European Journal of Nuclear Medicine and Molecular Imaging
https://doi.org/10.1007/s00259-019-04606-y

ORIGINAL ARTICLE

Deep neural network for automatic characterization of lesions on ^{68}Ga-PSMA-11 PET/CT

Yu Zhao[1] · Andrei Gafita[2] · Bernd Vollnberg[3] · Giles Tetteh[1] · Fabian Haupt[3] · Ali Afshar-Oromieh[3] · Bjoern Menze[1] · Matthias Eiber[2] · Axel Rominger[3,4] · Kuangyu Shi[1,3]

Received: 21 July 2019 / Accepted: 7 November 2019

Abstract

Purpose This study proposes an automated prostate cancer (PC) lesion characterization method based on the deep neural network to determine tumor burden on ^{68}Ga-PSMA-11 PET/CT to potentially facilitate the optimization of PSMA-directed radionuclide therapy.

Methods We collected ^{68}Ga-PSMA-11 PET/CT images from 193 patients with metastatic PC at three medical centers. For proof-of-concept, we focused on the detection of pelvis bone and lymph node lesions. A deep neural network (triple-combining 2.5D U-Net) was developed for the automated characterization of these lesions. The proposed method simultaneously extracts features from axial, coronal, and sagittal planes, which mimics the workflow of physicians and reduces computational and memory requirements.

Results Among all the labeled lesions, the network achieved 99% precision, 99% recall, and an F1 score of 99% on bone lesion detection and 94%, precision 89% recall, and an F1 score of 92% on lymph node lesion detection. The segmentation accuracy is lower than the detection. The performance of the network was correlated with the amount of training data.

Conclusion We developed a deep neural network to characterize automatically the PC lesions on ^{68}Ga-PSMA-11 PET/CT. The preliminary test within the pelvic area confirms the potential of deep learning methods. Increasing the amount of training data should further enhance the performance of the proposed method and may ultimately allow whole-body assessments.

Keywords Prostate cancer · PSMA · Deep learning · Lesion detection · PET/CT

Introduction

Prostate cancer (PC) is one of the most common cancers worldwide [1] and has become the third most frequent cause of cancer-related mortality among men in developed countries [2]. By introducing serum PSA-levels as a screening tool, PC is usually diagnosed in the early stage, with a 5-year survival rate of almost 100% for local disease. If not detected and treated at stage 1, when the cancer is confined to the prostate gland, malignant tumor cells may spread to other regions by invading the hematic and lymphatic systems, whereupon the 5-year survival rate declines to 29% for metastatic PC [3, 4]. Chemotherapy imparts some increase in survival in patients with treat advanced metastatic PC, but current treatments are not curative [5, 6].

Prostate specific-membrane antigen (PSMA) is a type II transmembrane glycoprotein, which is constitutively expressed by normal prostate cells and significantly upregulated in prostate cancer cells [2]. Internalization of PSMA can

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence)

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-019-04606-y) contains supplementary material, which is available to authorized users.

✉ Kuangyu Shi
kuangyu.shi@dbmr.unibe.ch

[1] Department of Informatics, Technische Universität München, Munich, Germany

[2] Department of Nuclear Medicine, Technische Universität München, Munich, Germany

[3] Department of Nuclear Medicine, University of Bern, Freiburgstrasse 18, 3010 Bern, Switzerland

[4] Department of Nuclear Medicine, Ludwig Maximilian University of Munich, Munich, Germany

Published online: 07 December 2019

concentrate bound ligands within the cancer cell, thus presenting a mechanism for targeted radiotherapy. Indeed, PSMA has emerged as a major target for the theranostic approach [7]. In various studies, diagnostic positron emission tomography (PET) imaging with the PSMA ligand ^{68}Ga-PSMA-11 is followed by treatment with ^{131}I, ^{177}Lu, ^{213}Bi, and ^{225}Ac labeled PSMA-ligands for therapy in PC [8–12]. The efficacy of ^{177}Lu-PSMA-617 has been recently validated in a phase II clinical trial [13].

Despite the encouraging early results for PSMA-targeted radioligand therapy (RLT), treatment planning of this novel therapy is very challenging compared with planning for conventional external beam radiotherapy, due to abundance and systemic spread of the lesions. Indeed, RLT proved to be suboptimal for 30% of a group treated PC patients [14]. Therefore, there is a need for improved treatment planning to optimize the RLT outcome. A critical step for treatment planning is to assess with some accuracy the tumor burden, which necessarily entails detection and segmentation of the lesions to diagnostic PET. Usually, patients who undergo PSMA-targeted RLT have a high number of metastases. Therefore, time-consuming manual segmentation methods are impractical in routine practice. A first approach towards a semiautomatic segmentation method was developed to characterize the tumor burden of bone metastases, namely the bone-PET-index (BPI), on ^{68}Ga-PSMA-11 PET/CT images. This procedure informed the planning of ^{223}Ra-dichloride therapy by segmentation of osseous lesions using an SUV-based threshold on the PET image, with masking of the skeleton based on the CT images [15]. However, this method does not generalize to other types of lesions such as lymph node metastases, where prior anatomical information is more difficult to obtain. Indeed, it is extremely challenging to segment a high number of PSMA-positive lesions of heterogeneous size and tracer uptake, with distribution in a variety of anatomical contexts with different background activity. Until now, there are no successful computer-aided methods to evaluate tumor load for the treatment planning of PSMA-targeted RLT.

Deep learning has demonstrated its superiority in exceeding the human capacity for processing certain kinds of data and has outperformed conventional machine learning methods in many applications such as image recognition and robotics [16]. Unlike conventional imaging processing methods, which usually rely on features specified a prior by experts, convolutional neural networks (CNNs) have the advantage of being able to learn automatically and effectively the salient feature representations [17–19]. Initial applications of deep classification networks for image segmentation made use of sliding patches [20], which leads to redundant computations and long inference times. The fully convolutional network (FCN) [21] was later developed in the context of speech recognition to realize semantic segmentation using fully convolutional layers, deconvolution layers, and skip architecture. FCN-like networks were then applied to medical image segmentation, the proposed U-Net [22] obtained performance in neuronal structure and cell segmentation that matched conventional methods. Increasing numbers of FCN-based methods have been proposed, and some have gained significant success in different medical image segmentation problems [23–26]. For example, the two cascaded FCN, i.e., W-Net, has been applied to PET/CT images with ^{68}Ga-Pentixafor for automatic detection and segmentation of multiple myeloma lesions in bone [27]. However, this approach is still restricted to the characterization of relatively unchallenging bone lesions.

