Deep learning approaches for bone and bone lesion segmentation on 18 FDG PET/CT imaging in the context of metastatic breast cancer*

Noémie Moreau, Caroline Rousseau, Constance Fourcade, Gianmarco Santini, Ludovic Ferrer, Marie Lacombe, Camille Guillerminet, Mario Campone, Mathilde Colombié, Mathieu Rubeaux, et al.

Deep learning approaches for bone and bone lesion segmentation on ^{18}FDG PET/CT imaging in the context of metastatic breast cancer*

Noémie Moreau[1,2], Caroline Rousseau[3,4], Constance Fourcade[2], Gianmarco Santini[2], Ludovic Ferrer[3,4], Marie Lacombe[4], Camille Guillerminet[4] Mario Campone [3,4], Mathilde Colombié [4], Mathieu Rubeaux[2] and Nicolas Normand[1]

***Abstract*— ^{18}FDG PET/CT imaging is commonly used in diagnosis and follow-up of metastatic breast cancer, but its quantitative analysis is complicated by the number and location heterogeneity of metastatic lesions. Considering that bones are the most common location among metastatic sites, this work aims to compare different approaches to segment the bones and bone metastatic lesions in breast cancer.**

Two deep learning methods based on U-Net were developed and trained to segment either both bones and bone lesions or bone lesions alone on PET/CT images. These methods were cross-validated on 24 patients from the prospective $EPICURE_{seinmeta}$ metastatic breast cancer study and were evaluated using recall and precision to measure lesion detection, as well as the Dice score to assess bones and bone lesions segmentation accuracy.

Results show that taking into account bone information in the training process allows to improve the precision of the lesions detection as well as the Dice score of the segmented lesions. Moreover, using the obtained bone and bone lesion masks, we were able to compute a PET bone index (PBI) inspired by the recognized Bone Scan Index (BSI). This automatically computed PBI globally agrees with the one calculated from ground truth delineations.

***Clinical relevance*— We propose a completely automatic deep learning based method to detect and segment bones and bone lesions on ^{18}FDG PET/CT in the context of metastatic breast cancer. We also introduce an automatic PET bone index which could be incorporated in the monitoring and decision process.**

I. INTRODUCTION

Breast cancer is the most common cancer in women and the second most frequent cancer overall. One in eight women will be diagnosed with invasive breast cancer in their lifetime. Breast cancer survival varies according to cancer staging at diagnosis. If detected early, the overall 5-years survival rate is 98% but it goes down to 27% with metastatic involvement [1].

^{18}FDG positron emission tomography combined with computed tomography (^{18}FDG PET/CT) whole-body imaging is widely used for diagnosis and follow-up [2]. Based on these imaging techniques, lesion segmentation can provide information to assess a treatment effect and to adapt the treatment over time. However, manual segmentation methods are time-consuming and subject to inter and intra-observer variability. Automatic and semi-automatic PET segmentation methods based on computer vision have been developed but their performance are affected by low contrast or heterogeneity in tumoral tissues [3]. Recently, deep learning based methods have been shown to outperform more conventional approaches [4].

Even if deep learning methods show good results for lesion segmentation in solid tumors, the heterogeneity of metastatic lesions in location, contrast and form can be very difficult to learn for a network. Since most deep learning methods are specialized on a single body part [4], an interesting approach would be to train different networks according to the location of the metastases. In the case of breast cancer, metastatic lesions are mostly located in the bones (69%), lungs (27%), liver (27%) [5], and lymph nodes [6]. As a first step towards metastatic breast cancer lesion segmentation and characterization, we chose to concentrate our effort on the prevalent bone lesion detection and segmentation.

Two deep learning methods based on convolutional neural networks (CNN) to segment bone lesions are presented: the first one is trained with expert bone lesion annotations solely, while the second also uses bone masks during the training phase to force the network to focus on the sites of interest.

Finally, the bone and the bone lesion segmentations allow to compute the percentage of total skeletal mass taken up by the tumors as done with the Bone Scan Index (BSI) [7]. The BSI was originally developed for prostate cancer but was then applied in the context of metastatic breast cancer [8, 9]. Building upon these works, we propose to transpose the BSI to PET imaging to automatically assess breast cancer metastatic burden in the bones.

II. METHODS

A. Preprocessing

PET images were converted to Standard Uptake Value (SUV) using the injected radioactivity's concentration and the patient's body weight according to the standard developed by Kim et al. [10] and were also normalized with the mean and the standard deviation and resampled by the nnU-Net preprocessing (see Fig. 1).

CT images and bone masks (when used during training) were resampled to match PET data.

