European Journal of Radiology 113 (2019) 89–95

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Deep learning for segmentation of 49 selected bones in CT scans: First step in automated PET/CT-based 3D quantification of skeletal metastases

Sarah Lindgren Belal[a,*], May Sadik[b], Reza Kaboteh[b], Olof Enqvist[c], Johannes Ulén[d], Mads H. Poulsen[e], Jane Simonsen[f], Poul F. Høilund-Carlsen[f], Lars Edenbrandt[b], Elin Trägårdh[a,g]

[a] Department of Translational Medicine, Lund University, Malmö, Sweden
[b] Department of Clinical Physiology, Sahlgrenska University Hospital, Göteorg, Sweden
[c] Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden
[d] Eigenvision AB, Malmö, Sweden
[e] Department of Urology, Odense University Hospital, Odense, Denmark
[f] Department of Nuclear Medicine, Odense University Hospital, Odense, Denmark
[g] Wallenberg Center for Molecular Medicine, Lund University, Malmö, Sweden

ARTICLE INFO

Keywords:
PET/CT
Artificial intelligence
Bone
Metastases
Prostate cancer

ABSTRACT

Purpose: The aim of this study was to develop a deep learning-based method for segmentation of bones in CT scans and test its accuracy compared to manual delineation, as a first step in the creation of an automated PET/CT-based method for quantifying skeletal tumour burden.

Methods: Convolutional neural networks (CNNs) were trained to segment 49 bones using manual segmentations from 100 CT scans. After training, the CNN-based segmentation method was tested on 46 patients with prostate cancer, who had undergone ^{18}F-choline-PET/CT and ^{18}F-NaF PET/CT less than three weeks apart. Bone volumes were calculated from the segmentations. The network's performance was compared with manual segmentations of five bones made by an experienced physician. Accuracy of the spatial overlap between automated CNN-based and manual segmentations of these five bones was assessed using the Sørensen-Dice index (SDI). Reproducibility was evaluated applying the Bland-Altman method.

Results: The median (SD) volumes of the five selected bones were by CNN and manual segmentation: Th7 41 (3.8) and 36 (5.1), L3 76 (13) and 75 (9.2), sacrum 284 (40) and 283 (26), 7th rib 33 (3.9) and 31 (4.8), sternum 80 (11) and 72 (9.2), respectively. Median SDIs were 0.86 (Th7), 0.85 (L3), 0.88 (sacrum), 0.84 (7th rib) and 0.83 (sternum). The intraobserver volume difference was less with CNN-based than manual approach: Th7 2% and 14%, L3 7% and 8%, sacrum 1% and 3%, 7th rib 1% and 6%, sternum 3% and 5%, respectively. The average volume difference measured as ratio volume difference/mean volume between the two CNN-based segmentations was 5–6% for the vertebral column and ribs and ≤3% for other bones.

Conclusion: The new deep learning-based method for automated segmentation of bones in CT scans provided highly accurate bone volumes in a fast and automated way and, thus, appears to be a valuable first step in the development of a clinical useful processing procedure providing reliable skeletal segmentation as a key part of quantification of skeletal metastases.

1. Introduction

Bone scan is commonly used as a first-line modality to detect bone changes due to metastatic disease in prostate cancer [1]. It is also used in many clinical trials to evaluate treatment effect in patients with metastatic prostate cancer. The interpretation of bone scans is subjective [2]. A research group at Memorial Sloan Kettering Cancer Center in New York has presented a quantitative method to assess tumor burden to the skeleton called the bone scan index (BSI) [3]. An automated method to calculate the BSI has been developed, making it more

Abbreviations: BSI, bone scan index; 3D, three-dimensional; PET/CT, positron emission tomography/computed tomography; CNN, convolutional neural network; SDI, Sørensen-Dice index; SD, standard deviation; NaF, ^{18}F-sodium fluoride; mCRPC, metastatic castration-resistant prostate cancer; SUV, standardized uptake value

* Corresponding author at: Department of Translational Medicine Division of Nuclear Medicine, Lund University, c/o Elin Trägårdh, Carl-Bertil Laurells gata 9, 202 05 Malmö, Sweden.

E-mail address: sarah.lindgren_belal@med.lu.se (S. Lindgren Belal).

https://doi.org/10.1016/j.ejrad.2019.01.028
Received 5 October 2018; Received in revised form 11 January 2019; Accepted 26 January 2019

clinically useful as shown in several papers [4]. In a recent meta-analysis based on 14 studies with data from 1295 patients with metastatic prostate cancer, the authors found that baseline BSI and change in BSI during treatment were associated with survival [5]. This strongly suggests that a quantitative imaging biomarker can carry prognostic information of value for clinical management and future clinical trials.

BSI is derived from a two-dimensional technique while the skeleton is a three-dimensional (3D) structure. With increasing use of in particular positron emission tomography/computed tomography (PET/CT), a corresponding fast and quantifiable measure of bone changes assessed by PET/CT is of obvious clinical relevance.

