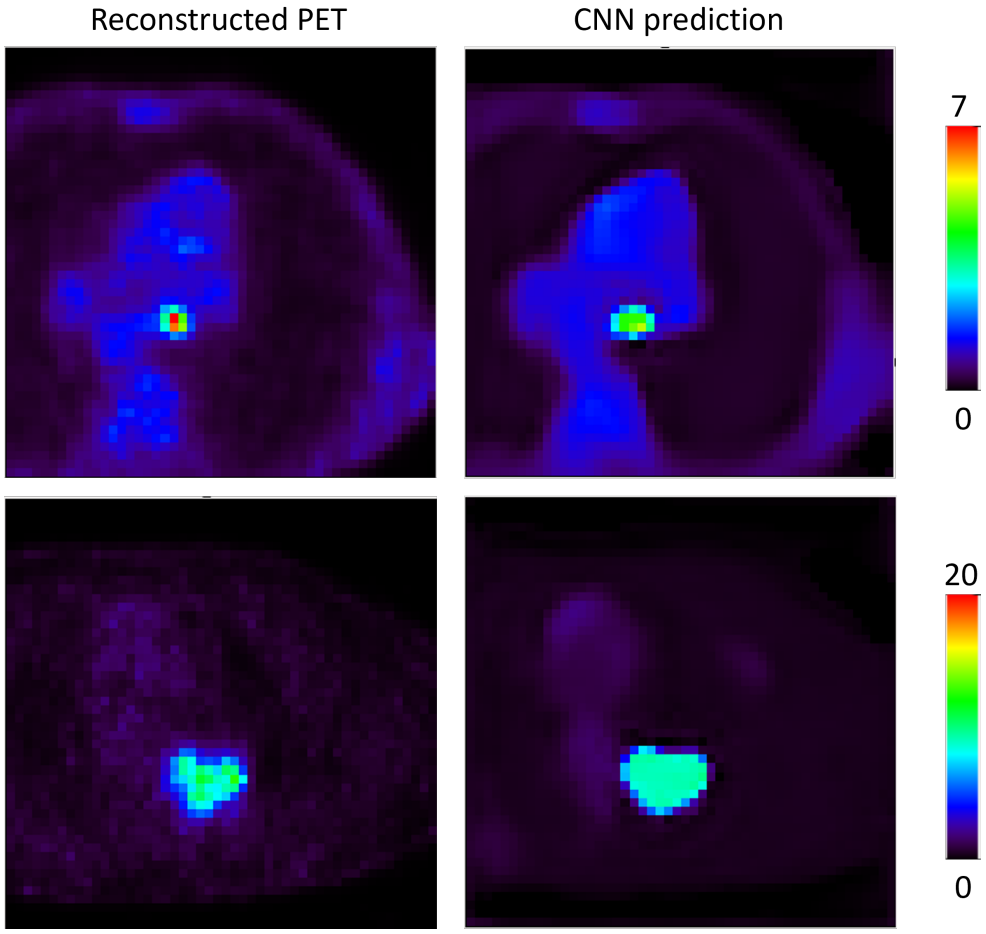


Figure 5.2: Transverse views of two NSCLC lesions. The CNN predicts denoised images, where the tumours have better defined shapes and more uniform activity distributions. In each row the images are shown with the same SUV colour scale.