*This work is partially financed through "Programme opérationnel regional FEDERFSE Pays de la Loire 2014-2020" noPL0015129 (EPICURE) and by the SIRIC ILIAD Nantes-Angers-INCADGOS-Inserm 12558 grant.

[1]Université de Nantes, CNRS, LS2N, F-44000 Nantes, France

[2]Keosys Medical Imaging, Nantes, France

[3]University of Nantes, CRCINA, INSERM UMR1232, CNRS-ERL6001, Nantes, France

[4]ICO Cancer Center, Nantes - Angers, France

CT images were also clipped to their 0.5^{th} and 99.5^{th} percentile values and normalized by the nnU-Net preprocessing.

B. CNN architecture

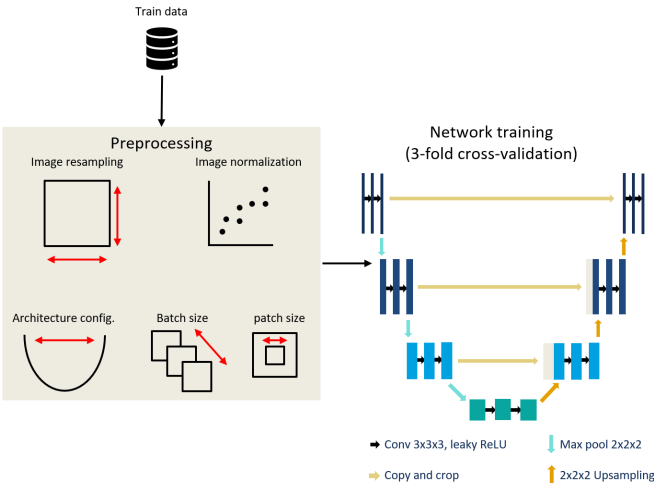


Fig. 1. The nnU-Net preprocessed the data with image resampling and normalization. It automatically sets the batch size (min. size of 2), patch size and number of pooling operations, while maximizing the amount of spatial context. (Figure inspired from [11].)

The proposed architecture is based on a recent 3D U-Net implementation called "not new U-Net" (nnU-Net) [12]. The U-Net is an encoder-decoder network, developed by Ronneberger, Fischer, and Brox [13] and commonly used for semantic segmentation of medical images. The nnU-Net takes advantage of recent improvements of the original U-Net such as leaky ReLU activation, instance normalization and padded convolutions. This network achieved state of the art performance in recent challenges [14].

The originality and interest of the nnU-Net implementation lies particularly in its capability to automatically set a number of hyper-parameters, taking into consideration several constraints such as input data features and memory consumption as shown in Fig. 1. The loss is defined as the sum of the cross entropy loss (L_{CE}) and the multi-class Dice loss (L_{Dice}): $L_{Total} = L_{Dice} + L_{CE}$, with L_{CE} as in [13] and L_{Dice} as in [12]. Stochastic gradient descent with Adam optimizer is employed with an initial learning rate of 3×10^{-4}, a drop factor of 0.2 after 30 epochs without improvement (1 epoch being defined as 250 batches), and a weight decay of 3×10^{-5}. Training ends after 1000 epochs or when the learning rate falls below 10^{-6}. Data augmentation composed of elastic deformations, random scaling, random rotations, and gamma augmentation was also used during training to counteract the relatively small database.

III. DATA AND EXPERIMENTS

A. Data acquisition and annotation

Twenty-four patients were recruited in the context of the prospective $EPICURE_{seinmeta}$ metastatic breast cancer study (NCT03958136). Data acquisition was performed in two institutions with different imaging systems. At the Integrated Center for Oncology (ICO) of Nantes, PET/CT images were obtained using either a Siemens Biograph mCT40 or mCT64 PET/CT imaging system. At the ICO of Angers, PET/CT images were obtained using a dual-slice GE discovery 690 or IQ PET/CT.

The $EPICURE_{seinmeta}$ study was approved by the French Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM, 2018-A00959-46) and the Comité de Protection des Personnes (CPP) IDF I, Paris, France (CPPIDF1-2018-ND40-cat.1), and a written informed consent was signed by each participant.

Bone lesions were manually annotated using Keosys Viewer [15] by one expert from each institution. Bone segmentations were first extracted automatically using a set of traditional morphological and thresholding procedures inspired by Banik, Rangayyan, and Boag [16] and then manually corrected by 4 non-specialist image processing researchers. Seven of the twenty-four patients did not have any bone lesions.

B. Experiments

All training experiments were conducted using a 3-fold cross-validation. Two methods for bone lesion segmentation are compared. First, only lesion annotations were used as ground truth with PET and CT images as 2-channel input for the training ($U\text{-}Net_L$). A second network was trained with both the reference bone and lesion masks as ground truth ($U\text{-}Net_{BL}$).