In this pilot study, we developed a triple-combining 2.5D U-Net to automatically detect and segment local prostate tumors, bone lesions, and lymph node metastasis. For the proof-of-concept, this work focuses on the detection and segmentation of lesions in the pelvic area. We tested the deep learning method that was tested on a dataset of PSMA PET-CT scans collected from three different centers.

Material and methods

Patient and imaging

This pilot study included a dataset of 193 patients (mean age 69.6 ± 7.9 years, range 50–85 years) with metastatic castration-resistant prostate cancer (mCRPC), who were examined at three university medical centers, i.e., Technical University of Munich (TUM; 44 Patients), University of Munich (48 Patients), and University of Bern (101 Patients). The study was performed in accordance with the requirements of local research committees and ethical guidelines. A quality control of the consistency of the protocol at each site prior to the study revealed no sign of bias between centers. All patients underwent PSMA PET/CT imaging from the head to the thigh approximately 60 min after intravenous injection of ^{68}Ga-PSMA-11. A low-dose CT was obtained for attenuation correction. PET emission data were acquired using a 3D model, followed by decay and scatter correction, and was iteratively reconstructed with attenuation correction. The PET images were intensity normalized for injected activity and body weight to standard uptake values (SUVs). For proof of concept, we restricted the focus of this study to the pelvic region. Nuclear medicine physicians (one from TUM and one from the University of Bern) manually delineated all the discernible lesions within the PET/CT scans of the pelvic region, with refinement and concordance of results by both physicians. A total of 1003 bone lesions, 626 lymph node lesions, and 127 local lesions were annotated in the entire group of 193 patients.

Figure 1 illustrates the typical examples of the PET/CT images from patients with different types of PC lesions

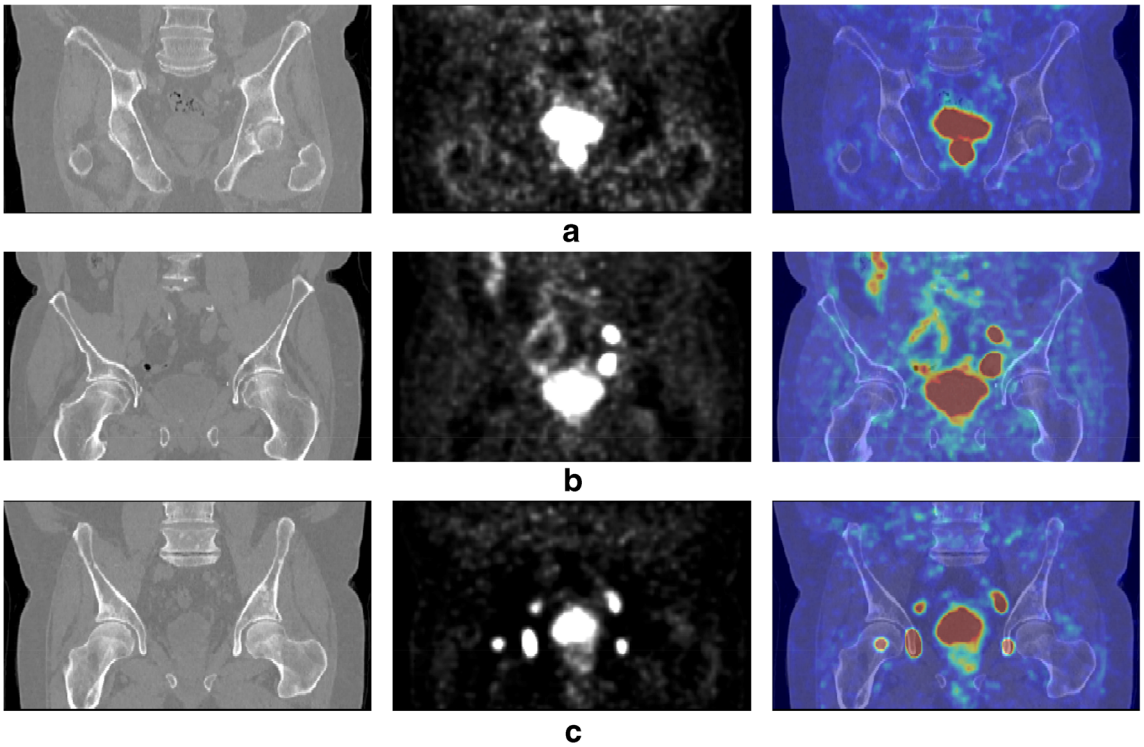

Fig. 1 Representative examples for recurrent PC patients with lesions in the residual prostate, bone, and lymph node (coronal plane). The left-hand column shows CT slices, the middle column shows PET slices, the right-hand column shows fused PET/CT scan. The top row **a** presents a patient with prostate and local tumour. The middle row **b** depicts a patient without local tumour, but numerous lymph node lesions. The bottom row **c** shows a patient with all three types of lesions

including local, bone metastasis, and lymph node metastasis. In the dataset from 193 patients, 89 patients have primary PC and locally recurrent tumor (Fig. 1a). Most patients underwent prostatectomy, which may have recurrence of lymph node or bone metastases despite curative surgical treatment of the primary disease (Fig. 1b); 38 patients (out of complete set) had all three types of lesions mentioned above (Fig. 1c). Given this heterogeneity, it is necessary to detect and evaluate a wide range of lesion types, which is a challenging task for several reasons. First, the low contrast of the local tumor against a background of healthy prostate tissue impairs the detection of the local tumor. Second, the unexpected occurrence of lymph node or bone lesions anywhere in the body with large heterogeneities in shape, size, and tracer uptake intensity make their detection more difficult. Finally, compared with the background, the lesions usually occupy a small fraction of the image volume, which leads to segmentation difficulties arising from this imbalance.