The first step to achieve this goal is accurate segmentation of the skeleton from PET/CT scans. A logic approach would be to use the CT part for segmentation and the PET part to record the tracer activity in the thus segmented skeletal volumes. The various approaches proposed for segmentation of bony structures in CT scans [6] are mostly semi-automated and focusing on parts of the skeleton. To be of clinical use, segmentation of bone in CT scans needs to be fast, reliable and comprising relevant major parts of the skeleton.

Machine learning, a branch of artificial intelligence, has gained increased usage in medical imaging. A recent approach in machine learning uses convolutional neural network (CNN) deep learning algorithms. The parameters of a deep networks are updated through a general learning procedure exploiting the input data [7]. CNN-based methods have been used for automated volumetric CT scan segmentation [8], but remains to our knowledge to be introduced into the field of PET/CT for 3D quantification of skeletal metastases.

The aim of this project was to develop a CNN-based method for automated segmentation of bones from CT scans, to test its accuracy against a set of manual segmentations, and to evaluate its reproducibility from repeat automated CNN-based segmentation sessions.

2. Material and methods

2.1. Study design

CNNs were trained to automatically segment bones in the axial skeleton. Manual segmentations performed by a single physician in 100 CT scans were used as training material. Forty-nine bones (12 thoracic vertebrae, 5 lumbar vertebrae, sacrum, 2 hip bones, 24 ribs, 2 scapulae, 2 clavicles and the sternum) were segmented in each CT scan. Cervical vertebrae were not included due to these bones' complex shape and small size relative to the slice thickness of the CT scans. Femora and humeri were not segmented as the whole bone was not always included in the CT studies.

The CNN-based segmentation method was validated using a separate group of 46 patients with prostate cancer. All 46 patients in the validation group had undergone ^{18}F-choline PET/CT (CT_1) and ^{18}F-sodium fluoride (NaF) PET/CT (CT_2) within a period of three weeks as part of a previous research project. The same 49 bones were automatically segmented by the CNN in both CT studies from all 46 patients. Additionally, a specialist in medical radiology and nuclear medicine manually segmented five bones (7th thoracic vertebrae [Th7], 3rd lumbar vertebrae [L3], sacrum, right 7th rib, and sternum) in both CT studies from five randomly selected patients in the validation group.

The accuracy of the automated segmentations was assessed against the manual segmentation by the radiology and nuclear medicine specialist. The reproducibility of the CNN-based method was evaluated through comparison of the two automated segmentation sessions.

2.2. Patients

2.2.1. Training group

As training material, ^{18}F-FDG PET/CT examinations from 100 patients were consecutively collected between August and December 2010 at Sahlgrenska University Hospital, Göteborg, Sweden. The cause of referral was known or suspected tumours. The patients signed an informed consent and the study was approved by the regional ethical review board in Göteborg, Sweden (295-08).

2.2.2. Validation group

A group of patients with biopsy-proven prostate cancer who had previously been included in a study at Odense University Hospital, Denmark, and in a Swedish study, was used [9,10]. The aim of the original Danish study was to compare whole-body bone scans, ^{18}F-choline PET/CT and NaF PET/CT with magnetic resonance imaging. A current bone scan with a minimum of one metastasis was required for inclusion in the study. A total of 50 patients were included between May 2009 and March 2012. Four patients missed ^{18}F-choline PET/CT imaging because of radiotracer production failure. The remaining 46 patients with two CT scans from the PET/CT studies comprised the validation group. Ethical approval was granted by the Danish Patient Safety Authority (3-3013-1692/1) and the regional ethical review board in Lund, Sweden (2016/443).

2.3. Imaging protocols

2.3.1. Training group

PET/CT scans were obtained by a Biograph 64 TruePoint (Siemens Healthineers). A low-dose CT scan (120 kV, 30 mAs) was obtained. The CT scan was reconstructed in a field of view of 70 cm using filtered back projection, slice thickness of 5 mm and spacing of 3 mm.

2.3.2. Validation group

PET/CT scans were obtained by a Discovery VCT 64 (GE Healthcare). A low-dose CT scan was acquired using tube current modulation (SmartmA, 120 kV, 80–400 mAs) and was reconstructed in a field of view of 50 cm using filtered back projection, slice thickness of 3.75 mm and spacing of 3.27 mm.

2.4. Automated segmentation

The core of the automated segmentation approach is a CNN [11,12] that essentially gives every voxel in the input image a label. Recently, CNNs have risen in popularity and are now state of the art for virtually all of the major segmentation challenges [13,14].

For a CNN, the resulting label for each pixel is typically only affected by a sub-patch of the input image, the receptive field. Due to the resolution of the 3D data it is infeasible to use a model with a receptive field covering the full image. The network we use has a receptive field of 52 × 92 × 92 voxels (Fig. 1). This leads to problems such as differentiating between adjacent vertebrae or ribs. To handle this issue, our approach consists of three steps:

1 A set of anatomical landmarks are detected using a CNN. The identity of each landmark is determined using a shape model.
2 The identified landmarks are fed as auxiliary input to a CNN that performs the final bone segmentation.
3 Post-processing to remove spurious pixels.