C. Evaluation

1) Bone segmentation: Bone segmentation was evaluated i) with a side-by-side visual comparison with the traditional automatic bone segmentation method (see Fig. 3) and ii) with the computation of a mean Dice score over all the cases.

2) Bone lesion segmentation: Detection as well as segmentation metrics have been computed since both tasks are closely linked. For some patients who do not present any bone lesion, the Dice score per case was not evaluated.

a) Detection metrics: A lesion is considered detected when the overlap between ground truth and its prediction reaches a certain threshold, fixed to 50% hereafter.

While lesions are generally easy to separate and match, bone lesions are often connected and hard to isolate which complicates the recall and precision computation. Thus, a connected components algorithm was employed to separate the lesions of the reference mask, which were then overlapped with the predicted binary mask as shown on Fig. 2.

This allows to count the true positives (TP), the false positives (FP), false negatives (FN) and to compute the detection recall ($\frac{TP}{TP+FN}$) and precision ($\frac{TP}{TP+FP}$).

b) Segmentation metrics: Two metrics based on the Dice score, which evaluates the degree of overlap between a reference mask and a predicted one, were used.

First, the mean Dice score per case was computed by averaging the Dice score obtained individually for the patients presenting bone lesions. In the second case, a global measurement of the dice was performed combining all cases

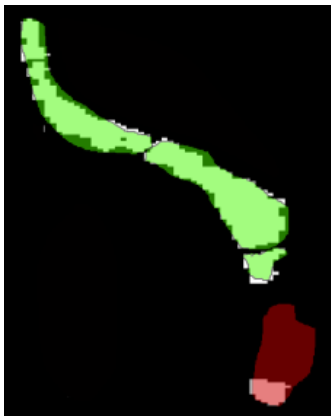


Fig. 2. Illustration of the true positives, false positives and false negatives computation on the spine. First, a connected components method is applied on the ground truth (GT) binary mask (white) to isolate bone lesions. Then, a lesion is considered to be correctly detected if the overlap between the predicted mask and and the GT is $> 50\%$ (green). Otherwise, it is considered as an undetected lesion (red).

in a single one. The global Dice score tends to be more affected by large lesions compared to small ones while the Dice score per case penalizes prediction error in cases with fewer lesions.

3) PET bone index: Using the automatically computed bone and bone lesion segmentation, we calculated the PET bone index as the ratio of the bone lesion volume compared to the total volume of all bones.

IV. RESULTS

The networks converged in approximately 500 epochs and it took between 4 and 5 days to train each fold depending on the GPU and the amount of data given as input.

A. Bone segmentation

Bone segmentation achieves a mean Dice score of 0.94 ± 0.03. Fig. 3 shows side-by-side comparison with the traditional automatic bone segmentation method for one patient.

B. Bone lesion segmentation and detection

The results obtained by the 2 different networks are given in Table I. The best score in terms of precision and Dice score was achieved by the *U-Net*$_{BL}$ network. Fig. 4 shows some visual examples of lesion segmentation for each network.

TABLE I

RESULTS OF THE BONE LESION SEGMENTATION. *U-Net*$_L$ IS THE NETWORK WITH PET AND CT AS INPUT AND ONLY BONE LESIONS AS GROUND TRUTH. *U-Net*$_{BL}$ IS THE NETWORK WITH PET AND CT AS INPUT AND BONE MASK WITH BONE LESIONS AS GROUND TRUTH.

Methods	**Recall det.**	**Preci. det.**	**Mean Dice**	**Global Dice**
U-Net$_L$	0.67	0.82	0.58 ± 0.17	0.59
U-Net$_{BL}$	0.67	0.88	0.61 ± 0.16	0.61

C. PET bone index

Per patient PET bone indexes are given in Fig. 5.