Deep learning method

We develop a triple-combining 2.5D U-Net (Convolutional Neural Network) to mimic the workflow of physicians for characterizing PC lesions. The 2.5D Convolutional Neural Network (2.5D CNN) is an alternative end-to-end learning approach to reduce the computational cost in the 3D image learning tasks [28]. This structure of 2.5D CNN is similar to the conventional 2D CNN. However, the 2.5D CNN takes three neighboring slices as the input to predict the result of the middle slices, which makes the network able to learn the information of the third dimension (context information between slices) from the channels. The employed triple-combining 2.5D U-Net in this work further enhances the ability to learn 3D context information. It works by extracting salient features in PET and CT scans simultaneously from the axial, coronal, and sagittal planes simultaneously and then combining the obtained information to detect and segment lesions automatically. As shown in Fig. 2, the proposed framework consists of an information-extraction component and a fusion component. The information-extraction component includes three 2.5D U-Net [22, 29] networks, which are designed for extracting features from three different planes, respectively. The fusion component then synthesizes all information to predict the final characterization result by majority voting strategy. The architecture of each U-Net employed in this study is also illustrated in Fig. 2. It comprises a down-sampling path including three repeated encoder stacks and an up-sampling path including three repeated decoder stacks. The down-sampling path aggregates increasingly abstract information and the up-sampling path then recombines this information with shallower features to localize precisely the structures of interest. The skip-connection (concatenation

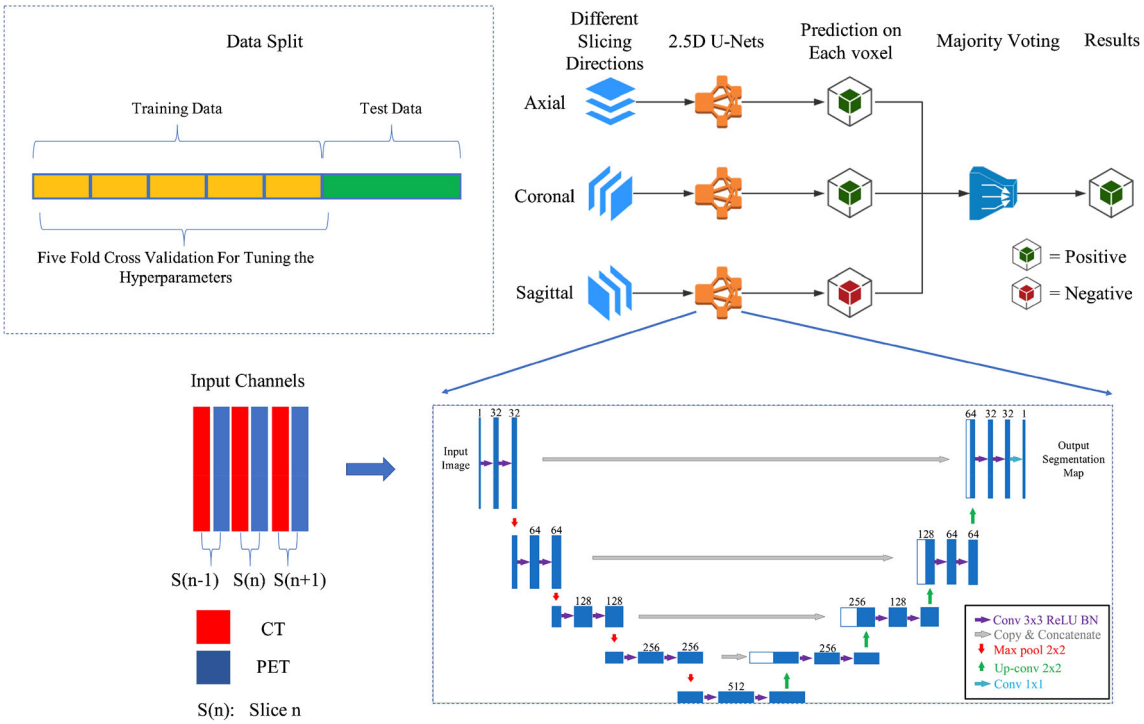

Fig. 2 The architecture of the proposed framework, which consists of an information-extraction component and a fusion component. The information-extraction component includes three 2.5D U-Nets, which are designed for extracting features from three different plans, respectively. The fusion component then synthesizes all information to predict the final characterization result by a majority voting strategy. This figure also presents the U-Net architecture. Different arrows denote specific operations in the architecture. The multi-channel feature maps are shown in blue and the copied feature maps are shown in write. The digit above the feature maps denotes the number of channels.

operation) between each decoder-encoder pair brings the features with higher spatial resolution from shallow layers of the encoder part directly to the layers of the decoder part for detection and segmentation. In the U-Net architecture, the number of down-sampling stages is an important factor influencing the performance. On the one hand, if the network has too many down-sampling layers, the information about small objects may disappear after passing one of the deeper down-sampling layers. On the other hand, an appropriate number of down-sampling stages correspond to a large valid receptive field, which helps the network capture more contextual information. In this work, we utilize four down-sampling processes to find a balance between the spatial resolution and feature representations.

In each encoder stack, there are two 3 × 3 convolutions and a 2 × 2 max pooling operation with stride 2 for down-sampling, where each convolution is followed by a rectified linear unit (ReLU) (as the activation function) and batch normalization [30]. Each decoder stack consists of a transposed convolution with kernel size 2 × 2 and a stride of 2, a concatenation operation, and two 3 × 3 convolutions with ReLU and batch normalization. The transposed convolution is utilized for up-sampling. With the concatenation operations, feature maps from the encoder stack are fused into the corresponding decoder stack. At the last layer, the sigmoid activation is employed to map obtained features to the segmentation probability map.

Loss function

One challenge in medical image segmentation arises from class imbalance in the data. In the present data set, PC lesions, although sometimes numerous, occupy only a small part of the image volume. This class-imbalance between healthy and lesion tissue will result in sub-optimal segmentation performance when using the conventional categorical cross-entropy loss. The Dice similarity coefficient (DSC) is a typical criterion used to evaluate the accuracy of anatomic segmentation [31]. The Dice loss function is a modification of the DSC first proposed to tackle the class-imbalance issue [32], and subsequently applied in variants forms for diverse medical image segmentation tasks [21, 33, 34]. Since Dice loss has proved to be well adapted to accommodate the high imbalance problem, we chose in this work to use a multiclass Dice loss function [35] to tackle this issue. Assuming p is the output of the network, and g is the encoding of the ground truth segmentation map, the multiclass Dice loss is defined as:

$$L_{dc} = -\frac{2}{|K|}\sum_{k\in K}\frac{\sum_i p_i^k g_i^k}{\sum_i p_i^k \sum_i g_i^k} \quad (1)$$

where, p and g have shape $I \times K$ with $i \in I$ being the number of pixels in the training patch and $k \in K$ being the classes. The term $|K|$ denotes the number of elements in one set.

Lesion detection based on the segmentation result

During staging and re-staging of PC, it is of great importance to know the location of all tumor lesions. Our proposed lesion detection approach is based on the segmentation result obtained by a deep learning network. In contrast to detection task conventionally used in the computer vision community, which has a regular bounding box or sphere as the ground truth, we use the manually annotated lesion areas as the ground truth, which is geometrically more irregular but anatomically more accurate. In this work, we assume that the lesion is a topologically connected volume, such that non-contiguous volumes qualify as separate lesions. During lesion detection, all contiguous voxel clusters designated as different types of lesion (bone, lymph node, or local lesion) are recognized after receiving the segmentation result of the deep neural network. We then filter out contiguous structures of volume less than a cutoff threshold ($T_V = 25$ mm^3), to avoid possible false positive predictions caused by noise.