2.4.1. Step 1: Landmark detection

A CNN was trained to detect a number of anatomical landmarks (Fig. 2). More precisely, the CNN outputs a probability map with high values in a blob at each landmark. Non-maximum suppression is then applied to get discrete detections. The CNN detects types of landmarks such as rib joints and vertebral processes but due to the limited receptive field it cannot detect their identity, i.e. which rib it belongs to. To determine this and remove possible spurious detections, classical active shape models are used to find plausible relative positions for groups of landmarks. A second network was trained to detect rib centerlines and an iterative tracking scheme was used to track each rib from a starting point at the corresponding rib joint.

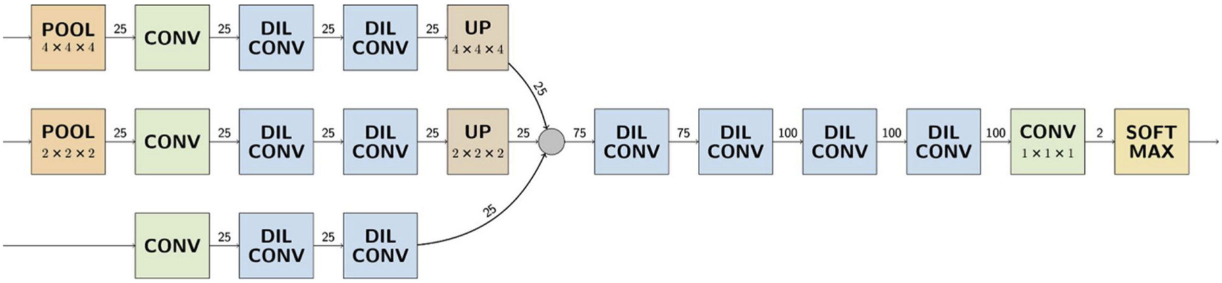

Fig. 1. Structure of the fully convolutional neural network used. All convolutions are 3 × 3 × 3 except the last one. For the dilated convolutions the first ones have dilation rate (1,2,2) whereas all the other have (2,4,4). The difference in the first dimension is due to lower spatial resolution in the images. The numbers on the lines indicate the number of data channels after each operation. The number of channels after the last convolution is the number of output labels for each network.

2.4.2. *Step 2: Voxel-wise segmentation*

The important outputs from Step 1 are the positions of the transverse processes of the lumbar vertebrae and centerlines of the twelve pair of ribs. Using binary coding, these 29 landmark pairs were used to form a five-dimensional image mask that was fed to the final CNN together with the original image. The five landmark channels provide the CNN additional information used to correctly differentiate between similar bones such as adjacent ribs. The CNN was trained using stochastic gradient descent over the 100-scan training set.

2.4.3. *Step 3: Post-processing*

The raw segmentation from the CNN may contain some spurious voxels. To mitigate this problem, some very simple post-processing was performed.

Each bone was associated with either zero or at most one type of landmark. For bones associated with one landmark, only the connected component overlapping the corresponding landmark mask was kept. For the other bones, the largest connected component was kept. As a final step, binary hole filling was performed in order to remove small errors.

2.4.4. *Network training*

All networks are trained by stochastic gradient descent using the ADAM optimizer [15] with Nesterov momentum and categorical cross entropy as loss function.

When creating training data for the different CNN models patches are chosen such that the number of voxels for each class is roughly equal and the background class consist of 50% of all voxels.

2.5. *Manual segmentation*

2.5.1. *Training group*

A single physician manually segmented 49 bones in 100 CT scans, within a time frame of two months. A customized cloud-based segmentation tool was used (eScan Research, eScan Academy AB, Lund, Sweden) to label all voxels corresponding to a specific bone in any of the three planes (axial, coronal or sagittal) in each slice.

