

**ACCEPTED MANUSCRIPT • OPEN ACCESS**

# Improved 3D tumour definition and quantification of uptake in simulated lung tumours using deep learning

To cite this article before publication: Laura Dal Toso *et al* 2022 *Phys. Med. Biol.* in press <https://doi.org/10.1088/1361-6560/ac65d6>

## Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2022 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

# Improved 3D Tumour Definition and Quantification of Uptake in Simulated Lung Tumours Using Deep Learning

Laura Dal Toso<sup>1</sup>, Zacharias Chalampalakis<sup>2,6</sup>, Irène Buvat<sup>3</sup>, Claude Comtat<sup>2</sup>, Gary Cook<sup>1</sup>, Vicky Goh<sup>1</sup>, Julia A Schnabel<sup>1,4,5</sup> and Paul K Marsden<sup>1</sup>

<sup>1</sup> School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

<sup>2</sup> Laboratoire d'Imagerie Biomédicale Multimodale (BioMaps), Université Paris-Saclay, CEA, CNRS, Inserm, Service Hospitalier Frédéric Joliot, Orsay, France

<sup>3</sup> Laboratoire d'Imagerie Translationnelle en Oncologie, Inserm, Institut Curie, Orsay, France

<sup>4</sup> Helmholtz Center Munich, Munich, Germany

<sup>5</sup> Faculty of Informatics, Technical University of Munich, Munich, Germany

<sup>6</sup> ABX-CRO Advanced Pharmaceutical Services Forschungsgesellschaft M.B.H., Dresden, Germany.

E-mail: laura.dal\_toso@kcl.ac.uk

## Abstract.

**Objective** In clinical positron emission tomography (PET) imaging, quantification of radiotracer uptake in tumours is often performed using semi-quantitative measurements such as the standardised uptake value (SUV). For small objects, the accuracy of SUV estimates is limited by the noise properties of PET images and the partial volume effect. There is need for methods that provide more accurate and reproducible quantification of radiotracer uptake.

**Approach** In this work, we present a deep learning approach with the aim of improving quantification of lung tumour radiotracer uptake and tumour shape definition. A set of simulated tumours, assigned with “ground truth” radiotracer distributions, are used to generate realistic PET raw data which are then reconstructed into PET images. In this work, the ground truth images are generated by placing simulated tumours characterised by different sizes and activity distributions in the left lung of an anthropomorphic phantom. These images are then used as input to an analytical simulator to simulate realistic raw PET data. The PET images reconstructed from the simulated raw data and the corresponding ground truth images are used to train a 3D convolutional neural network.

**Results** When tested on an unseen set of reconstructed PET phantom images, the network yields improved estimates of the corresponding ground truth. The same network is then applied to reconstructed PET data generated with different point spread functions. Overall the network is able to recover better defined tumour shapes and improved estimates of tumour maximum and median activities.

**Significance** Our results suggest that the proposed approach, trained on data simulated with one scanner geometry, has the potential to restore PET data acquired with different scanners.

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*  
3  
4  
5  
6  
7  
8  
9

10     *Keywords:* PET imaging, CNN, Analytical simulator Submitted to: *Phys. Med. Biol.*  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

## Introduction

12     Positron emission tomography (PET) is an imaging modality which is extensively used  
13     in oncology to detect and stage tumours, and to monitor response to treatment. In  
14     clinical routine, image interpretation is usually performed by visual inspection of PET  
15     images, which leads to inter- and intra-observer variability. Although in many cases  
16     visual assessment may be sufficient, more challenging tasks such as the evaluation of  
17     the response to therapy of solid tumours require some form of quantification [1]. In  
18     PET, uptake quantification is usually performed using semi-quantitative measurements  
19     of tumour radiotracer uptake, such as the standardised uptake value (SUV). The SUV  
20     is defined as the ratio of the radioactivity concentration in a region of interest (ROI)  
21     to the average radioactivity concentration in the whole body and can be calculated as  
22     follows:  
23  
24

$$25 \quad SUV = \frac{\text{Activity Concentration}_{ROI} (\text{kBq/ml})}{\text{Injected Activity (MBq) / Body Weight (kg)}} \quad (1)$$

26  
27     Usually a normalisation by a mass density of 1 (g/ml) is assumed, and SUV is  
28     presented as a dimensionless metric. Different SUV metrics, such as  $SUV_{max}$ ,  $SUV_{mean}$   
29     and  $SUV_{peak}$  can be defined. These semi-quantitative metrics are affected by the noise  
30     properties of PET images as well as partial volume effects (PVE), which compromise  
31     the accuracy and reproducibility of the measurements [1] [2]. The PVE, which results  
32     from the poor spatial resolution of PET scanners and from image sampling, degrades  
33     the quantitative accuracy of PET images and it can result in large biases on measures  
34     of tracer uptake in small tumours [3] [4]. Most commonly  $SUV_{max}$ , which is calculated  
35     using the most intense tumour voxel within the ROI, is reported.  $SUV_{max}$  is widely used  
36     as it is user-independent, but it is also affected by noise. A metric that is less dependent  
37     on noise is  $SUV_{mean}$ , but its disadvantage is that it depends on the delineation of the  
38     volume of interest (VOI) in which the measurement is performed.  $SUV_{peak}$  measures  
39     the average activity concentration within a VOI. Various definitions of  $SUV_{peak}$  using  
40     different VOI shapes, sizes and locations can be found in literature [5]. In this paper,  
41     the maximum  $SUV_{peak}$  value measured in a  $1 \text{ cm}^3$  spherical VOI within the tumour is  
42     reported.  $SUV_{peak}$  is less affected by image noise than  $SUV_{max}$ , but it presents some  
43     issues when applied to small tumours, especially if they are smaller than the VOI in  
44     which the peak is measured.  
45  
46

47     In this paper, PET images are reconstructed using the ordered subsets expectation  
48     maximisation (OSEM) algorithm. To improve PET image quality a model of the system  
49     resolution can be integrated in the reconstruction algorithm. Resolution modelling  
50     typically results in enhanced images where small lesions and narrow structures are  
51     characterised by a better contrast. Although resolution modelling can significantly  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     3  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14

improve image resolution and contrast it can also introduce artefacts near sharp edges in the PET image. These artefacts resemble the familiar Gibbs artefact and can lead to overestimation of quantitative indices - this is a particular issue for small lesions where PSF-reconstruction can lead to enhanced detectability but poor quantitative accuracy including overestimation of SUV max and peak indices.

There is need for reproducible methods that allow for more accurate quantification of tumour radiotracer uptake. Improved uptake quantification would lead to a more accurate assessment of response to treatment, especially at the early stages.

The use of artificial intelligence in the field of medical imaging has increased dramatically over the last decade. In PET imaging, machine learning and deep learning methods have been successfully applied to tumour segmentation, classification, automatic detection and image reconstruction [6, 7, 8, 9]. In recent years, deep learning methods have been used to denoise static PET images, and they have demonstrated better performance than traditional denoising approaches for various tracers and tasks. The two main deep learning architectures that have been used for denoising are convolutional neural networks (CNNs) [10] and generative adversarial networks (GANs) [11]. In previous work we developed a deep learning algorithm using a 3D CNN with the aim to improve quantification of tumour radiotracer uptake in simulated PET images [12]. The network was trained on simulated “ground truth” images that presented 3D shapes with typical tumour activity distributions found in clinical FDG images and on a corresponding set of simulated PET images. The network was able to robustly estimate the original activity, yielding improved images in terms of shape, activity distribution and quantification of activity. The main limitation of our previous work was that the PET images were simulated in a simplistic way, which did not take into account many of the effects that degrade the image quality in PET images.

Supervised deep learning methods require large amounts of labelled data, which are hard to obtain in PET imaging. Usually PET studies only comprise a relatively small number of patients. Furthermore, the true radionuclide distribution, which would correspond to the “label” of the PET images, is very difficult to obtain and rarely known. This limitation affects not only deep learning-based methods, but all the PET data processing methods (i.e. image reconstruction) which can never be fully evaluated *in vivo*. The use of phantoms and realistic PET simulators partially overcomes the lack of large labelled datasets. Monte Carlo simulation is the most commonly used technique to generate realistic PET data, but it has the disadvantage of being computationally very demanding. A number of analytical simulators have been developed to generate simulated PET images in a shorter time [13, 14]. While analytical simulators are not as accurate as Monte Carlo based simulators, they enable fast generation of PET data with realistic noise properties, which makes them particularly useful for creating large numbers of datasets. The simulation of realistic tumours also has some limitations. Real tumours are often characterised by inter- and intra-tumour heterogeneity, and by complex spatial structures. The mathematical and computational models of cancer that have been implemented so far mainly focus on describing a few specific aspects

## 1 2      *Improved 3D Tumour Definition and Quantification of Uptake*      4

3  
4      of the disease [15]. These models are not able to capture all the characteristics of  
5      tumour biology. In clinical practice, biopsies are performed to obtain ground truth  
6      information on tumour composition, but this method only provides limited information  
7      as the samples are extracted from a small region, and they cannot provide a description  
8      of the whole tumour. In this work, we generated a dataset of tumours assigned with  
9      three different activity patterns. The simulated dataset offers a wide variety of tumour  
10     activity values and tumour shapes, located at different positions within the left lung  
11     of an anthropomorphic phantom. This dataset is then used to train and test a 3D  
12     convolutional neural network (CNN) that can recover the real activity distribution from  
13     the signal seen on the PET image. We build on our previous deep learning approach  
14     described in [12] by significantly enhancing the simulation of PET images using an  
15     analytical simulator developed by Stute *et al.* [16] and by testing the proposed approach  
16     on three datasets simulated with different characteristics.  
17  
18