Furthermore, we evaluated the lesion detection accuracy based on the overlap between a predicted lesion and ground truth. A lesion was considered to be correctly detected whenever the overlap ratio exceeded a threshold ($T_E = 10\%$), in consideration of the likely scale of misregistration between PET and CT images. We selected this threshold setting by comparing the detection performance with various other thresholds (i.e., 20%, 30%), noting that the current threshold leads to the best detection accuracy, which remained relatively stable at the changes (see supplemental Fig. 2). We then calculated the magnitudes of the indices of detection accuracy known as precision (positive predictive value), recall (sensitivity), and F1-score (harmonic mean of precision and recall). The segmentation performance was evaluated using dice similarity coefficient, positive predictive value, and the specificity.

Experimental setup and parameter selection

The dataset consisting of PET-CT scans from 193 PC patients was randomly divided into a training set from 130 patients and a test set from 63 patients, thus about 30% of the entire dataset (Each patient has one PET image and one CT image). The test set was exclusively reserved for independent validation of the developed method and was constructed to include nearly equal proportions of patients from the three research sites, i.e., 15 patients from Technical University of Munich, 16 from University of Munich, and 32 from the University of Bern. During the training of the network using data from the 130 patients, we employed fivefold cross validation [36] on the training set to optimize the hyperparameters of the network and to test for influences of random differences in the training dataset composition on the network performance.

Each PET image was co-registered to its corresponding CT, and all the imaging data of the three centers were resampled to the same pixel size $1 \times 1 \times 1$ mm^3 using b-spline interpolation, to accommodate slight differences instrumentation. Z-score normalization (standardizations) of CT Hounsfield units of CT and PET tracer uptake applied to both PET and CT images. The network takes three neighboring slices as the input. Each input slice consists of one corresponding PET slice and one CT slice as two channels. The details of the input strategy are illustrated in Fig. 2. The ADAM optimizer [37] was used during training with an initial learning rate $lr_{init} = 10^{-2}$. The learning rate was reduced by a factor of two upon stagnation of the learning. To regularize the network, we utilized the early stopping strategy with the patience of 20, which is a criterion often employed to detect the convergence of training and thereby avoiding overfitting [38]. The term "patience" refers to the number of epochs to wait before early stop when there is no progress in the validation set [38]. The fully convolutional nature of our network allows the processing of arbitrarily sized inputs. Therefore, we segment the entire target region for each patient during the test phase.

Results

Detection accuracy

After obtaining optimized hyperparameters of the proposed network from the abovementioned cross-validation strategy, we then utilized the entire training dataset of images from 130 patients to train the network to calculate the optimized weights of each network layer. The detection results of all the annotated lesions of the reserved test set from 63 patients for the three types of PC lesions are listed in Table 1. The developed method achieved more than 99% accuracy for precision, recall and F1-score in the detection of bone lesions, versus more than 90% for lymph node lesions. The detection accuracy of local lesions remained relatively low (< 80%).

Table 1 Detection accuracy of the proposed network for three different kinds of prostate cancer lesions. The accuracy of each type of lesion is calculated for all the annotated lesions

	Bone lesion	Lymph node lesion	Local lesion
Precision	0.99	0.94	0.79
Recall	0.99	0.90	0.61
F1 score	0.99	0.92	0.69

In order to investigate the influence of the different training data on the performance, we evaluated the performance of the trained models obtained during the cross-validation phase on the reserved test dataset. These models were trained on different subsets of the entire training dataset (in the cross-validation procedure, one fold is selected as the validation set each time and the remaining four folds are for training). Supplemental Fig. 1 demonstrates the influence of the training data set composition on the detection accuracy by illustrating the detection results of different trained models of each fold (called fivefold cross test for the convenience of following description). There was some variation between different folds in the detection accuracy for both bone and lymph node lesions; for bone lesions, the developed method achieved more than 95% detection accuracy (precision, recall, and F1-score) in four different folds, where fold 1 had a precision of 71.7% and F1-score of 83.5%. For the detection of lymph node lesions, four folds had a detection accuracy around 90%, with fold 3 having slightly worse performance of 78.8% precision and 83.5% F1-score.

We further compared the detection results on the reserved test dataset for the model trained with the entire training data (denoted as the fully trained model) and the best performing model arising from the fivefold cross-test phase (denoted as cross-best model) in Fig. 3. The fully trained model achieved better precision, recall, and F1-score in the detection of bone lesions, and better precision but worse recall than the cross-best model for lymph node lesions.

Segmentation accuracy

The segmentation accuracy of the developed approach on the ^{68}Ga-PSMA-11 PET/CT dataset are presented in Fig. 4. The deep learning method achieved average DSCs of 64.5% for bone lesions and 54.4% for lymph node lesions. The positive predictive value (PPV) was 79.9% and the specificity 60.6% for the segmentation of bone lesions, versus 66.9% and 55.2% for lymph node lesions.

Figure 4 also shows the comparison of segmentation accuracy when deep learning was applied on combined PET/CT and PET or CT data alone. The segmentation accuracy learning with only the CT modality was less than 10% for bone lesions and lymph node lesions. With PET alone, there was a considerable improvement in the segmentation accuracy learning, which reached 51.5% (DSC), 61.9% (PPV), and 54.8% (specificity) for bone lesions and 50.3% (DSC), 59.7% (PPV), and 54.5% (specificity) for lymph node lesions. Learning on combined PET/CT gave much better results, with DSC and PPV exceeding than the sum of individual DSC and PPV for separate PET and CT modalities.

Figure 5 illustrates an example of the segmentation results obtained by the developed method for bone and lymph node lesions, where the true positive, false positive, and false negative are marked with different colors. Typically, false negatives occurred with small lesions, insufficient contrast compared with background. False positives occurred in relation to signal intensity, with misattribution of regions of high non-specific tracer uptake.