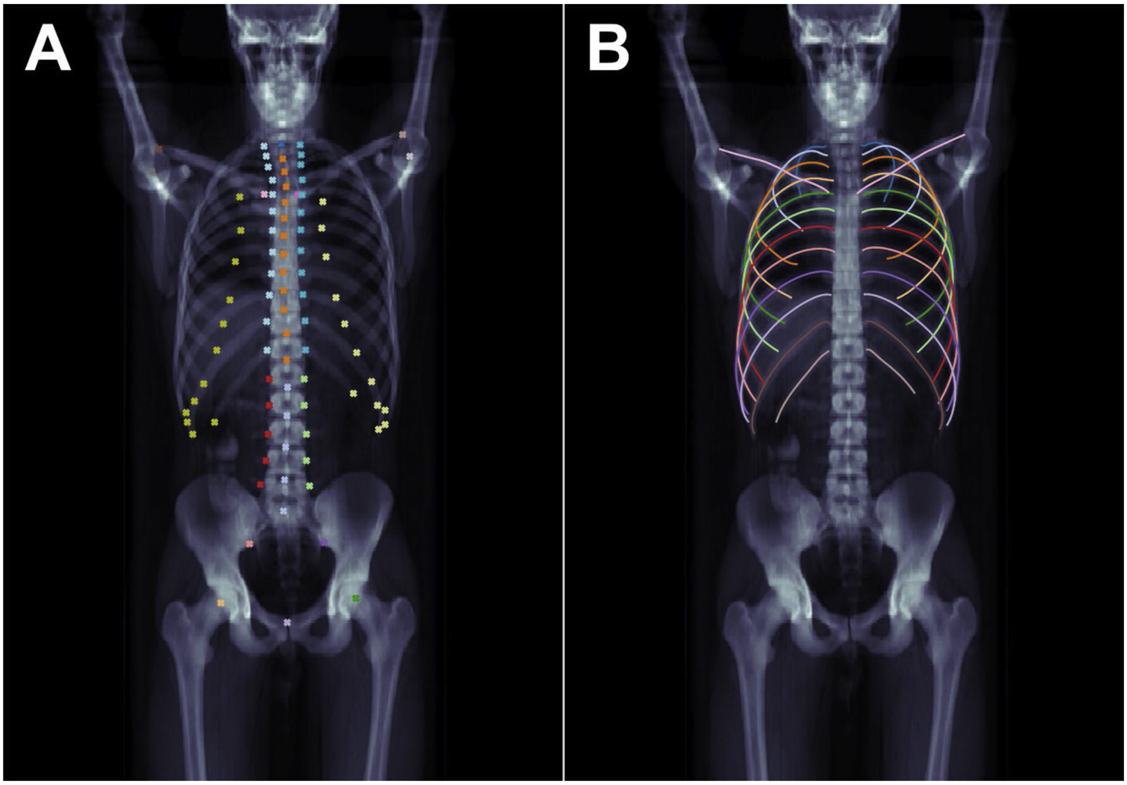

Fig. 2. (A) Maximum intensity projection, in the range 0–400 Hounsfield units, of the CT scan together with the annotated landmarks. Landmarks with identical markers belong to the same class and are not separated by the detector. (B). Detected center lines for ribs and clavicles.

2.5.2. Validation group

Five randomly chosen patients from the validation group were used for manual segmentation. An experienced specialist in medical radiology and nuclear medicine manually segmented five different bones (Th7, L3, sacrum, right 7th rib and sternum) in both CT scans from these five patients. This generated a total of 25 paired segmentations of the same bones from the two studies. The physician was blinded to the automated segmentations of the validation group. The same segmentation tool was used as for the training group.

2.6. Statistical methods

To evaluate the accuracy of the CNN-based method, Sørensen-Dice index (SDI) [16] was used to analyze the overlap between the automated and the manual segmentations in the training group. The SDI evaluates how well two different segmentations of the same bone agree with each other. A voxel is defined as belonging to the overlap of two segmentations if it is classified as bone in both segmentations. The SDI is defined as two times the number of overlapping voxels divided by the sum of the total amount of voxels that are classified as bone in both segmentations. The value ranges between 0 and 1, where 1 reflects a perfect segmentation.

To assess the reproducibility of the volume calculation by the CNN, the Bland-Altman method was used [17]. The difference between the two automated volume measurements was calculated for the 49 bones. The mean and standard deviation (SD) of the difference between the two automated volume calculations were used to calculate the upper and lower confidence limit.

All calculations were performed using the software IBM SPSS Statistics 24.

3. Results

Patient characteristics for the training and validation groups are presented in Table 1. The training group consisted of both women and men with various diagnoses, while the validation group consisted of men with prostate cancer. The patient age was on average 11 years lower in the training group than in the validation group.

3.1. Segmentation-based volume calculation

In the validation group, volumes of the 49 bones were calculated based on the automated segmentations from CT_1 and CT_2 in the 46 patients. The volumes for an individual bone ranged from 7 ± 2.8 mL (mean ± SD) for the 12th rib, to 436 ± 44.5 mL (mean ± SD) for the pelvic bone. The total volume for all 49 bones in one patient was 3018 ± 220 mL (mean ± SD).

A reconstruction of the 49 segmented bones in one patient from the validation group is illustrated in Fig. 3.

Table 1
Patient characteristics.

	Training group Value	Validation group Value
Number of patients	100	46
Age, years		
Mean (range)	62 (27–85)	73 (53–92)
Gender		
Male	60	46
Female	40	–
Diagnosis		
Prostate cancer	–	46
Lung cancer	19	–
Colorectal cancer	15	–
Head and neck cancer	14	–
Other	52	–

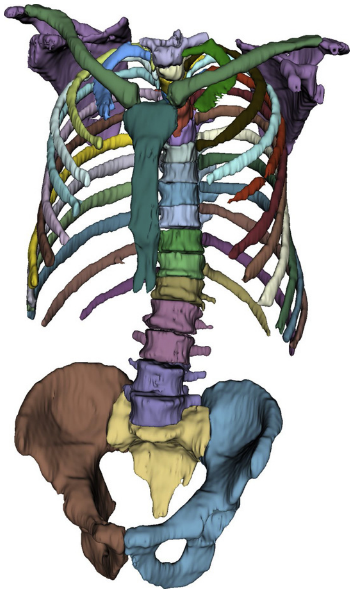

Fig. 3. Reconstruction of automated convolutional neural network-based segmentations of 49 bones in one patient from the validation group.

Table 2
Accuracy assessment. Sørensen-Dice index (SDI) for assessment of spatial overlap between convolutional neural network-based and manual segmentations in the same CT scan from five patients in the validation group.