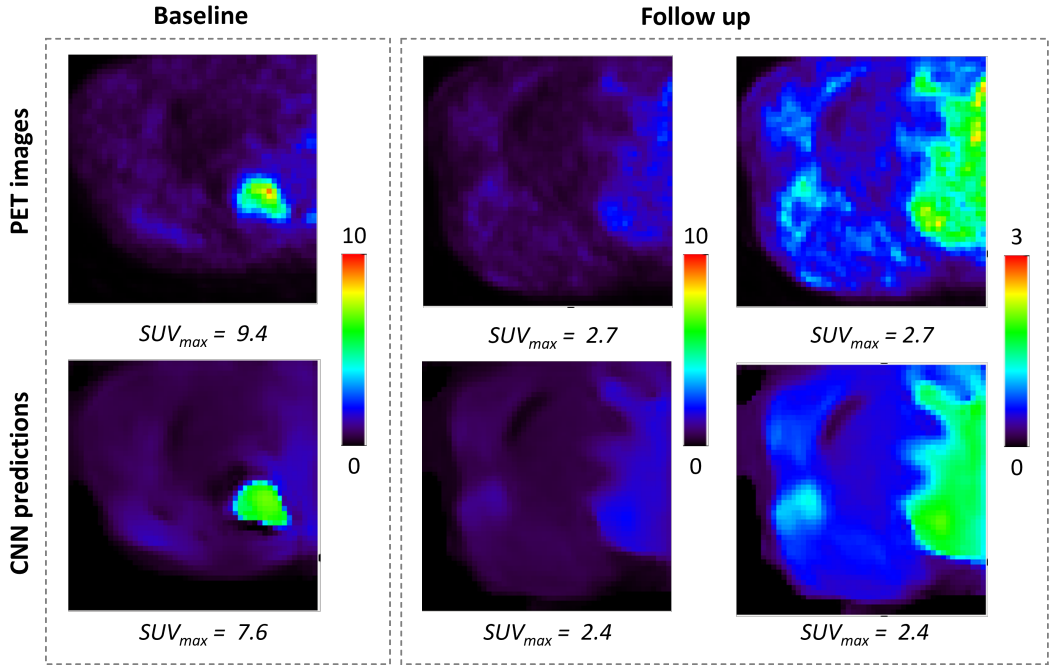


Figure 5.3: Transverse views of PET images of a lymphoma patient. The baseline scan and the follow-up scan are shown in the top row. The corresponding images predicted by the CNN are presented in the bottom row. The CNN predicts denoised images. The images in the same column are shown using the same SUV colour scale.

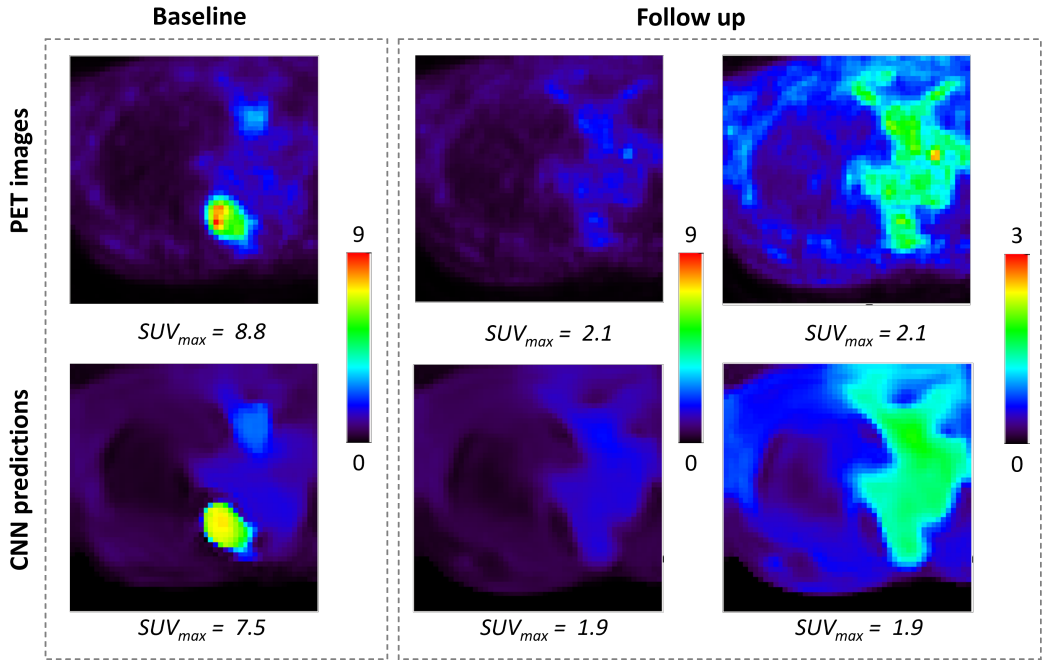


Figure 5.4: Transverse views of PET images of a lymphoma patient. The baseline scan and the follow-up scan are shown in the top row. The corresponding images predicted by the CNN are presented in the bottom row. The CNN predicts denoised images. The images in the same column are shown using the same SUV colour scale.