Discussion

Clinical research findings show that PC tumor burden and corresponding therapeutic radioligand dose received by the tumor correlate with the treatment response of PSMA-directed radioligand therapy [39]. As in the case of external beam radiation therapies, the identification and characterization of target lesions of the treatment is increasingly important for treatment planning of internal radionuclide therapy. However, the PSMA radionuclide therapy is typically reserved for patients with multiple or disseminated metastases. In such cases, it can be impractical to delineate manually all

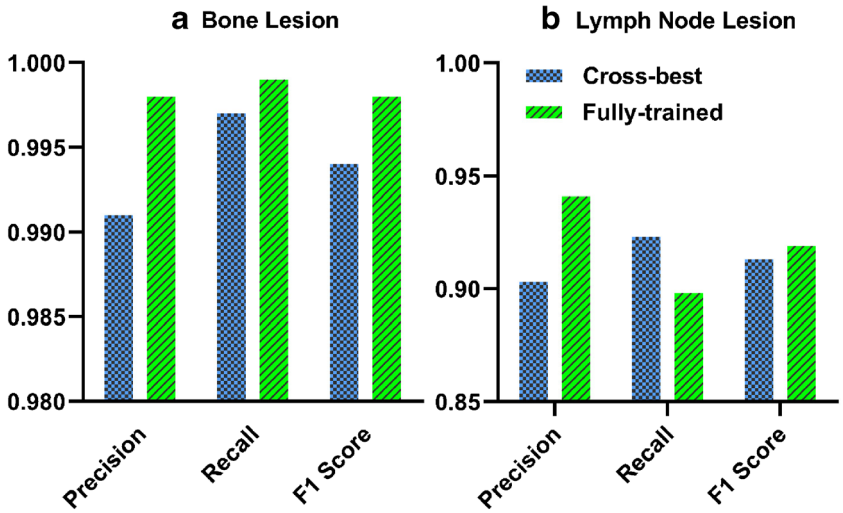

Fig. 3 Comparison of the detection performance between the cross-best (best performing model arising from the fivefold cross-test phase) model and fully-trained model (model trained with the entire training data). **a** Bone lesions, **b** lymph node lesion

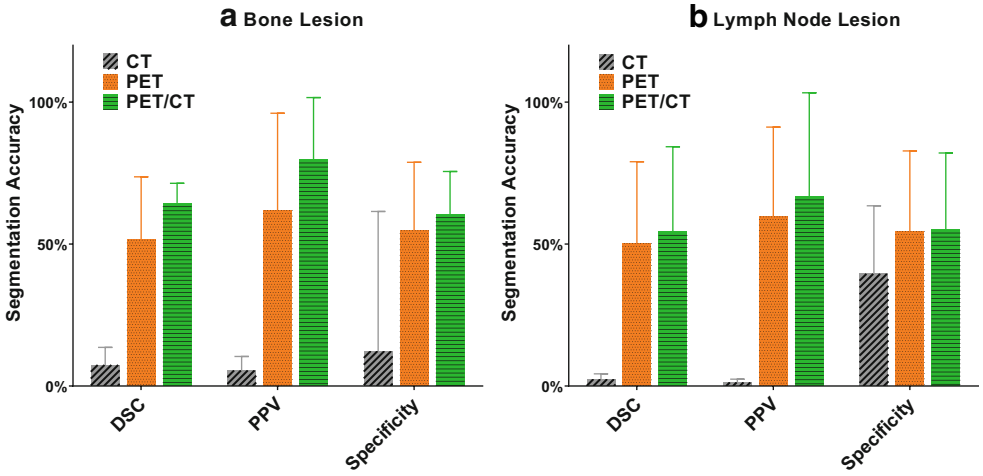

Fig. 4 Comparison of the segmentation accuracy of the deep learning network when applied to CT alone, PET alone, or PET/CT. DSC is dice similarity coefficient and PPV represents positive prediction value

lesions present, which may take many hours of expert attention for a subject. Therefore, the automation of lesion detection and segmentation has the potential to assist profoundly in the treatment planning (and response monitoring) of radionuclide therapy. In this study, we developed a triple-combining 2.5D U-Net method for automatic detection and segmentation of PC lesions on ^{68}Ga-PSMA-11 PET/CT images.

Although deep learning has demonstrated its advantages in many applications for image analysis, the amount and variability of the training data plays a critical role in the procedure's performance. The variation we saw in lesion detection accuracy for the cross-fold test shows how distinctly the selection of training data can influence the performance of the deep learning network. Specifically, there was a fold for bone

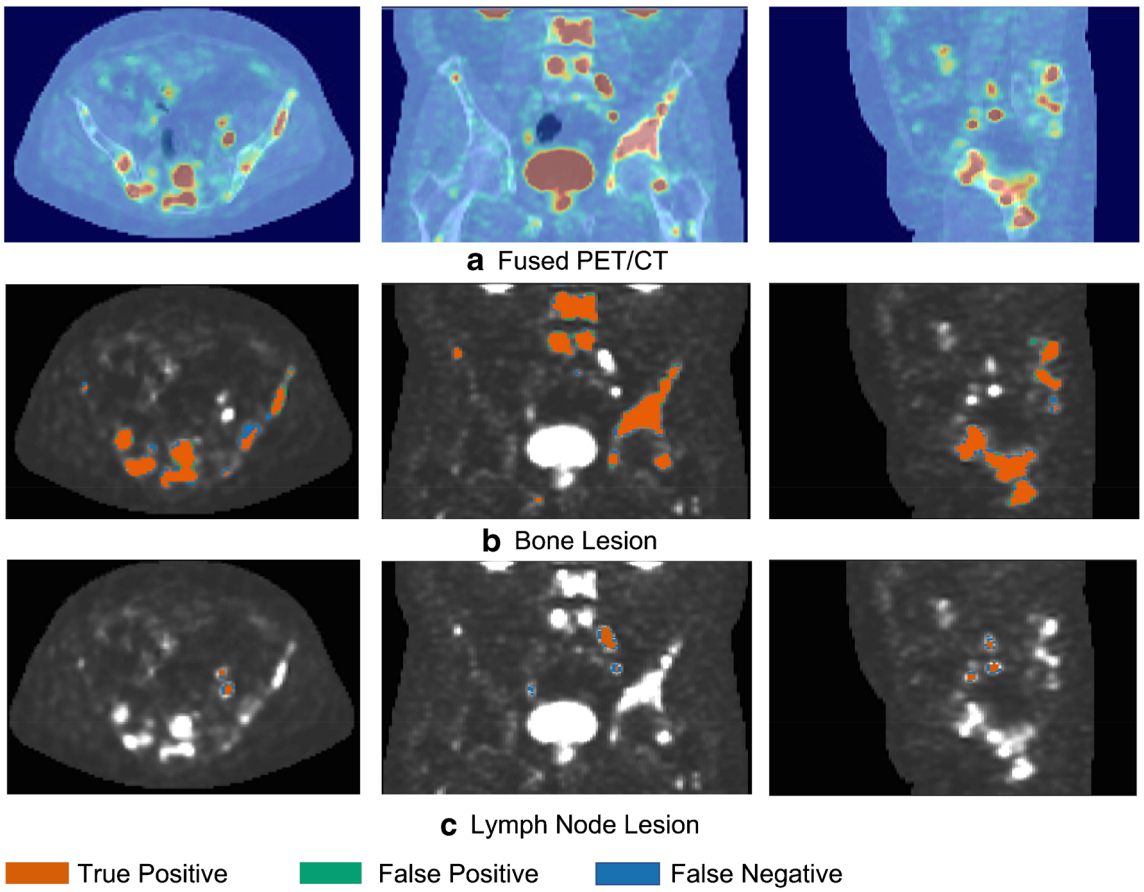

Fig. 5 Examples of the segmentation results of the proposed triple-combining 2.5D U-Net for the bone and lymph node lesions. The first row shows the fused PET/CT slices, the second row and the third row show the detection result of the proposed network (bone lesion and lymph node lesion respectively). Axial, sagittal, and coronal views were illustrated from left to right