SDI	Median	Range
Th7	0.86	0.42-0.89
L3	0.85	0.72-0.90
Sacrum	0.88	0.76-0.89
Right 7th rib	0.84	0.82-0.86
Sternum	0.83	0.80-0.87

3.2. Accuracy

For the subset of five cases, the SDI assessing the overlap between the CNN-based segmentations and the manual segmentations by the physician are presented in Table 2. An example of both the CNN-based and the manual segmentations in one CT study is displayed in Fig. 4. In one of the cases, the CNN-based method and the physician segmented different but adjacent vertebrae as Th7.

The volumes derived from the manual and the CNN-based segmentations in the subset of five cases are presented in Table 3. The volume of each individual bone was higher when measured by the CNN-based compared to the manual method.

3.3. Reproducibility

A scatter plot (Fig. 5A) and a Bland-Altman plot (Fig. 5B) displaying the correlation and agreement between the bone volumes from the two automated segmentations in the 46 patients are illustrated in Fig. 5. The mean volume difference between the two automated segmentations of the same bone was 0.3 mL with a SD of 3.3 mL. The average volume difference measured as ratio volume difference/mean volume between the two CNN-based segmentations was 5–6% for the vertebral column and ribs and ≤3% for all other bones.

Volume difference for both the CNN-based and manual segmentations between CT_1 and CT_2 in the five subset cases are displayed in Table 3.

For the CNN-based method to generate automated segmentations took approximately two minutes per CT scan. The manual segmentations of five bones in the validation set took approximately 50 min per

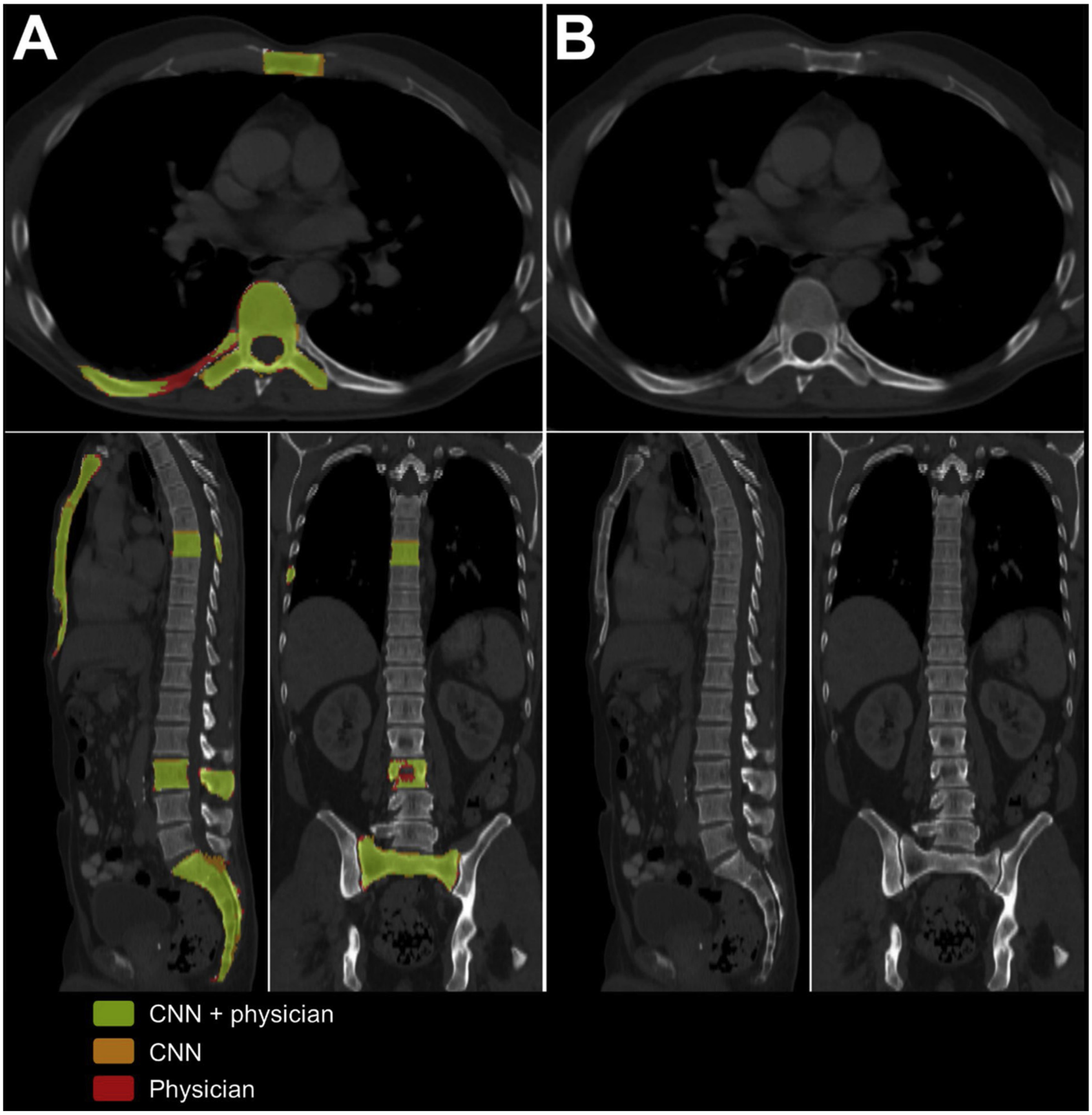

Fig. 4. (A) Axial (upper half), sagittal and coronal reconstructions (lower half) of one CT study showing convolutional neural network (CNN)-based and manual segmentations of five selected bones (Th7, L3, sacrum, right 7th rib and sternum). Green represents the area segmented as the selected bone by both the CNN-based method and the physician, orange by only the CNN-based method and red by only the physician. (B) The same reconstructions without segmentations for comparison.