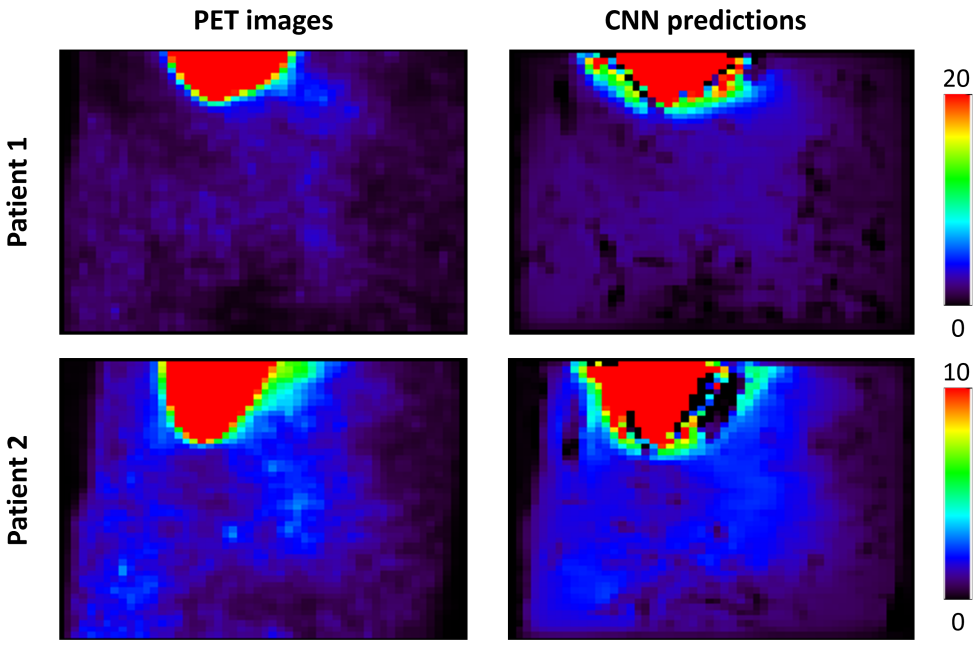


Figure 5.5: Sagittal views of PET images belonging to the PROSPER dataset, acquired using PSMA as radiotracer. Images relative to different patients are presented in each row. The original reconstructed PET images are shown in the first column, whereas the CNN predictions are presented in the second column. The CNN predicts denoised images but it does not recover better defined tumour shapes. The images in the same row are shown using the same SUV colour scale.

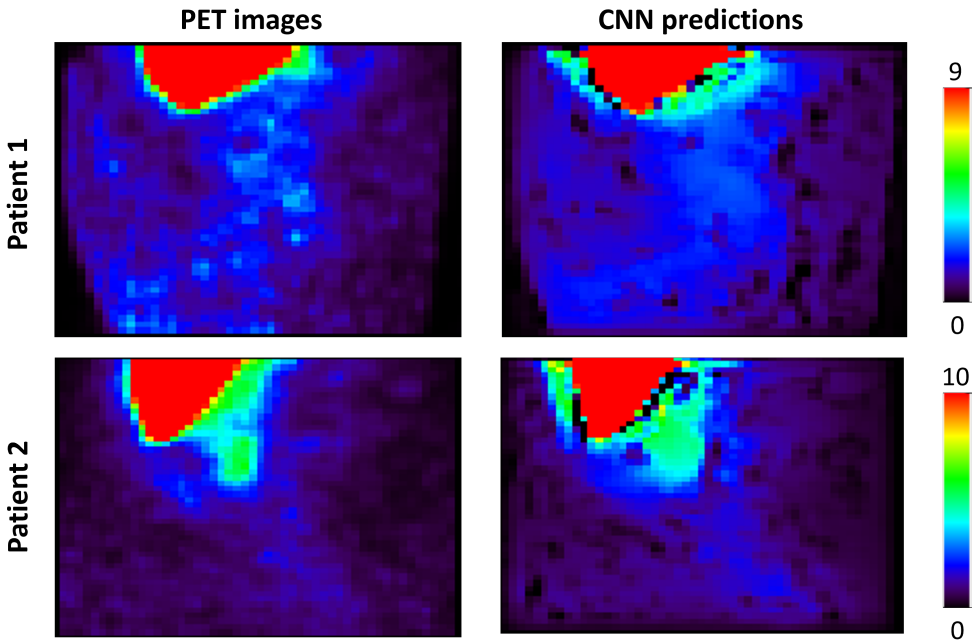


Figure 5.6: Sagittal views of PET images belonging to the PROSPER dataset, acquired using ^{11}C-Choline as radiotracer. Images relative to different patients are presented in each row. The original reconstructed PET images are shown in the first column, whereas the CNN predictions are presented in the second column. The CNN predicts denoised images but it does not recover better defined tumour shapes. The images in the same row are shown using the same SUV colour scale.