lesion and lymph node lesion detection, which had discernibly worse performance in the cross-fold test. This phenomenon could relate to the scale and composition of the training set, which may have been insufficient to give a stable solution to the deep learning training. Alternately, small site differences in instrumentation or imaging protocol may have introduced heterogeneity into the data, resulting in sensitivity to the training-set composition. Furthermore, lesions are not homogeneously distributed across patients; some may have an abundance of bone and lymph node lesions, whereas other may have only isolated lesions of one type. This could result in unbalanced division for the cross-fold test, leading to reduced performance. Nevertheless, our entire cohort of 193 cases was sufficiently large to support accurate detection of bone and lymph node lesions. Therefore, the training and test sets were adequate to achieve relatively good results for the development and validation.

Currently, the accuracy of detection and segmentation is distinctly more challenging for local (recurrent) lesions than for bone and lymph node metastases. This accuracy difference is consistent with the different numbers of the lesions in the dataset, which confirms that the amount of data plays a critical role in the accuracy. Furthermore, there was relatively greater heterogeneity for the size and intensity of tracer uptake in the local lesions, and most had no discernible local lesions, as expected after surgical resection. As such, present results most likely do not yet depict the full capability of deep learning methods in the detection and segmentation of local lesions.

Our results also demonstrated that the segmentation accuracy when trained with PET alone is much superior to that with CT alone, whereas having the combination of PET and CT modalities had a super-additive effect on the accuracy attained. This is consistent with the superior performance of PSMA PET/CT in the detection and characterization of bone and lymph node metastasis in mCRPC patients by manual interpretation of the fusion images [8]. Inclusion of CT brings additional anatomical information, which evidently assists the deep learning algorithm better to identify the lesions.

We included PSMA PET/CT data from three centers to assemble a large dataset and to test the robustness of the deep learning methods to small local differences in protocol. Theoretically, deep learning models are able learn a hierarchy of features that can distinguish low-level domain-specific features of the scanner and imaging protocols from high-level object-specific features of the patient [40]. Therefore, deep learning is more flexible and robust to cross-domain differences such as might arise from slightly different scanners or protocols. Despite possible sources of variability between centers, the admixture of data from three centers could still support network training with admirable certain accuracy, which confirms the method`s relative robustness compared with conventional machine learning approaches. The collaboration of three centers in this development enhances the potential of the algorithm to be applied to other centers and to accelerate its clinical translation. Each of the participating centers possesses a relatively large database of PSMA PET/CT images. The possibility of compiling retrospective data from different centers enhances the potential of further development of the deep learning methods with very large data sets containing thousands of cases, which can further increase the performance. However, the annotation of the data remains a logistic bottleneck for the compilation of very large databases. There is at present no substitute for manual labeling. However, there might be scope for improving the efficiency and interoperator reliability of data annotation, perhaps by an iterative process with an initial automatic processing followed by manual editing of the annotation.

Compared with the lesion detection accuracy, the segmentation accuracy of our methods remained relatively low, in particular for lymph node metastases and local tumors. This could be due to several reasons: (1) an inadequately sized training set, which limits the potential of the network; (2) mixture of local lesions that are primary PC and local recurrent tumors after prostatectomy, which have differing contrast against background; (3) The large variability of number, size, shape, and uptake intensity of the lymph node metastases. On the other hand, we used the manual delineation as the reference standard to calculate the Dice similarity coefficients, despite the known propensity for interobserver differences in delineation [41]. We also note that, due to partial volume effects, the visible lesion boundaries may deviate from the location of pathological abnormalities. Therefore, manual segmentation cannot be a perfect ground truth in the estimation of segmentation accuracy. Theoretically, only pathological boundaries can be considered as ground truth, which is missed in this study. 3D pathology distribution of the prostatectomy specimen and fusion with preoperative PET/CT using an advanced co-registration algorithm can be used to determine the ground-truth boundary for primary lesions [42]. However, it is very difficult to obtain the histopathological boundary of metastatic lesions, which are the main focus of the current treatment of radionuclide therapy. Specific resection, pathology, and co-registration algorithm should be further developed. Since it is basically impossible to obtain the pathology for all the lesions, alternatives, according to the crowd-sourcing theory of deep learning such as training with a large data set with manual annotations by a large number of diverse experts may assemble the knowledge and converge the segmentation to the underlying pathology [43].

The physicians, who helped to analyze the data of our study, are well experienced and familiar with possible pearls and pitfalls of PSMA-imaging. This includes ganglia as well as inflammatory processes within the prostate or in lymph nodes. Although recently, a magnitude of PSMA-positive lesions of non-prostatic origin were published (almost all of them as case reports), it must be stated that these entities

represent exceptions and come up with a very low number compared with the probable thousands of true-positive lesions detected every day worldwide. In addition, the PSMA-positive lesions of non-prostatic origin usually present with as significantly lower tracer uptake compared with prostate cancer lesions. This statement is also valid for inflammatory lymph nodes [44]. In addition, we notice the multitude of studies confirming the excellent specificity (95–97%) of ^{68}Ga-PSMA-11 [45–51].

We chose to confine our proof-of-concept study to the pelvic area, but it shall be necessary eventually to extend the PC lesion detection and segmentation to the whole body. We do not expect that the data assembled thus far is sufficient to support training the network for a whole-body search. This preliminary test has already demonstrated the capability of the proposed network in lesion detection, and we are enlarging the dataset. The dataset will include more than 1000 cases, which should be adaptable to train our model for whole-body lesion detection and segmentation (we conducted a survey and found that deep learning methods have achieved decent performance in different medical image segmentation tasks with training datasets including around 1000 samples or, in most cases, less than this number [52–58].). Furthermore, we will also collect and analyze the histological data to make the study more comprehensive. A possible direction for further methodological development is to employ a region proposal network (like R-CNN) for initial localization, followed by application of a segmentation-specific network, which might improve the currently inadequate performance for assessing local lesions. In addition, using the generative adversarial network to generate more useful training samples or applying the semi-supervised concept in the network could alleviate some issues arising from insufficient training data. Assembling several different networks also has the potential to enhance further the detection and segmentation accuracy. Finally, while our procedure gives a considerable degree of accuracy relative to the ground truth, we have not yet evaluated clinical benefits of automated lesion detection and segmentation for treatment planning. This is the topic of ongoing research at our sites.