CT study.

Table 3
Bone volumes and reproducibility assessment. Volumes of five selected bones from five patients in the validation group. For each case, repeated segmentations of the selected bones were performed separately from two CT scans (CT_1 and CT_2) by both the convolutional neural network (CNN) and the physician. The median volume is derived from the 10 individual segmentations. The differences within the five pairs were calculated.

	Median volume, mL (SD*, CV† [%])		Difference, %	
	CNN	Manual	CNN	Manual
Th7	41 (3.8, 0.1)	36 (5.1, 0.1)	2	14
L3	76 (13, 0.2)	75 (9.2, 0,1)	7	8
Sacrum	284 (40, 0.1)	283 (26, 0.1)	1	3
Right 7th rib	33 (3.9, 0.1)	31 (4.8, 0.2)	1	6
Sternum	80 (11, 0,1)	72 (9.2 0.1)	3	5

*Standard deviation, †Coefficient of variation

4. Discussion

4.1. Main findings

We present a new CNN-based method for automated segmentation of bone in CT scans focusing on skeletal parts most commonly affected by metastases from prostate cancer. Using a separate validation material, we found that the performance of the CNN-based method in segmenting 49 bones in the axial skeleton was comparable to that of an experienced physician in medical radiology and nuclear medicine, both in terms of accuracy and reproducibility.

Without a known gold standard to measure bone volume, manual segmentations by a specialist in radiology and nuclear medicine were used for comparison with the CNN-based automated segmentations. The automated segmentations showed a very high accuracy when compared to manual segmentations and yielded bone segmentations and volume calculations about 25 times faster than the physician.

The agreement for segmentation-based volume calculation was

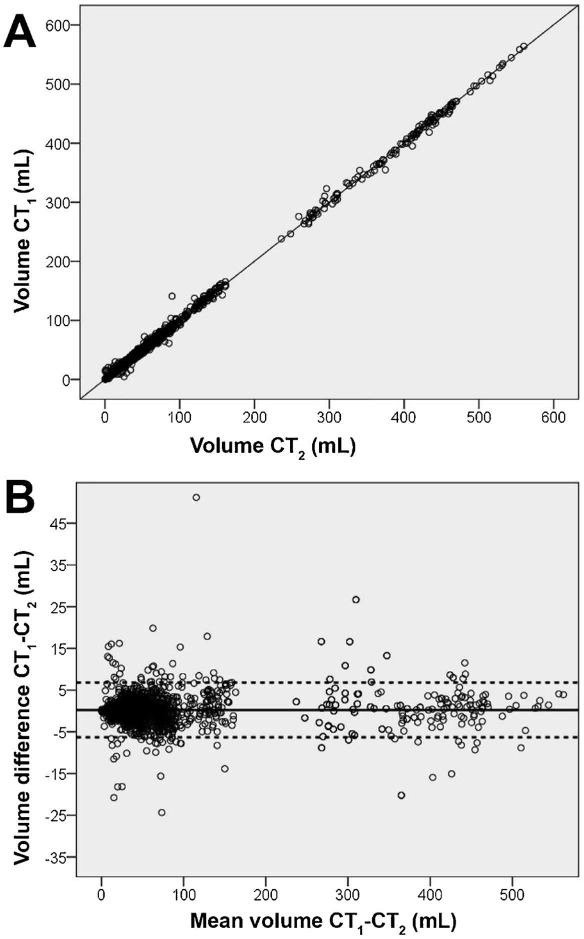

Fig. 5. (A) Scatter plot illustrating the correlation between the automated volume calculations of 49 bones in 46 patients in two CT studies (CT_1 and CT_2). (B) Bland-Altman plot for reproducibility assessment of the volume calculations. Mean volume difference 0.3 mL (solid horizontal line), with upper confidence limit of 6.8 mL and lower confidence limit of -6.3 mL (dotted horizontal lines).

higher for the CNN-based than the manual method. The mean volume difference of the same bone in the two different CT scans was less than half a milliliter. This suggests excellent reproducibility for the CNN-based method.

In one case, the CNN-based method and the physician segmented different, but adjacent, vertebrae as Th7, contributing to a lower SDI for this bone. The CNN tries to identify the first set of ribs and assumes that the corresponding vertebra is Th1. The reason for not starting from C1 is that the cervical vertebrae are harder to detect in the images. The physician always labelled the vertebrae starting cranially, and so the explanation in this particular case could be that the patient had a lot of degenerative changes which made counting of the vertebrae, especially the cervical, challenging. We also found that in relatively many patients, the vertebrae count differed from the usual number of 33.