## 19      **Material and methods**

20  
21      An overview of the proposed deep learning approach is presented in Figure 1. Tumours  
22      with different sizes, shapes and activity distributions were simulated and subsequently  
23      placed in the left lung of an anthropomorphic phantom [17]. Different activities  
24      were assigned to each organ to create the “ground truth” images. Additionally, the  
25      corresponding attenuation maps were created. These two were used as input to  
26      an analytical PET simulator to generate PET raw data, which were subsequently  
27      reconstructed to provide the simulated PET images. The ground truth images and  
28      simulated PET images were used to train a 3D CNN.  
29  
30

### 31      *Simulation of ground truth images*

32  
33      At first, tumour-like 3D shapes with different volumes, spanning 0.01 to 200 ml, were  
34      simulated. This set of tumours was composed of three groups, each corresponding to  
35      a different activity pattern: tumours filled with uniform activity, tumours split into  
36      halves (each assigned with a different activity), and hollow tumours with background  
37      activity in the inner part, to mimic necrotic regions. The ratio between the activities  
38      assigned to the two halves of the heterogeneous tumours was variable, and the tumour  
39      activities ranged from 6 kBq/ml to 35 kBq/ml. The same activity range was used for  
40      the uniform tumours and for the hollow tumours. The thickness of the external layer  
41      of hollow tumours was set to half the radius of the tumour, and the inner core of the  
42      tumour was assigned with background activity. The background activity was set to  
43      1/10<sup>th</sup> of the maximum tumour activity, so in each image the tumour to background  
44      ratio was equal to 10. The choice of this specific set of activity patterns was based on  
45      the work of E. Pfaehler *et al.* [19], in which realistic phantom inserts were designed  
46      according to Non-Small Cell Lung Cancer (NSCLC) tumours extracted from patient  
47      studies. The simulated tumours were subsequently placed in random positions into the  
48  
49

1  
2      *Improved 3D Tumour Definition and Quantification of Uptake*      5  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

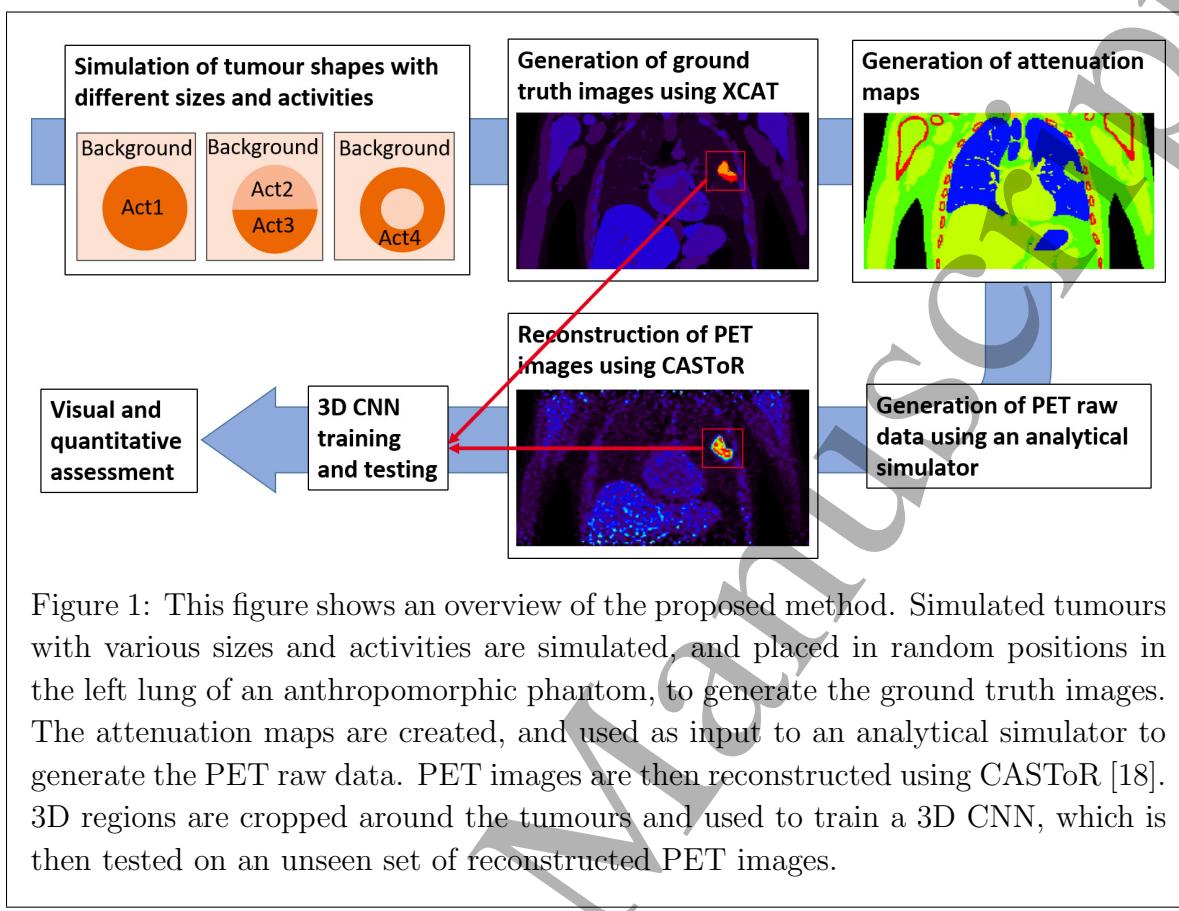


Figure 1: This figure shows an overview of the proposed method. Simulated tumours with various sizes and activities are simulated, and placed in random positions in the left lung of an anthropomorphic phantom, to generate the ground truth images. The attenuation maps are created, and used as input to an analytical simulator to generate the PET raw data. PET images are then reconstructed using CASToR [18]. 3D regions are cropped around the tumours and used to train a 3D CNN, which is then tested on an unseen set of reconstructed PET images.

left lung region of the XCAT phantom [17]. In order to shorten the simulation times, the XCAT phantom was reduced to  $344 \times 344 \times 127$  voxels including only the torso. The voxel size of the cropped image was  $[2.09, 2.09, 2.03] \text{ mm}^3$ . Realistic activity values drawn from a range of radioactivity concentrations measured from real patients PET images were assigned to each simulated tumour. Attenuation maps were generated for each XCAT image using the XCAT phantom attenuation values.

#### *PET data simulation and reconstruction*

To generate realistic PET raw data we used an analytical PET simulator developed by Stute *et al.* [16]. The geometry, detector resolution, and sensitivity of a positron emission tomography/magnetic resonance (PET/MR) system, more specifically the Siemens mMR scanner (Siemens Biograph mMR, Erlangen, Germany), were used in the simulation. In this work only the PET component of the simulator was modelled and we did not use any anatomical information from the MR. PET raw data were generated from the ground truth images each with 100 million total prompt counts, including scatter and random events. This corresponds to 12.4 million Noise Equivalent Counts (NEC). Subsequently, image reconstruction was performed using the open source fully quantitative reconstruction platform CASToR [18], with an iterative ordered subset

### 1 2      *Improved 3D Tumour Definition and Quantification of Uptake*      6 3 4 5

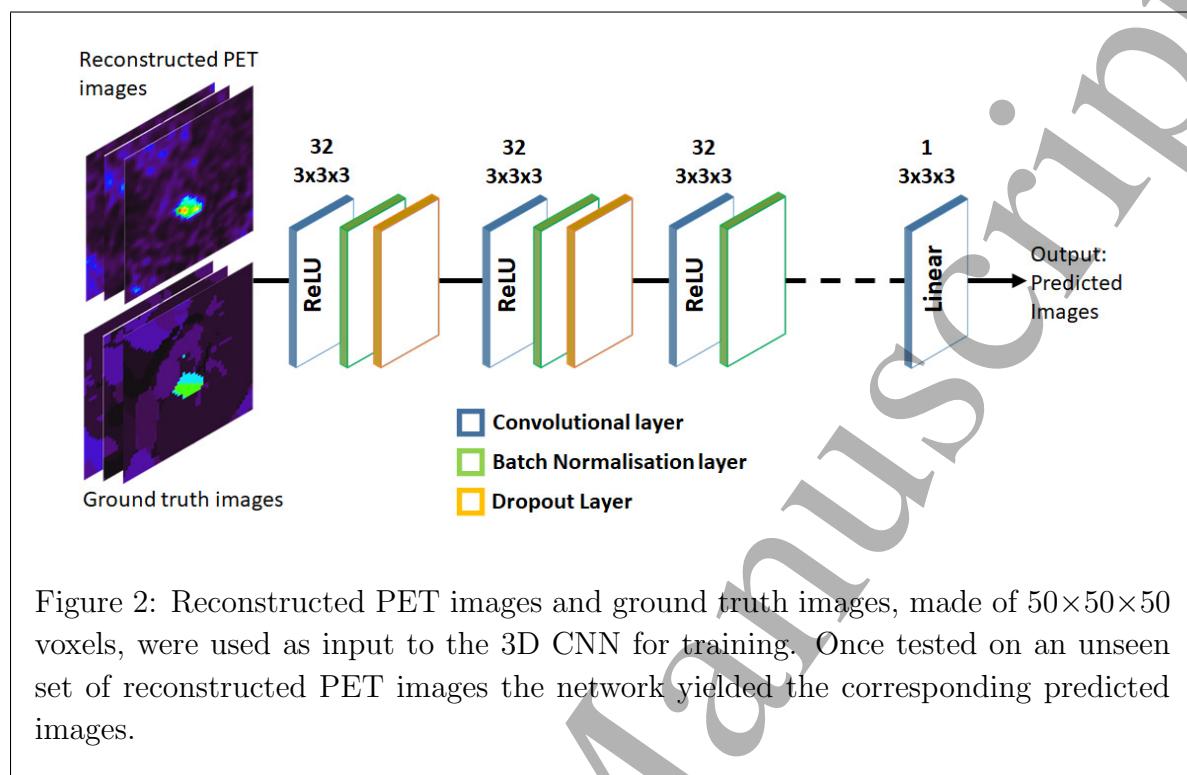
6  
7 expectation maximization (OSEM) algorithm run to 6 iterations with 21 subsets. The  
8 reconstructed voxel size was [2.09, 2.09, 2.03] mm<sup>3</sup>, as in the XCAT phantom images.  
9 Image reconstruction was performed with image-based PSF modelling, which is known  
10 to lead to increased SUV measurements (i.e.  $SUV_{max}$  and  $SUV_{mean}$ ) especially for  
11 small lesions [20]. No filter was applied to the images post reconstruction. Two  
12 different datasets were generated using the analytical simulator. The first one, called  
13 dataset 1, was composed of 800 uniform tumours, 721 tumours split in half and 589  
14 hollow tumours. In this case the analytical simulator generated the PET data with an  
15 anisotropic, spatially invariant PSF with FWHM (4.5, 4.5, 4.0) mm, which was also used  
16 in the reconstruction of the simulated PET images. A total of 2110 simulated PET raw  
17 datasets and images were generated.