Conclusion

This pilot study explores the capacity of deep learning methods for lesion detection and segmentation of PC lesions in ^{68}Ga-PSMA-11 PET/CT scans. The preliminary test confined to the pelvic area confirmed the potential of deep learning methods for this task, setting the stage for a more extensive data collection and annotation. Increasing the amount of training data should further enhance the performance of the developed deep learning methods, especially in light of the requirement for whole-body assessments.

Acknowledgements We would like to gratefully acknowledge NVIDIA Corporation for the donation of a Titan XP GPU used for this research and thank Professor Paul Cumming for critical reading of the manuscript.

Funding information This study was funded by UniBern Forschungstiftug and Swiss Krebsliga (KFS-4723-02-2019).

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not describe any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Maurer T, Eiber M, Schwaiger M, Gschwend JE. Current use of PSMA–PET in prostate cancer management. Nat Rev Urol. 2016;13:226.
3. Howlader N, Noone A, Krapcho M, Neyman N, Aminou R, Altekruse S, et al. SEER cancer statistics review, 1975-2009 (vintage 2009 populations), National Cancer Institute. Bethesda 2012.
4. Bernacki KD, Fields KL, Roh MH. The utility of PSMA and PSA immunohistochemistry in the cytologic diagnosis of metastatic prostate carcinoma. Diagn Cytopathol. 2014;42:570–5.
5. Fizazi K, Faivre L, Lesaunier F, Delva R, Gravis G, Rolland F, et al. Androgen deprivation therapy plus docetaxel and estramustine versus androgen deprivation therapy alone for high-risk localised prostate cancer (GETUG 12): a phase 3 randomised controlled trial. Lancet Oncol. 2015;16:787–94.
6. Attard G, Parker C, Eeles R, Schroffder F, Tomlins SA, Tannock I, et al. Prostate cancer. Lancet. 2016;387:70–82.
7. Weineisen M, Schottelius M, Simecek J, Baum RP, Yildiz A, Beykan S, et al. 68Ga-and 177Lu-labeled PSMA I&T: optimization of a PSMA-targeted theranostic concept and first proof-of-concept human studies. J Nucl Med. 2015;56:1169–76.
8. Afshar-Oromieh A, Holland-Letz T, Giesel FL, Kratochwil C, Mier W, Haufe S, et al. Diagnostic performance of 68 Ga-PSMA-11 (HBED-CC) PET/CT in patients with recurrent prostate cancer: evaluation in 1007 patients. Eur J Nucl Med Mol Imaging. 2017;44:1258–68. https://doi.org/10.1007/s00259-017-3711-7.
9. Kratochwil C, Bruchertseifer F, Giesel FL, Weis M, Verburg FA, Mottaghy F, et al. 225Ac-PSMA-617 for PSMA-targeted a-radiation therapy of metastatic castration-resistant prostate cancer. J Nucl Med. 2016;57:1941–4.
10. Kratochwil C, Schmidt K, Afshar-Oromieh A, Bruchertseifer F, Rathke H, Morgenstern A, et al. Targeted alpha therapy of mCRPC: dosimetry estimate of 213Bismuth-PSMA-617. Eur Jo Nucl Med Mol Imaging. 2018;45:31–7.