4.2. Study strengths

The deep learning approach that was used in this project is not dependent on a pre-conceived feature model and, therefore, requires extensive training to learn and improve. It is largely dependent upon the quality and size of the data set that it is being fed [7,18]. Using CT scans from patients with heterogeneous biological characteristics for training increases the performance of the method in recognizing different skeletal patterns. Thus, it was a strength of the study that the training group comprised of patients with different gender and ages and various diagnoses.

The method was validated using CT studies from a different hospital, acquired by different equipment and another acquisition protocol. Given the positive results in this study, it points toward good consistency of the method irrespective of different locations and certain technical aspects. However, it would require further studies to properly evaluate the consistency and robustness of the method. This was outside the scope of the present study.

4.3. Clinical implications

The presented very fast automated segmentation approach represents a first promising step toward a highly needed automated PET/CT-based imaging methodology for quantification of skeletal metastases in prostate cancer. Other authors have recognized the need for objective methods to quantify the skeletal tumor burden in PET/CT. Rohren et al. proposed a method for quantification of skeletal tumor burden from NaF PET/CT scans [19]. The same authors showed in a retrospective clinical study that quantification of tumor burden at baseline is predictive of overall survival in patients with castration-resistant prostate cancer (mCRPC) treated with radium [20]. However, their method is semi-automated and involves placement of a large volume of interest surrounding the entire skeleton in a PET scan, and manual exclusion of non-metastatic uptake areas. In another recent study, Harmon et al. showed that metrics derived from NaF PET/CT early in treatment in patients with mCRPC, receiving chemotherapy or androgen receptor inhibitors, are correlated with clinical and progression-free survival [21]. The authors presented a new semi-automated analysis tool identifying lesions by a threshold-defined standardized uptake value (SUV), where extra-skeletal tissue is excluded by means of an anatomic CT mask, after which different SUV metrics are derived. A required step in this process is correction by a nuclear medicine specialist to exclude benign uptake. The manual component in both these approaches makes processing more time-consuming than with a fully automated method like the present one and probably results in higher intra- and inter-observer variability.

In order to obtain a PET/CT-based index of skeletal tumor burden it is essential to have an accurate segmentation of relevant skeletal structures as well as metastatic changes. To escape manual interference, we are planning to train our fully automated CNN-based method to also detect and measure radiotracer uptake due to metastases in PET/CT scans. The goal is subsequently to develop a quantitative measure of tumor burden in the skeleton. Metastatic lesions are often located in cancellous bone which is rich in red bone marrow, while degenerative disease is often located near the joint surface. Therefore, we will refine the method to detect uptake due to metastatic lesions not solely based on standardized uptake values, but also based on location.

In the setting of prostate cancer, the benefit of CNN-based segmentation reaches beyond the assessment of radiotracer uptake due to metastatic skeletal disease in NaF PET/CT. Prostate cancer can also spread to regional lymph nodes and more uncommonly can also metastasize to viscera [22]. We are currently working on developing similar CNN-based segmentation methods for other organs of interest. The methods will not be radiotracer specific, but can be applied to PET/CT scans using different radiotracers such as ^{18}F-choline and prostate-specific membrane antigen.

An imaging biomarker is in line with personalized cancer management, can aid in both diagnosis and risk stratification and act as decision-support when it comes to treatment planning. The implications for clinical practice could be substantial in terms of improved radiologist performance and cost-efficiency.

4.4. Study limitations

For the accuracy assessment, we used segmentations performed by a specialist in medical radiology and nuclear medicine. We had the physician perform repeated manual segmentations of five bones in both CT scans in five different cases. We then compared the overlap of these manual segmentations with the CNN-based automated segmentations, to calculate the SDI. Ideally, more manual segmentations could have been used for this assessment. However, being such a labor-intensive task, this was not feasible within the time-frame of our project. The manual segmentations were based on the performance of one physician and, therefore, did not necessarily represent the absolute truth. However, we considered analysis of 25 pairs acceptable and sufficiently informative at this early stage of the development.

The time-consuming nature of the manual segmentation procedure was also a major limiting factor for the number of CT studies available for training. In addition, it is a challenge to harvest large enough patient image data sets for this purpose due to healthcare IT, ethical and legal restrictions [23]. Despite this, we were able to establish 100 manual segmentations of 49 bones and plan to include even larger training materials in the future to optimize our model.

A third limitation was that not all bones in the axial skeleton could be segmented, due to rather large slice thicknesses in the training set and since the whole skull was not included in the CT acquisition. Also, the humeri and femora were not segmented, since the whole bones were not always included in the CT. Excluding the cervical spine, femora, humeri, and skull is concerning for the ultimate goal of assessing tumor burden as these bones contain on average approximately 14% of bone metastases [24]. These bones will therefore be considered in future editions of our CNN-based approach.

5. Conclusion

This new deep learning-based method for automated segmentation of bones in CT scans is a fast, accurate, and reproducible and performs equally well as experienced image reader, except at much higher speed. A robust and fast way to automatically measure bone volume is a key component in the development of a PET/CT-base index relating radiotracer uptake to bone volume in metastasizing prostate cancer.