18 One of the aims of this work was to test if the proposed algorithm trained on a  
19 given dataset could generalise well to data acquired with different scanners. One way  
20 to simulate the diversity between different scanners is to generate and reconstruct PET  
21 images using a range of PSFs. As a result, a new dataset (called dataset 2) composed  
22 of 100 images, with 33 uniform tumours, 34 tumours in halves and 33 hollow tumours  
23 was generated using a range of anisotropic spatially-invariant PSFs. A value randomly  
24 drawn from a Gaussian distribution (mean  $\mu = 4.5$  mm and sigma  $\sigma = 0.2$  mm) was  
25 assigned to the transaxial components of the PSF, both in the simulation and in the  
26 reconstruction, with a perfect match. The axial component was calculated by dividing  
27 this value by 1.125 and a constant ratio between transaxial and axial PSF components  
28 was maintained.

### 34 35      *Network architecture* 36

37 Convolutional neural networks (CNNs) are among the most commonly used algorithms  
38 for medical imaging applications [21]. These networks are composed of a series of  
39 convolutional layers, which can extract features from the input images, at multiple  
40 levels of abstraction. In this work 3D CNNs with different depths were tested and  
41 a visual and quantitative assessment suggested that a 3D CNN with 7 convolutional  
42 layers yielded the best results on our dataset. The proposed 3D CNN, presented in  
43 Figure 2, is composed of 7 convolutional layers each followed by a batch normalisation  
44 layer except for the final layer. The convolutional layers are characterised by 32 filters  
45 with dimensions  $3 \times 3 \times 3$ , and by ReLU activation functions except for the final one  
46 which has linear activation function. Two dropout layers, with dropout rate 0.3, were  
47 added after the first and second batch normalisation layers. Mean squared error was  
48 used as loss function during training and the optimizer was Adam [22]. The learning  
49 rate was set to the default value 0.001. The training was performed using a NVIDIA  
50 Tesla K40 GPU and the network architecture was implemented in the Keras Framework  
51 with Tensorflow [23].

1  
2      *Improved 3D Tumour Definition and Quantification of Uptake* 7  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



### Experiments

All the simulated PET images were cropped around the tumours to a final dimension of  $50 \times 50 \times 50$  voxels, before being processed by the convolutional neural network. The network inputs were normalised using the MinMaxScaler, which is a function provided in the scikit-learn Python package [24]. Using this function, the training, testing and validation datasets were scaled in the range [0,1]. The normalisation factors were stored and subsequently applied to the network's predictions to rescale the resulting images, before performing any quantitative analysis. In all experiments, a visual assessment of the predicted images was at first performed using a free software tool for multimodality medical image analysis (AMIDE) [25]. Subsequently, the images were quantitatively assessed. In order to provide a baseline comparison we used multiple metrics for the quantitative assessment of the results. The maximum, median and peak values were estimated for each tumour in order to quantitatively compare the reconstructed PET images to the images predicted by the CNN. Tumour masks defined on the ground truth images were subsequently applied to the reconstructed PET images and to the CNN's predicted images to measure the median and peak values. To present the results in a more compact way, the recovery coefficients defined as the ratio between the observed activity and the ground truth activity, were calculated using the maximum, median and peak values as shown in Eq. 2, Eq. 3 and Eq. 4 respectively.

$$RC_{max} = \frac{\text{Max activity}_{prediction}}{\text{Max activity}_{ground truth}} \quad (2)$$

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*  
3  
4

$$RC_{median} = \frac{Median\ activity_{prediction}}{Median\ activity_{ground\ truth}} \quad (3)$$

$$RC_{peak} = \frac{Peak\ activity_{prediction}}{Peak\ activity_{ground\ truth}} \quad (4)$$

10     When testing the network on simulated data, the structural similarity index  
11 measure (SSIM) was also calculated. The structural similarity allows the comparison of  
12 two images by taking into account their luminance  $l$ , contrast  $c$  and structure  $s$ . The  
13 SSIM is defined as:  
14

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma \quad (5)$$

18     where  $x$  and  $y$  are two non-negative image signals, which could be for example  
19 two image patches.  $\alpha$ ,  $\beta$  and  $\gamma$  are used to adjust the relative importance of the three  
20 components. In this paper,  $\alpha$ ,  $\beta$  and  $\gamma$  are equal to 1. The luminance comparison  
21 function  $l$  between two image signals depends on the mean intensities of the two signals  
22  $\mu_x$  and  $\mu_y$ . After estimating the luminance, the mean intensity is removed from the  
23 initial image signal. At this point, contrast comparison between the two signals is  
24 performed by measuring their standard deviations  $\sigma_x$  and  $\sigma_y$ . Finally, each signal is  
25 normalised by its standard deviation and the structure comparison is made on the  
26 normalised signals. In practice, when two images ( $X, Y$ ) are compared, the overall image  
27 quality can be estimated using the MSSIM, which it is expressed as:  
28

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M (SSIM(x_j, y_j)) \quad (6)$$

32     where  $x_j$  and  $y_j$  are the image contents at the  $j$ -th local window and  $M$  is the  
33 number of local windows in the image. In this work, the MSSIM was used to compare  
34 the shape and texture of the reconstructed PET images and the predicted images to  
35 the ground truth images. The window used to estimate the MSSIM were composed of  
36  $7 \times 7 \times 7$  voxels.  
37

43     *CNN training and testing using PET data generated with a single PSF*     This experiment  
44 was performed to optimise the network's architecture and test its performance on the  
45 simulated data. Dataset 1 was used to train and test the network. In this dataset,  
46 the PET images were generated and reconstructed using a spatially invariant PSF with  
47 FWHM (4.5 mm, 4.5 mm, 4.0 mm). The simulated images were split into training and  
48 testing datasets, with ratio 80/20. 20 % of the training data were used for validation.  
49 The training dataset was augmented by scaling some of the bigger tumours with scaling  
50 factors ranging from 0.5 to 0.8, with the aim to increase the number of small tumours.  
51 As a result, the training dataset was composed of 645 uniform tumours, 645 tumours  
52 split in halves (of which 70 were augmented) and 645 hollow tumours (of which 177 were  
53 augmented). In this experiment, the network was trained for 500 epochs with batch size  
54 50. The test dataset was made of 422 non-augmented images. This experiment is  
55 henceforth referred to as experiment 1.  
56  
57

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     9  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12

13  
14     *Application to PET data generated with different PSFs*     The second aim of this work  
15 was to test if the proposed 3D CNN could generalise well to PET data generated with  
16 different PSFs and restore these images accurately. In this experiment the 3D CNN was  
17 trained on dataset 1 generated with one PSF, and subsequently applied to dataset 2,  
18 generated with a range of PSFs which were not learned during training. This experiment  
19 is henceforth referred to as experiment 2.  
20  
21

22     *Application to PET data generated with different noise levels*     As a proof of concept,  
23 the proposed 3D CNN was applied to two small datasets characterised by two different  
24 noise levels. These two datasets were generated starting from the same ten ground truth  
25 hollow tumours. These tumours were assigned with activity concentrations ranging  
26 from 6 kBq/ml to 35 kBq/ml, and they all had volumes larger than 5 ml. Two sets of  
27 PET images were generated from the same ground truth images, using the analytical  
28 simulator. This was done by setting the number of total prompts to 50 millions and  
29 to 200 millions respectively. Both datasets were generated using a spatially invariant  
30 PSF with FWHM (4.5, 4.5, 4.0) mm. The CNN, trained on PET images simulated with  
31 100 million total prompt counts, was applied to the two small datasets generated with  
32 different statistics. This experiment is henceforth referred to as experiment 3.  
33  
34

35     **Results**  
36  
37

38     In this section, the results obtained training and testing the 3D CNN on data generated  
39 with a single PSF are presented. Subsequently, the same network is applied to PET  
40 data generated with a range of PSFs and the results are qualitatively and quantitatively  
41 compared to those obtained in the first experiment.  
42  
43

44     *CNN training and testing using PET data generated with a single PSF*  
45  
46

47     The first experiment was performed using the simulated training and test datasets  
48 generated with a single PSF. Three representative volumes, each characterised by a  
49 different activity pattern, are presented in figure 3. The CNN yielded better defined  
50 tumour shapes and denoised images in all three cases. This visual assessment was  
51 followed by a quantitative analysis. In figure 4, 150 randomly selected MSSIM values  
52 are shown. The average MSSIM value measured using the ground truth images and the  
53 reconstructed PET images is  $0.33 \pm 0.06$ . An improved average MSSIM value equal to  
54  $0.47 \pm 0.06$  is measured using the CNN predictions and the ground truth images. The  
55 predicted images are overall characterised by higher MSSIM values. To further assess the  
56 performance of the network, we then measured the maximum, median and peak recovery  
57 coefficients (RCs). The median value was only estimated for the tumours with uniform  
58 uptake and for the hollow tumours, which also had uniform uptake. In figure 5 the RCs  
59 are plotted against the tumour volume expressed in ml. The maximum, median and  
60 peak RC were not well recovered for tumours that have a volume smaller than around

1  
2 *Improved 3D Tumour Definition and Quantification of Uptake* 10  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