11. Rahbar K, Schmidt M, Heinzel A, Eppard E, Bode A, Yordanova A, et al. Response and tolerability of a single dose of 177Lu-PSMA-617 in patients with metastatic castration-resistant prostate cancer: a multicenter retrospective analysis. J Nucl Med. 2016;57:1334–8.
12. Afshar-Oromieh A, Haberkorn U, Zechmann C, Armor T, Mier W, Spohn F, et al. Repeated PSMA-targeting radioligand therapy of metastatic prostate cancer with 131 I-MIP-1095. Eur J Nucl Med Mol Imaging. 2017;44:950–9.
13. Hofman MS, Violet J, Hicks RJ, Ferdinandus J, Thang SP, Akhurst T, et al. [177Lu]-PSMA-617 radionuclide treatment in patients with metastatic castration-resistant prostate cancer (LuPSMA trial): a single-centre, single-arm, phase 2 study. Lancet Oncol. 2018;19:825–33.
14. Eiber M, Fendler WP, Rowe SP, Calais J, Hofman MS, Maurer T, et al. Prostate-specific membrane antigen ligands for imaging and therapy. J Nucl Med. 2017;58:67S–76S.
15. Bieth M, Krönke M, Tauber R, Dahlbender M, Retz M, Nekolla SG, et al. Exploring new multimodal quantitative imaging indices for the assessment of osseous tumor burden in prostate cancer using 68Ga-PSMA PET/CT. J Nucl Med. 2017;58:1632–7.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436.
17. Leynes AP, Yang J, Wiesinger F, Kaushik SS, Shanbhag DD, Seo Y, et al. Zero-echo-time and dixon deep pseudo-CT (ZeDD CT): direct generation of pseudo-CT images for pelvic PET/MRI attenuation correction using deep convolutional neural networks with multiparametric MRI. J Nucl Med. 2018;59:852–8. https://doi.org/10.2967/jnumed.117.198051.
18. Spuhler KD, Gardus J 3rd, Gao Y, DeLorenzo C, Parsey R, Huang C. Synthesis of patient-specific transmission image for PET attenuation correction for PET/MR imaging of the brain using a convolutional neural network'. J Nucl Med. 2018. https://doi.org/10.2967/jnumed.118.214320.
19. Torrado-Carvajal A, Vera-Olmos J, Izquierdo-Garcia D, Catalano OA, Morales MA, Margolin J, et al. Dixon-VIBE Deep Learning (DIVIDE) Pseudo-CT Synthesis for Pelvis PET/MR Attenuation correction. J Nucl Med. 2018. https://doi.org/10.2967/jnumed.118.209288.
20. Ciresan D, Giusti A, Gambardella LM, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems; 2012. p. 2843-51.
21. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 3431-40.
22. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention: Springer; 2015. p. 234-41.
23. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48.
24. Hatt M, Laurent B, Ouahabi A, Fayad H, Tan S, Li L, et al. The first MICCAI challenge on PET tumor segmentation. Med Image Anal. 2018;44:177–95. https://doi.org/10.1016/j.media.2017.12.007.
25. Corral JE, Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Su1337-deep learning to diagnose intraductal papillary mucinous neoplasms (IPMN) with MRI. Gastroenterology. 2018;154:S-524–S-5.
26. Bi L, Kim J, Kumar A, Wen L, Feng D, Fulham M. Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies. Comput Med Imaging Graph. 2017;60:3–10. https://doi.org/10.1016/j.compmedimag.2016.11.008.
27. Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, et al. Automated whole-body bone lesion detection for multiple myeloma on 68Ga-pentixafor PET/CT imaging using deep learning methods. Contrast Media Mol Imaging. 2018;2018:11. https://doi.org/10.1155/2018/2391925.
28. Sironi A, Türetken E, Lepetit V, Fua P. Multiscale centerline detection. IEEE Trans Pattern Anal Mach Intell. 2015;38:1327–41.
29. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2016. p. 424-32.
30. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167. 2015.
31. Hwang D, Kim KY, Kang SK, Seo S, Paeng JC, Lee DS, et al. Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning. J Nucl Med. 2018;59:1624–9. https://doi.org/10.2967/jnumed.117.202317.
32. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 3D Vision (3DV), 2016 Fourth International Conference on: IEEE; 2016. p. 565-71.
33. Bailey J, Piert M. Performance of 68Ga-PSMA PET/CT for prostate cancer management at initial staging and time of biochemical recurrence. Curr Urol Rep. 2017;18:84.
34. Hoffmann MA, Miederer M, Wieler HJ, Ruf C, Jakobs FM, Schreckenberger M. Diagnostic performance of 68Gallium-PSMA-11 PET/CT to detect significant prostate cancer and comparison with 18FEC PET/CT. Oncotarget. 2017;8:111073.
35. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. International MICCAI Brainlesion Workshop: Springer; 2017. p. 287-97.
36. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv. 2010;4:40–79.
37. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
38. Chollet F. Keras. 2015.
39. Violet J, Jackson P, Ferdinandus J, Sandhu S, Akhurst T, Iravani A, et al. Dosimetry of (177)Lu-PSMA-617 in metastatic castration-resistant prostate cancer: correlations between pretherapeutic imaging and whole-body tumor dosimetry with treatment outcomes. J Nucl Med. 2019;60:517–23. https://doi.org/10.2967/jnumed.118.219352.
40. Ghafoorian M, Mehrtash A, Kapur T, Karssemeijer N, Marchiori E, Pesteie M, et al. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer; 2017. p. 516-24.
41. Karki K, Saraiya S, Hugo GD, Mukhopadhyay N, Jan N, Schuster J, et al. Variabilities of magnetic resonance imaging-, computed tomography-, and positron emission tomography-computed tomography-based tumor and lymph node delineations for lung cancer radiation therapy planning. Int J Radiat Oncol Biol Phys. 2017;99:80–9. https://doi.org/10.1016/j.ijrobp.2017.05.002.
42. Zamboglou C, Carles M, Fechter T, Kiefer S, Reichel K, Fassbender TF, et al. Radiomic features from PSMA PET for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate-and high-risk prostate cancer-a comparison study with histology reference. Theranostics. 2019;9:2595.
43. Albarqouni S, Baur C, Achilles F, Belagiannis V, Demirci S, Navab N. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. IEEE Trans Med Imaging. 2016;35:1313–21.
44. Afshar-Oromieh A, Sattler LP, Steiger K, Holland-Letz T, da Cunha ML, Mier W, et al. Tracer uptake in mediastinal and paraaortal thoracic lymph nodes as a potential pitfall in image interpretation

of PSMA ligand PET/CT. Eur J Nucl Med Mol Imaging. 2018;45:1179–87.
45. Afshar-Oromieh A, Avtzi E, Giesel FL, Holland-Letz T, Linhart HG, Eder M, et al. The diagnostic value of PET/CT imaging with the 68 Ga-labelled PSMA ligand HBED-CC in the diagnosis of recurrent prostate cancer. Eur J Nucl Med Mol Imaging. 2015;42:197–209.
46. Eiber M, Maurer T, Souvatzoglou M, Beer AJ, Ruffani A, Haller B, et al. Evaluation of hybrid 68Ga-PSMA ligand PET/CT in 248 patients with biochemical recurrence after radical prostatectomy. J Nucl Med. 2015;56:668–74.
47. Sahlmann C-O, Meller B, Bouter C, Ritter CO, Ströbel P, Lotz J, et al. Biphasic 68 Ga-PSMA-HBED-CC-PET/CT in patients with recurrent and high-risk prostate carcinoma. Eur J Nucl Med Mol Imaging. 2016;43:898–905.
48. Herlemann A, Wenter V, Kretschmer A, Thierfelder KM, Bartenstein P, Faber C, et al. 68Ga-PSMA positron emission tomography/computed tomography provides accurate staging of lymph node regions prior to lymph node dissection in patients with prostate cancer. Eur Urol. 2016;70:553–7.
49. Maurer T, Weirich G, Schottelius M, Weineisen M, Frisch B, Okur A, et al. Prostate-specific membrane antigen–radioguided surgery for metastatic lymph nodes in prostate cancer. Eur Urol. 2015;68:530–4.
50. Pfister D, Porres D, Heidenreich A, Heidegger I, Knuechel R, Steib F, et al. Detection of recurrent prostate cancer lesions before salvage lymphadenectomy is more accurate with 68 Ga-PSMA-HBED-CC than with 18 F-Fluoroethylcholine PET/CT. Eur J Nucl Med Mol Imaging. 2016;43:1410–7.
51. Hijazi S, Meller B, Leitsmann C, Strauss A, Meller J, Ritter C, et al. Pelvic lymph node dissection for nodal oligometastatic prostate cancer detected by 68Ga-PSMA-positron emission tomography/computerized tomography. Prostate. 2015;75:1934–40.
52. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2014;34:1993–2024.
53. Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ayed IB. HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation. IEEE Trans Med Imaging. 2018;38:1116–26.
54. Oktay O, Ferrante E, Kamnitsas K, Heinrich M, Bai W, Caballero J, et al. Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans Med Imaging. 2017;37:384–95.
55. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches. IEEE Trans Med Imaging 2019.
56. Leclerc S, Smistad E, Pedrosa J, Østvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Trans Med Imaging 2019.
57. Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, et al. Automatic multi-organ segmentation on abdominal CT with dense v-networks. IEEE Trans Med Imaging. 2018;37:1822–34.
58. Wang L, Nie D, Li G, Puybareau É, Dolz J, Zhang Q, et al. Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge. IEEE Trans Med Imaging 2019.