Summary statement

The deep learning-based method presented in this work for automated segmentation and volume calculation of bones in CT scans is important for the development of a PET/CT index relating volumes of abnormal PET tracer uptake to bone volume, which in turn can add information to the diagnosis, prognosis and treatment planning in patients with metastasized prostate cancer.

Declaration of interest

Professor Lars Edenbrandt works as a part-time consultant for EXINI Diagnostics AB.

Funding

This work was supported by the Swedish Research Council, Region Skåne, Knut and Alice Wallenberg Foundation and EXINI Diagnostics AB.

References

[1] H.I. Scher, M.J. Morris, W.M. Stadler, et al., Trial design and objectives for castration-resistant prostate Cancer: updated recommendations from the prostate Cancer Clinical trials working group 3, J. Clin. Oncol. 34 (12) (2016) 1402–1418.
[2] M. Sadik, M. Suurkula, P. Höglund, A. Järund, L. Edenbrandt, Quality of planar whole-body bone scan interpretations–a nationwide survey, Eur. J. Nucl. Med. Mol. Imaging 35 (8) (2008) 1464–1472.
[3] M. Imbriaco, S.M. Larson, H.W. Yeung, et al., A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index, Clin. Cancer Res. 4 (7) (1998) 1765–1772.
[4] D. Ulmert, R. Kaboteh, J. Fox, et al., A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index, Eur. Urol. 26 (1) (2012) 78–84.
[5] D. Li, H. Lv, X. Hao, Y. Dong, H. Dai, Y. Song, Prognostic value of bone scan index as an imaging biomarker in metastatic prostate cancer: a meta-analysis, Oncotarget 8 (48) (2017) 84449–84458.
[6] N. Sharma, L.M. Aggarwal, Automated medical image segmentation techniques, J. Med. Phys. 35 (1) (2010) 3–14.
[7] B.J. Erickson, P. Korfiatis, Z. Akkus, T.L. Kline, Machine learning for medical imaging, Radiographics 37 (2) (2017) 505–515.
[8] Q. Dou, L. Yu, H. Chen, et al., 3D deeply supervised network for automated segmentation of volumetric medical images, Med. Image Anal. 41 (2017) 40–54.
[9] M.H. Poulsen, H. Petersen, P.F. Høilund-Carlsen, et al., Spine metastases in prostate cancer: comparison of technetium-99m-MDP whole-body bone scintigraphy, [(18) F]choline positron emission tomography(PET)/computed tomography (CT) and [(18) F]NaF PET/CT, BJU Int. 114 (6) (2014) 818–823.
[10] S. Lindgren Belal, M. Sadik, R. Kaboteh, et al., 3D skeletal uptake of 18F sodium fluoride in PET/CT images is associated with overall survival in patients with prostate cancer, EJNMMI Res. 7 (1) (2017) 15.
[11] I. Goodfellow, Y. Bengio, Courville A. Deep Learning, MIT Press, Cambridge, MA, 2016.
[12] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Conference on Computer Vision and Pattern Recognition: IEEE (2015).
[13] T. Lin, M. Maire, S. Belongie, et al., Microsoft COCO: Common objects in context, European Conference on Computer Vision, Springer, Cham, 2014.
[14] M. Everingham, L. Van Gool, C. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[15] D. Kingma, J. Ba, Adam: a method for stochastic optimization, The International Conference on Learning Representations: arXiv.Org (2014).
[16] L. Dice, Measures of the amount of ecological association between species, Ecology. 26 (1945) 297–302.
[17] J.M. Bland, D.G. Altman, Measuring agreement in method comparison studies, Stat. Methods Med. Res. 8 (2) (1999) 135–160.
[18] K.J. Dreyer, J.R. Geis, When machines think: radiology's next frontier, Radiology 285 (3) (2017) 713–718.
[19] E.M. Rohren, E.C. Etchebehere, J.C. Araujo, et al., Determination of skeletal tumor burden on 18F-Fluoride PET/CT, J. Nucl. Med. 56 (10) (2015) 1507–1512.
[20] E.C. Etchebehere, J.C. Araujo, P.S. Fox, N.M. Swanston, H.A. Macapinlac, E.M. Rohren, Prognostic factors in patients treated with 223Ra: the role of skeletal tumor burden on baseline 18F-Fluoride PET/CT in predicting overall survival, J. Nucl. Med. 56 (8) (2015) 1177–1184.
[21] S.A. Harmon, T. Perk, C. Lin, et al., Quantitative assessment of early [(18)F]Sodium fluoride positron emission Tomography/Computed tomography response to treatment in men with metastatic prostate Cancer to bone, J. Clin. Oncol. 35 (24) (2017) 2829–2837.
[22] American Cancer Society, (2019) Retrieved January 2nd from https://www.cancer.org/cancer/prostate-cancer/treating/recurrence.html.
[23] M.A. Morris, B. Saboury, B. Burkett, J. Gao, E.L. Siegel, Reinventing radiology: big data and the future of medical imaging, J. Thorac. Imaging 33 (1) (2018) 4–16.
[24] C. Wang, Y. Shen, Study on the distribution features of bone metastases in prostate cancer, Nucl. Med. Commun. 33 (4) (2012) 379–383.