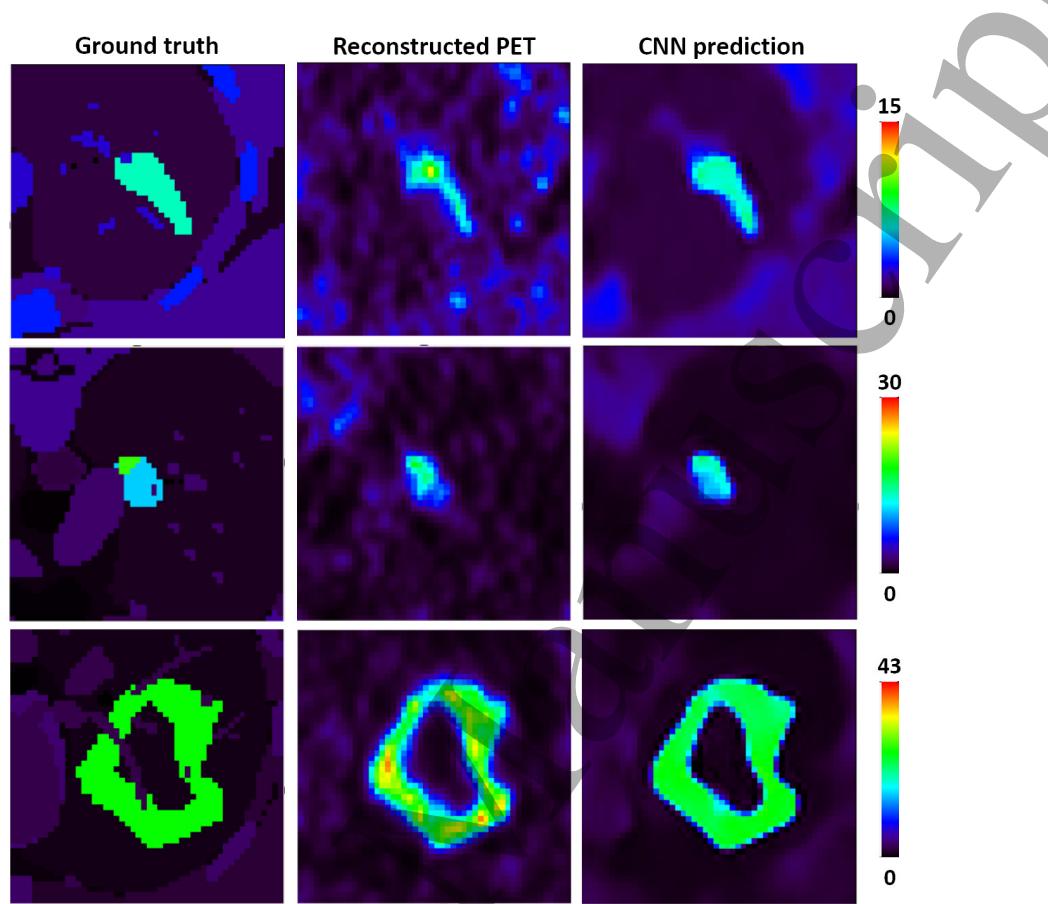


Figure 3: Transverse views of three representative volumes, belonging to experiment 1, where the reconstructed PET images were generated with one PSF. Each column shows the ground truth images, the reconstructed PET images and the CNN's predicted images respectively. In each row the images are shown with the same colour scale expressed in kBq/ml.

5 ml, so we performed a detailed investigation of the small tumours. Out of all the test images, the network did not recover any increased uptake for 5 tumours, which had a volume smaller than 0.18 ml (20 voxels). In these predicted images no tumour could be detected. The tumour activities were not accurately recovered for tumours with volumes between 0.18 and 1.33 ml that were located close to other structures (ribs, heart) and that were characterised by a low ground truth activity. To better describe the RC curves we present two separate measurements, one describing the recovery of bigger volumes and the second measured for smaller volumes. The first metric we used is the average RC value, measured only for tumours with volume exceeding 5 ml. These average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values measured on the reconstructed PET images and on the CNN predictions are summarised in table 1. All the recovery coefficients measured for the predicted images are approaching 1, meaning that the CNN yields

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     11  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

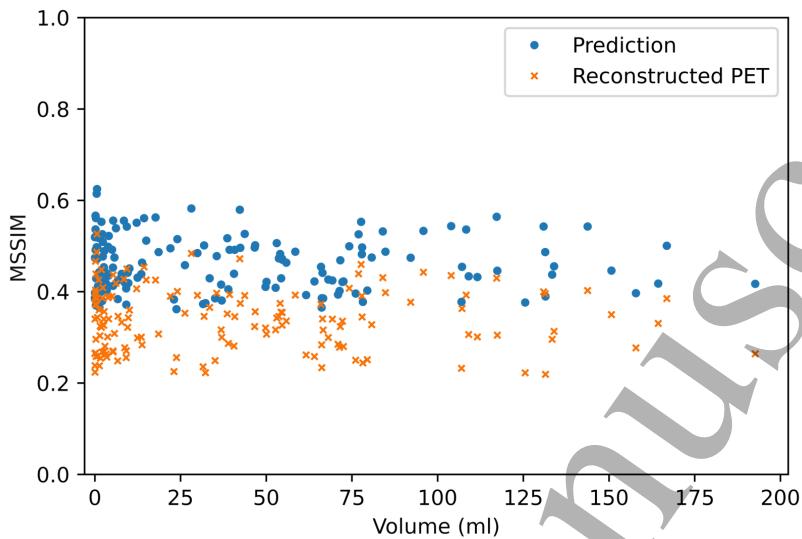


Figure 4: This figure shows 150 representative MSSIM values, obtained in experiment 1, where the reconstructed PET images were generated with one PSF. The MSSIMs between reconstructed PET images and ground truth images are shown in orange, the MSSIMs between predictions and ground truth are plotted in blue.

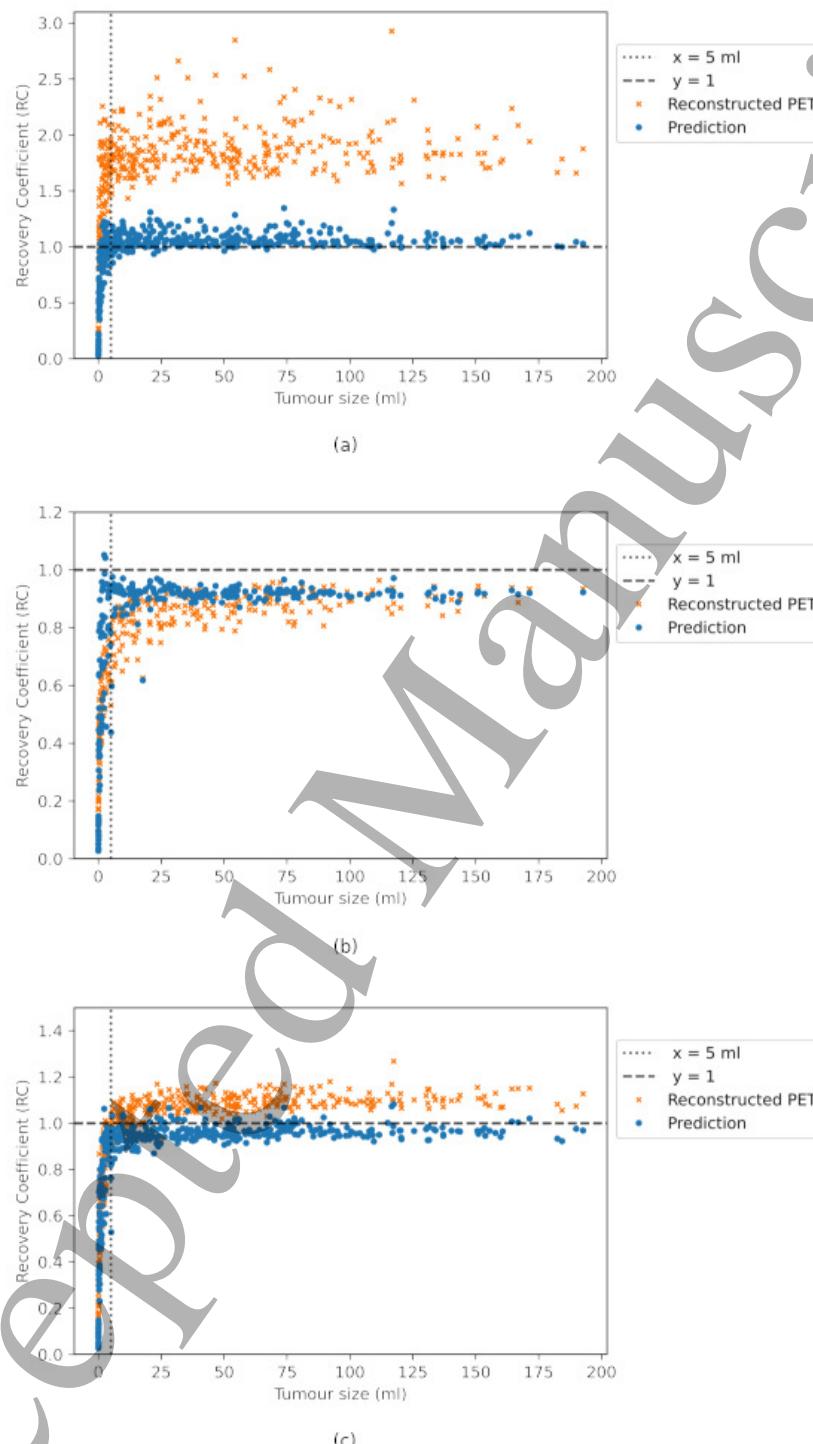
improved estimates of the maximum and median activity within the tumours. In order to quantitatively assess the performance of the CNN on tumours split into halves, we measured the ratio between the median activity values measured in each tumour half and compared the results. The CNN underestimated the activity ratio between the two tumour halves in 85% of the test cases. When tumours smaller than 5 ml were excluded from the analysis, the ratio between the median activities measured in the two tumour halves was underestimated by the CNN in 90% of the cases.

Table 1: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values relative to experiment 1, where the reconstructed PET images were generated with one PSF. These values are measured only for tumours larger than 5 ml.

volume $\geq$ 5 ml	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET	$1.87 \pm 0.22$	$0.86 \pm 0.07$	$1.08 \pm 0.05$
CNN's predictions	$1.06 \pm 0.07$	$0.91 \pm 0.05$	$0.96 \pm 0.04$

To describe the smaller volumes we performed an average RC measurement using tumours with volume spanning 1 to 2 ml. To calculate this value the chosen 1 ml interval is split in four smaller segments. The average RC is measured for each segment and subsequently an overall value is obtained by averaging these four RCs. The maximum,

1  
2 *Improved 3D Tumour Definition and Quantification of Uptake* 12  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



52 Figure 5: The  $RC_{max}$  and  $RC_{median}$  values, obtained training and testing the network  
53 on PET data generated with a single PSF, are plotted against the tumour volume in  
54 (a) and (b) respectively. The  $RC_{peak}$  values measured on the same dataset are shown  
55 in (c). The coefficients measured using the reconstructed PET images are shown in  
56 orange, whereas the ones measured using the predicted images are shown in blue.  
57 Tumours split in halves, each assigned with a different activity, were excluded from  
58 the calculation of the median values.  
59  
60

1  
2      *Improved 3D Tumour Definition and Quantification of Uptake*      13  
3  
4

5 median and peak average RCs are measured both using the reconstructed PET images  
6 and the CNN predictions, and the results are summarised in table 2.  
7

8 Table 2: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values relative to experiment 1, where  
9 the reconstructed PET images were generated with one PSF. These values are measured  
10 only for tumour volumes between 1 and 2 ml.  
11

$1 \text{ ml} \geq \text{volume} \geq 2 \text{ ml}$	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET	$1.46 \pm 0.17$	$0.55 \pm 0.06$	$0.68 \pm 0.09$
CNN's predictions	$0.81 \pm 0.11$	$0.67 \pm 0.13$	$0.73 \pm 0.08$

20 Although the RC values predicted by the CNN for small tumours are lower than  
21 1, the CNN yields improved estimates for all RC. We further investigated which factors  
22 affected the recovery of radiotracer activity in the simulated tumours. As the simulated  
23 tumours may be characterised by elongated shapes, one parameter that we chose for the  
24 analysis of small tumours was the sphericity, a dimensionless metric defined as:  
25

$$26 \quad sphericity = \frac{\sqrt[3]{36\pi \cdot V^2}}{A} \quad (7)$$

27 where  $V$  is the tumour volume and  $A$  the tumour surface area. The sphericity  
28 describes the roundness of the tumour shape relative to a sphere and its values range  
29 from 0 to 1. A sphericity value of 1 indicates that the tumour has a perfect spherical  
30 shape, a low sphericity value means that the tumour shape is more elongated. In figure  
31 6 the RC values are colour-coded based on the sphericity of the corresponding tumour  
32 volumes. In these plots, spherical tumours are shown in yellow whereas more elongated  
33 tumours are shown in blue. The plots show that higher recovery coefficients are often  
34 associated with a higher sphericity. This may be due to the fact that elongated tumours  
35 are more affected by partial volume effect as a larger fraction of tumour voxels is located  
36 closer to the tumour edges and is thus more prone to spilling-in and spilling-out [3]. For  
37 this reason, it may be more difficult for the CNN to recover the ground truth tumour  
38 shape and uptake.  
39

40      *Application to PET data generated with different PSFs*  
41

42 The network previously trained in experiment 1, where the reconstructed PET images  
43 were generated with one PSF, was then applied to PET data generated with different  
44 PSFs. The MSSIM values measured using the reconstructed PET images and the images  
45 predicted by the CNN are shown in figure 7. The average MSSIM measure between the  
46 reconstructed PET images and the ground truth images is  $0.34 \pm 0.07$  whereas the average  
47 MSSIM between the CNN predictions and the ground truth images is  $0.43 \pm 0.07$ . This  
48 value is comparable to the MSSIM obtained in the previous experiment. In figure 8,  
49 three volumes belonging to the test dataset are presented. The PSFs used to simulate the  
50 PET raw data and to reconstruct the PET images had FWHM (4.7, 4.7, 4.2) mm for the  
51

1  
2 *Improved 3D Tumour Definition and Quantification of Uptake* 14  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

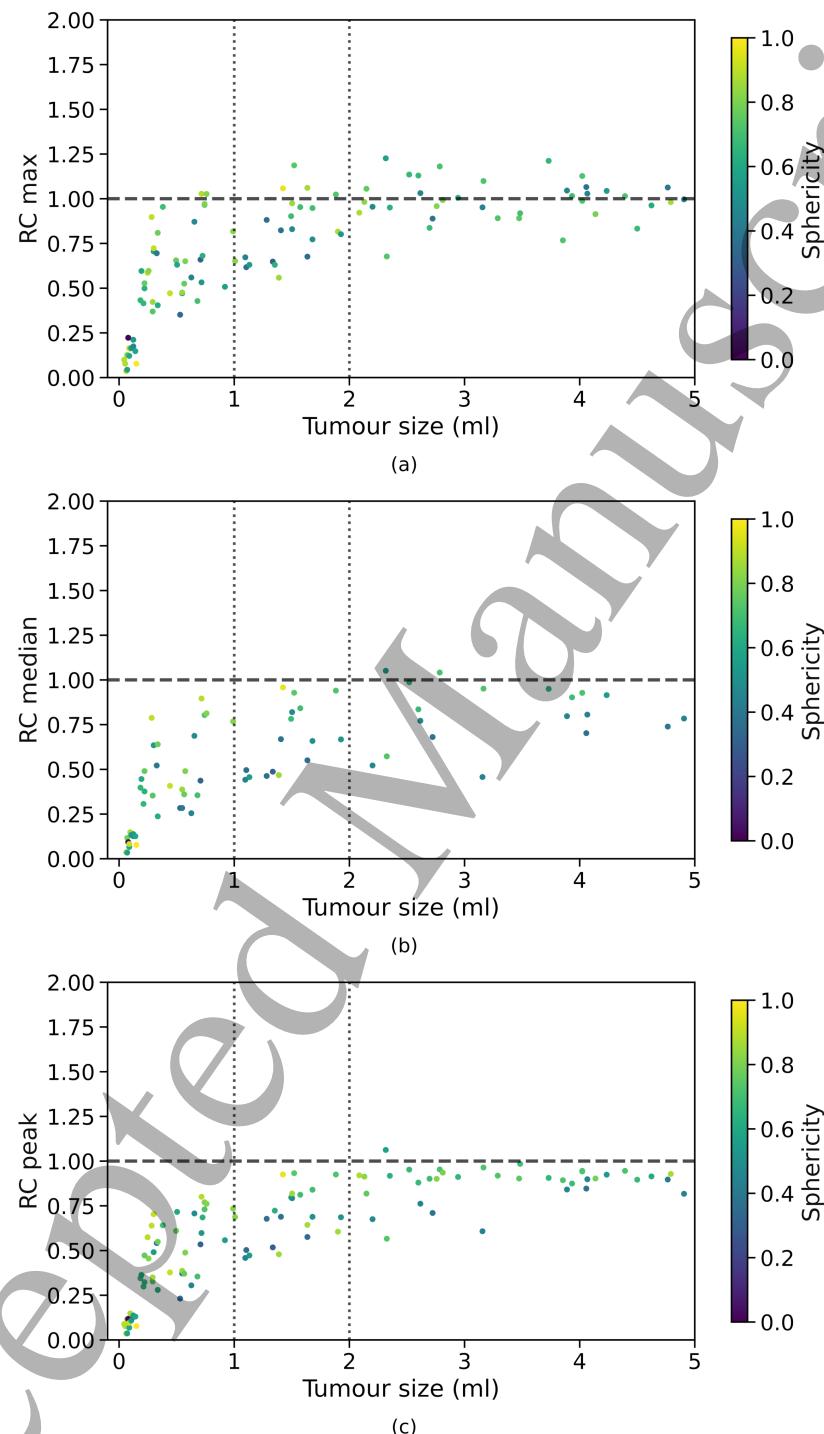


Figure 6: Analysis of tumour volumes smaller than 5 ml predicted by the CNN. The  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values, obtained training and testing the network on PET data generated with a single PSF, are plotted against the tumour volume in (a), (b) and (c) respectively. The tumour sphericity is colour coded using the diverging colour scale shown on the right.

1  
2      *Improved 3D Tumour Definition and Quantification of Uptake*      15  
3  
4

5      Table 3: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values relative to PET data generated  
6      with different PSFs. These values are averaged over tumours with volume larger than  
7      5 ml.

volume $\geq$ 5 ml	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET	$1.94 \pm 0.26$	$0.84 \pm 0.08$	$1.07 \pm 0.08$
CNN's predictions	$1.07 \pm 0.08$	$0.90 \pm 0.05$	$0.95 \pm 0.05$

16      images in the first row, FWHM (4.0, 4.0, 3.6) mm for the images in the second row and  
17      FWHM (4.5, 4.5, 4.0) mm in the third row. Even though the PSFs used in the first and  
18      second row did not match the PSF used to generate and reconstruct the PET images  
19      used for training, the network was still able to predict improved tumour shapes and  
20      activities. In figure 9 the  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  are plotted against the tumour  
21      volume expressed in ml. The recovery curves show similar behaviours as in the previous  
22      experiment and improved values are obtained when estimating the recovery coefficients  
23      on the images predicted by the CNNs. The average RC measurements calculated for  
24      tumour volumes larger than 5 ml are presented in table 3. The CNN yields improved  
25      recovery coefficients. The RC estimates shown in this table are comparable to the  
26      ones obtained in the previous experiment, showing that the network can successfully  
27      recover the maximum, median and peak activity in the tumours even when tested on  
28      images generated with different PSF values. A separate analysis was performed for  
29      tumour volumes smaller than 5 ml. Average RCs were measured for tumour volumes  
30      between 1 ml and 2 ml as in the previous experiment, and the results are presented  
31      in table 4. Although the recovery of small tumours was less accurate, the CNN always  
32      yielded improved RC estimates. Again, we measured the sphericity of the small tumours  
33      comprised between 1 and 2 ml and figure 10 shows the  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$   
34      measurements which are colour coded based on the tumour sphericity. Looking at the  
35      plots we can observe that in this dataset a lower recovery coefficient is often associated  
36      with lower sphericity, similarly to what we found in the previous experiment.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 4: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values obtained testing the network  
on the reconstructed PET images generated with a range of PSFs. These values are  
measured only for tumour volumes between 1 and 2 ml.

1 ml $\geq$ volume $\geq$ 2 ml	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET	$2.03 \pm 0.21$	$0.59 \pm 0.05$	$0.87 \pm 0.07$
CNN's predictions	$1.15 \pm 0.12$	$0.84 \pm 0.13$	$0.89 \pm 0.06$

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

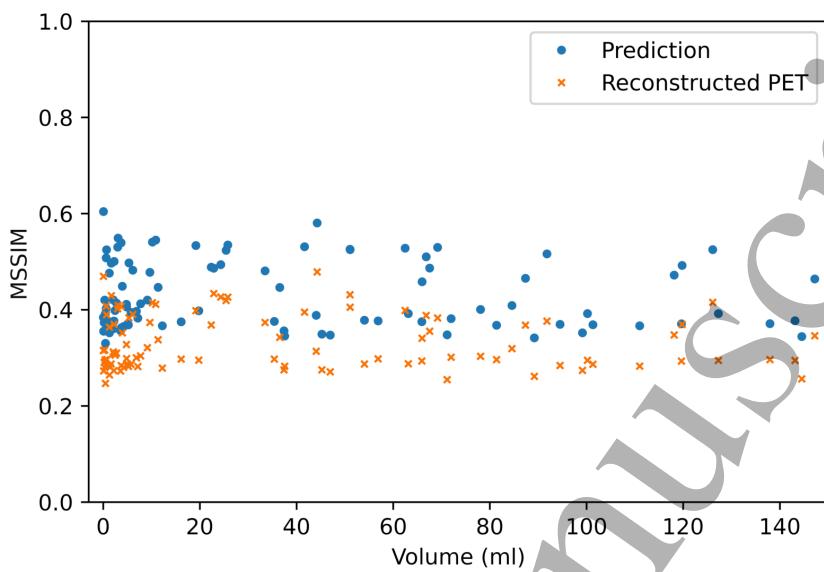


Figure 7: This figure shows all the MSSIM values, measured in the test dataset generated with multiple PSFs. The MSSIMs between reconstructed PET images and ground truth images are shown in orange, the MSSIMs between predictions and ground truth are plotted in blue.

#### *Application to PET data generated with different noise levels*

As a proof of concept, we performed an experiment using simulated PET images generated with different number of counts. In this case, a small dataset composed of 10 hollow tumours was used as ground truth. The two datasets were generated by setting the number of total prompts to 50 millions and to 200 millions respectively. The CNN that was trained in experiment 1, using PET images generated with 100 million total prompt events, was then applied to the two datasets.

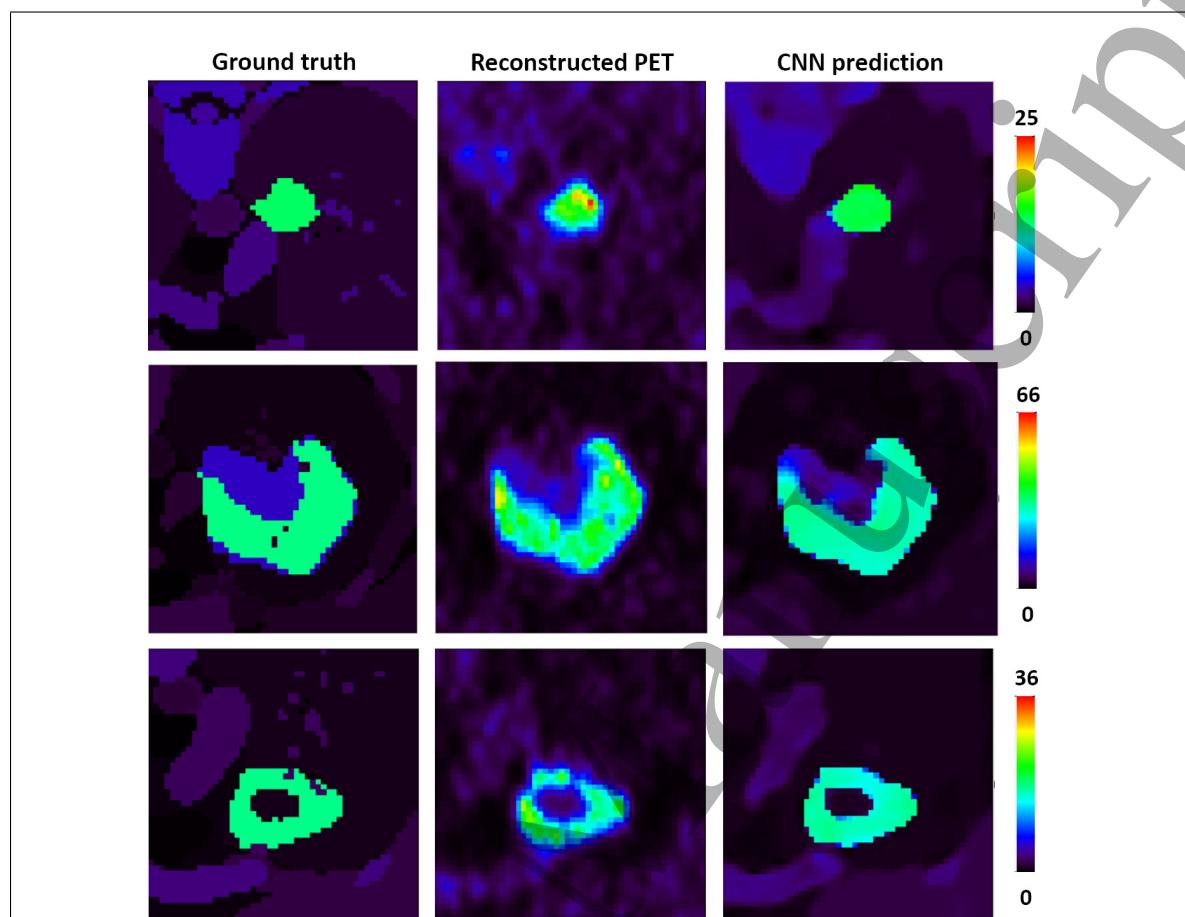
A representative example of a hollow tumour simulated with different number of counts is presented in figure 11. When the CNN was applied to the reconstructed PET images generated with high statistics and with low statistics, it yielded a denoised image with a better defined tumour shape and a more uniform tumour activity distribution.

Table 5: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values obtained testing the network on the reconstructed PET images generated with 50 million counts.

$V \geq 5 \text{ ml}$	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET	$2.28 \pm 0.32$	$0.79 \pm 0.05$	$1.08 \pm 0.09$
CNN's predictions	$1.16 \pm 0.07$	$0.93 \pm 0.04$	$1.00 \pm 0.07$

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*  
3  
4

17



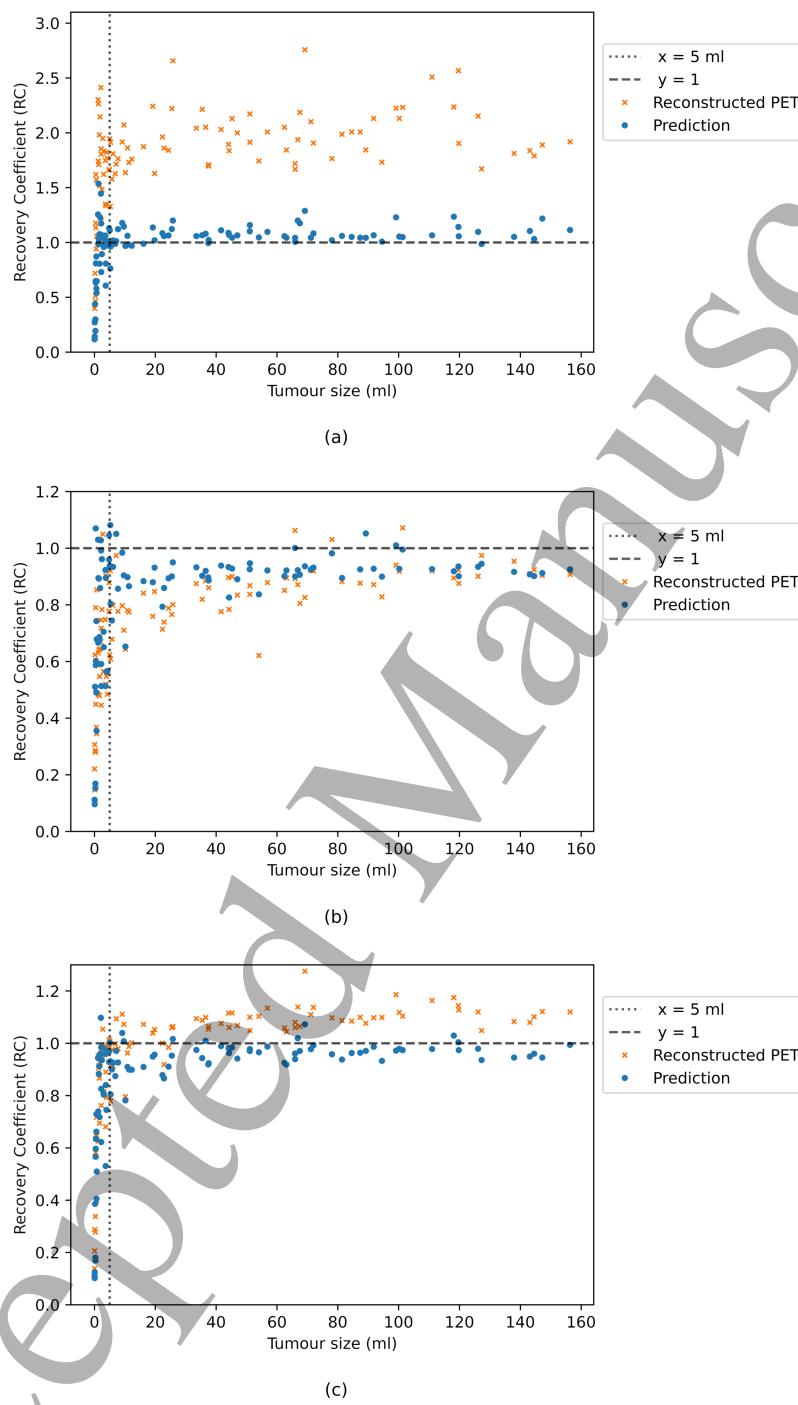
34  
35     Figure 8: Transverse views of three representative volumes, generated each with a  
36     different PSF. The PET images were generated with a PSF with FWHM (4.7, 4.7,  
37     4.2) mm in the first row, a PSF with FWHM (4.0, 4.0, 3.6) mm in the second row  
38     and PSF with FWHM (4.5, 4.5, 4.0) mm in the third row. Each column shows  
39     the ground truth images, the reconstructed PET images and the CNN's predicted  
40     images respectively. In each row the images are shown with the same colour scale,  
41     expressed in kBq/ml.  
42

43  
44  
45     Table 6: Average  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values obtained testing the network on  
46     the reconstructed PET images generated with 200 million counts.  
47

	$V \geq 5 \text{ ml}$	$RC_{max}$	$RC_{median}$	$RC_{peak}$
Reconstructed PET		$1.73 \pm 0.14$	$0.82 \pm 0.06$	$1.04 \pm 0.08$
CNN's predictions		$1.06 \pm 0.05$	$0.91 \pm 0.04$	$0.96 \pm 0.04$

56  
57     To quantitatively assess our results, the RCs were measured using the reconstructed  
58     PET images and the CNN predictions. As shown in table 5, when applied to the  
59     PET images generated with low statistics, the CNN yielded improved RCs. The RCs  
60

1  
2 *Improved 3D Tumour Definition and Quantification of Uptake* 18  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



51 Figure 9: The  $RC_{max}$  and  $RC_{median}$  values, obtained training and testing the network  
52 on PET data generated with a range of PSFs, are plotted against the tumour volume  
53 in (a) and (b) respectively. The  $RC_{peak}$  values measured on the same dataset are  
54 plotted in (c). The coefficients measured using the reconstructed PET images are  
55 shown in orange, whereas the ones measured using the predicted images are shown in  
56 blue. Tumours split in halves, each assigned with a different activity, were excluded  
57 from the calculation of the median values.  
58  
59  
60

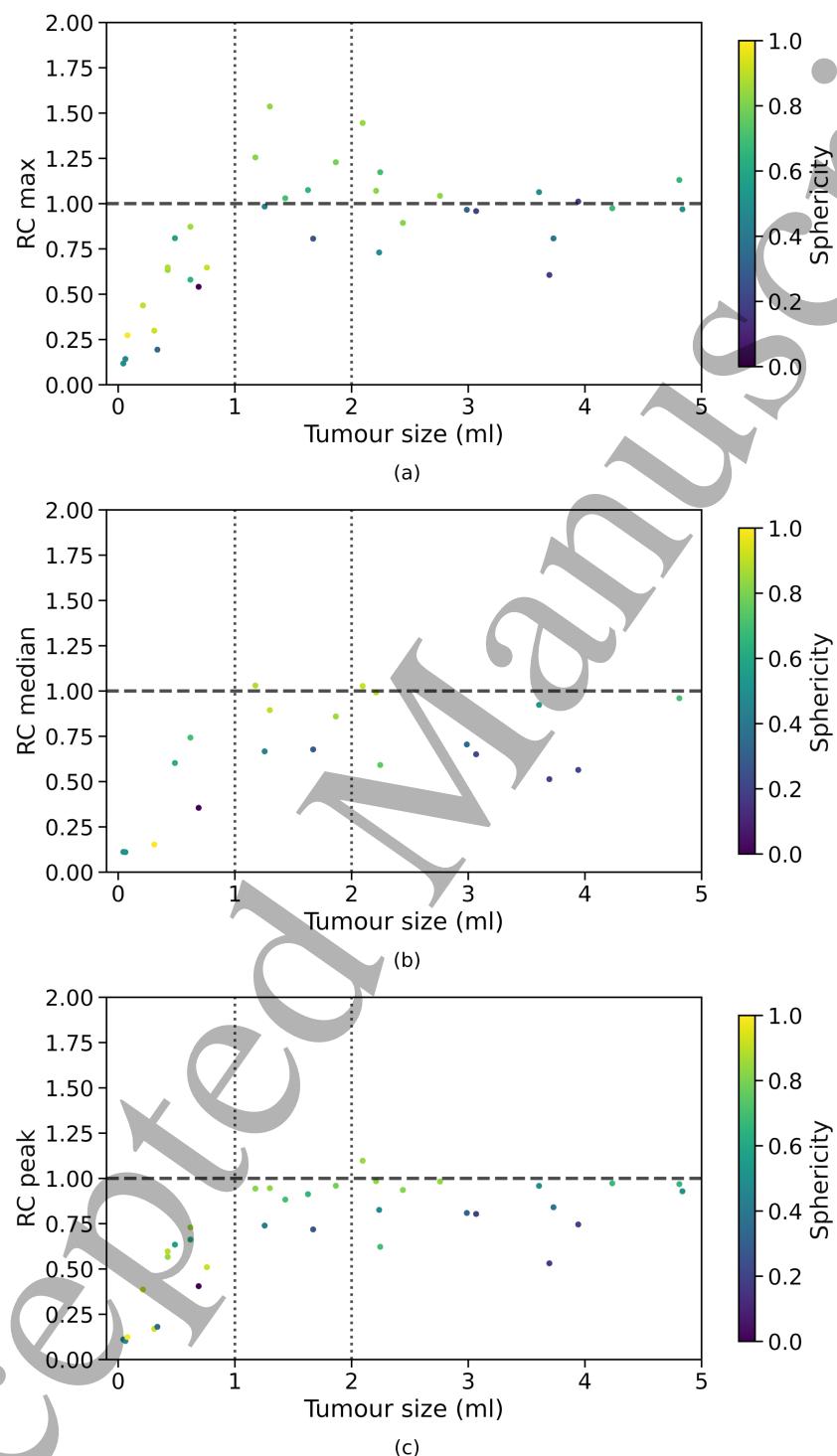
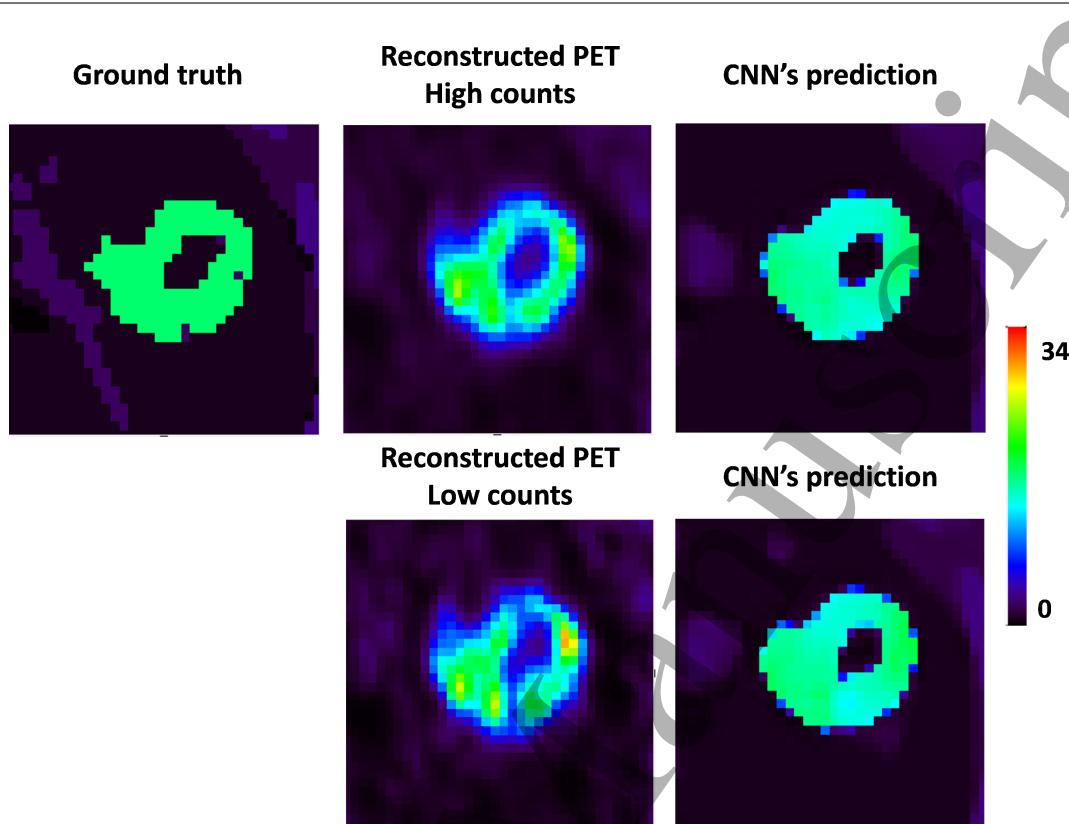
1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     19  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 10: Analysis of tumour volumes smaller than 5 ml predicted by the CNN. The  $RC_{max}$ ,  $RC_{median}$  and  $RC_{peak}$  values, obtained testing the network on PET data generated with a range of PSFs, are plotted against the tumour volume in (a), (b) and (c) respectively. The tumour sphericity is colour coded using the diverging colour scale shown on the right.

1  
2 *Improved 3D Tumour Definition and Quantification of Uptake* 20  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30



31 Figure 11: Transverse views of a hollow tumour. The reconstructed PET images in  
32 the top row were generated by setting the number of prompt counts to 200 million,  
33 the ones in the bottom row were generated using 50 million prompt counts. The  
34 CNN predictions are shown in the third column. All the images are shown with the  
35 same colour scale, expressed in kBq/ml.

36  
37  
38  
39  
40 measured using the reconstructed PET images generated with a high number of counts  
41 are presented in table 6. Again, the CNN was able to recover improved RCs. In both  
42 cases, the CNN's performance is comparable to the one obtained in experiment 1, when  
43 the CNN was trained and tested on data generated with the same noise level.  
44  
45

#### 46 Discussion and Conclusion

47  
48 In this paper we propose a deep learning approach to improve quantification of  
49 radiotracer uptake and tumour shape definition in PET images. A 3D CNN was  
50 successfully trained and tested on simulated data generated with a single PSF and  
51 applied to reconstructed PET images generated with a range of PSFs. The results  
52 indicate that the network is able to improve the definition of the tumour shapes and to  
53 denoise reconstructed PET images. A quantitative analysis of the results obtained using  
54 simulated data has shown that the images predicted by the 3D CNN yield improved  
55  
56  
57  
58  
59  
60

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     21  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

estimates of the maximum tumour activities. We observed that the peak, maximum and median activities were not accurately recovered for tumours with volumes smaller than 5 ml, so a more detailed analysis was performed on small tumours. Only tumour volumes smaller than 0.18 ml presented critical issues. Bigger volumes located close to other background structures and characterised by a low ground truth activity were generally associated with an inaccurate prediction of the peak, maximum and median activity. We measured the average RCs for tumours with volumes between 1 ml and 2 ml and the CNN yielded improved estimates for all RCs. We additionally measured the sphericity for each of these tumours and we observed that lower RCs were often related to a lower tumour sphericity. In future work, we plan to train the network on a bigger dataset and to further augment the training dataset, thus adding more small volumes to the training dataset. This might improve the performance of the network for this class of tumours, as the network would be able to learn from more small tumours during training. A secondary effect that we noticed in our experiments was an improvement in the recovery of background structures. This effect will be further investigated in future work. Our approach proved successful when the network was applied to a set of reconstructed PET images generated with a range of PSFs that did not match the PSF used to generate the training dataset. Preliminary results suggest that the proposed approach would be able to restore PET images acquired with different scanners and spatially varying PSF. Finally, a proof of concept experiment was performed applying the CNN on reconstructed PET images generated with two different noise levels. In both cases, the CNN's performance is comparable to the one obtained when the CNN was trained and tested on data generated with the same noise levels. This work has the potential to be extended to larger areas of the body, in order to improve the estimation of the total tumour burden. The proposed approach has been tested on images generated with two specific noise levels, the robustness of this method to data generated with other noise levels remains to be evaluated. Before this algorithm can be applied widely, it will be necessary to evaluate its performance on PET images reconstructed with a wider range of number of OSEM iterations and imaging situations in order to determine the range of acquisition, reconstruction and imaging conditions under which it remains valid. In this work the same spatially invariant PSF was used for the simulation of the PET raw data and for the reconstruction of PET images. On the contrary, in a clinical setting PET images would be characterised by spatially variant and non stationary PSFs and there may be a mismatch between the system's PSF and the one modelled in the reconstruction algorithm. These aspects will be investigated in future work. Further experimentation is also needed to assess the robustness of the proposed deep learning approach to data generated using PSFs characterised by a different anisotropy and to more complex tumour shapes associated with heterogeneous activities. In future work, we plan to extend the proposed approach so that it can be applied to a wider region of the torso.

1  
2     *Improved 3D Tumour Definition and Quantification of Uptake*     22  
3  
4  
5

6     **Acknowledgments**

7     This project has received funding from the European Union's Horizon 2020 research  
8     and innovation programme under the Marie Skłodowska-Curie grant agreement No.  
9     764458. This research was supported by the Wellcome/EPSRC Centre for Medical  
10    Engineering [WT 203148/Z/16/Z], and by the National Institute for Health Research  
11    (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation  
12    Trust and King's College London. The views expressed are those of the author(s) and not  
13    necessarily those of the NHS, the NIHR or the Department of Health. Cancer Research  
14    UK National Cancer Imaging Translational Accelerator Award (C4278/A27066).

15  
16     **ORCID iDs**

17  
18     Zacharias Chalampakis <https://orcid.org/0000-0003-3390-6979>

19     Claude Comtat <https://orcid.org/0000-0003-4691-7001>

20     Irène Buvat <https://orcid.org/0000-0002-7053-6471>

21     Julia A. Schnabel <https://orcid.org/0000-0001-6107-3009>

22     Paul K. Marsden <https://orcid.org/0000-0001-9892-9640>

23  
24     **References**

- 25  
26  
27  
28  
29  
30  
31  
32  
33     [1] Boellaard R, Krak N C, Hoekstra O S and Lammertsma A A 2004 *Journal of Nuclear Medicine*  
34       **45** 1519–1527
- 35     [2] Zaidi H and Karakatsanis N 2017 Nuclear medicine: physics special feature review article towards  
36       enhanced PET quantification in clinical oncology *Tech. rep.*
- 37     [3] Soret M, Bacharach S and Buvat I 2007 *Journal of Nuclear Medicine* **48** 932–945 ISSN 0161-5505
- 38     [4] Cysouw M, Kramer G, Schoonmade L, Boellaard R, De Vet H and Hoekstra O 2017 *European*  
39       *Journal of Nuclear Medicine and Molecular Imaging* **44** 2105–2116
- 40     [5] Vanderhoek M, Perlman S B and Jeraj R 2012 *Journal of Nuclear Medicine* **53** 4–11 ISSN 0161-  
41       5505
- 42     [6] Gong K, Berg E, Cherry S R and Qi J 2020 *Proceedings of the IEEE* **108** 51–68
- 43     [7] Kim K, Wu D, Gong K, Dutta J, Kim J H, Son Y D, Kim H K, El Fakhri G and Li Q 2018 *IEEE*  
44       *Transactions on Medical Imaging* **37** 1478–1487
- 45     [8] Litjens G, Kooi T, Bejnordi B E, Setio A A A, Ciompi F, Ghafoorian M, van der Laak J A, van  
46       Ginneken B and Sánchez C I 2017 *Medical Image Analysis* **42** 60–88 ISSN 1361-8415
- 47     [9] Shiyam Sundar L K, Muzik O, Buvat I, Bidaut L and Beyer T 2021 *Methods* **188** 4–19 ISSN  
48       1046-2023 artificial Intelligence Approaches for Imaging Biomarkers
- 49     [10] Gong K, Guan J, Liu C C and Qi J 2019 *IEEE Transactions on Radiation and Plasma Medical*  
50       *Sciences* **3** 153–161
- 51     [11] Wang Y, Yu B, Wang L, Zu C, Lalush D, Lin W, Wu X, Zhou J, Shen D and Zhou L 2018  
52       *NeuroImage* **174** 550–562 ISSN 1053-8119
- 53     [12] Dal Toso L, Pfaehler E, Boellaard R, Schnabel J A and Marsden P K 2019 Deep learning  
54       based approach to quantification of PET tracer uptake in small tumors *Machine Learning for*  
55       *Medical Image Reconstruction* ed Knoll F, Maier A, Rueckert D and Ye J C (Cham: Springer  
56       International Publishing) pp 181–192 ISBN 978-3-030-33843-5
- 57  
58  
59  
60

1  
2      *Improved 3D Tumour Definition and Quantification of Uptake*      23  
3  
4

- [13] Berthon B, Häggström I, Apte A, Beattie B J, Kirov A S, Humm J L, Marshall C, Spezi E, Larsson A and Schmidlein C R 2015 *Physica Medica* **31** 969–980 ISSN 1120-1797
- [14] Pfaehler E, Jong J, Dierckx R, van Velden F and Boellaard R 2018 *EJNMMI Physics* **5**
- [15] Bekisz S and Geris L 2020 *Journal of Computational Science* 101198 ISSN 1877-7503
- [16] Stute S, Tauber C, Leroy C, Bottlaender M, Brulon V and Comtat C 2015 Analytical simulations of dynamic pet scans with realistic count rates properties pp 1–3
- [17] Segars W, Sturgeon G, Mendonca S, Grimes J and Tsui B 2010 *Medical physics* **37** 4902–15
- [18] Merlin T, Stute S, Benoit D, Bert J, Carlier T, Comtat C, Filipovic M, Lamare F and Visvikis D 2018 *Physics in Medicine and Biology* **63**
- [19] Pfaehler E, Beukinga R, Jong J, Slump C, Dierckx R and Boellaard R 2018 *Medical Physics* **46** 665–678
- [20] Lasnon C, Hicks R, Beauregard J M, Milner A, Guizard A V, Bardet S, Gervais R, Lemoel G, Zalcman G and Aide N 2012 *Clinical nuclear medicine* **37** 971–976
- [21] Yamashita R, Nishio M, Do R and Togashi K 2018 *Insights Imaging* **9** 611–629
- [22] Kingma D and Ba J 2014 *International Conference on Learning Representations*
- [23] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I J, Harp A, Irving G, Isard M, Jia Y, Józefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D G, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P A, Vanhoucke V, Vasudevan V, Viégas F B, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y and Zheng X 2016 *CoRR* **abs/1603.04467**
- [24] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E 2011 *Journal of Machine Learning Research* **12** 2825–2830
- [25] Loening A and Sam Gambhir S 2003 *Molecular Imaging* **2** 131–137

Accepted