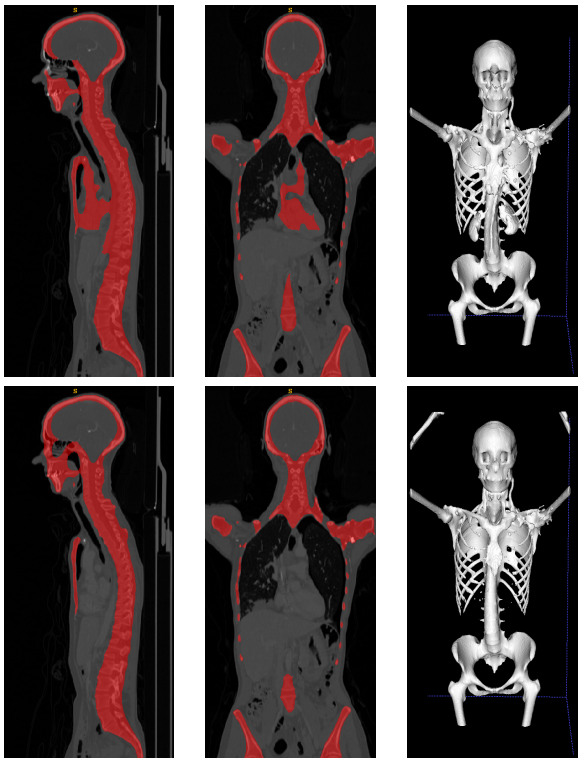


Fig. 3. Example of bone segmentation visual result. **Top**: traditional morphological and thresholding procedures inspired by [16]. **Bottom**: *U-Net*$_{BL}$ trained with both the reference bones and lesions masks as ground truth. Unlike the *U-Net*$_{BL}$, the traditional approach fails to dissociate active organs (top-left/middle: heart and top-right: kidneys) from bones.

V. DISCUSSION AND CONCLUSION

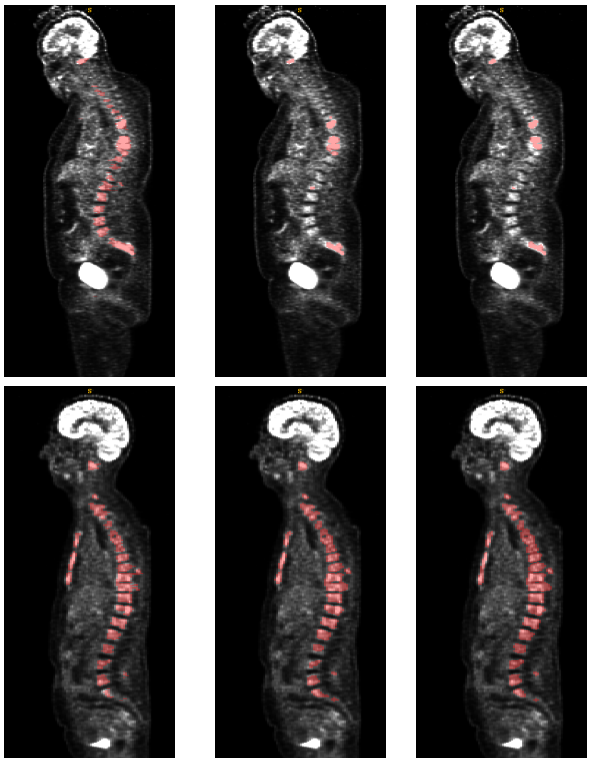
We presented an automatic PET/CT bones and metastatic bone lesion segmentation approach based on deep learning in the context of metastatic breast cancer, applied to 24 patients.

The bone segmentation algorithm achieves good results, without presenting large false positive regions, contrary to threshold-based automatic methods. Moreover, the use of the bone masks as ground truth during the training phase slightly improves the results of the automatic bone lesion segmentation in terms of Dice score, but even more in terms of precision: the network is constrained to look for lesions in the bones.

The segmentation results of the bone lesions are perfectible. Indeed, the large PET SUV heterogeneity found in these lesions tend to lower the bone lesion Dice score in our experiments, low fixing lesions being sometimes ignored by the presented U-Net architecture as shown on the middle row of Fig. 4.

The automatic PET bone index (PBI) shows relatively good agreement with the ground truth measurements, except for a few cases (see Fig. 5). These disagreements are also due to the lesions ^{18}FDG fixation differences.

Nevertheless, there is a real potential to improve the results. First, we already started working on a way to integrate the bone masks in the deep learning network more efficiently. Then, several paths would be explored to address the lesions with low ^{18}FDG fixation.



(a) Ground Truth (b) *U-Net*$_L$ (c) *U-Net*$_{BL}$

Fig. 4. Bone lesion segmentation visual results for 3 patients. (a) The ground truth (b) *U-Net*$_L$ trained only with the lesion masks, (c) *U-Net*$_{BL}$ trained with both the reference bones and lesions masks. **Top**: patient 15 with bad segmentation results (Dice score < 0.5 for both methods) due to the low ^{18}FDG fixation. The mean SUV difference between detected and undetected lesions is 2.7. **Bottom**: patient 14 presenting a lot of bone lesions with good segmentation results (Dice score > 0.76 for both methods).

Finally, we will be soon able to evaluate our methods on a more significant set of data of the ongoing *EPICURE*$_{seinmeta}$ study, and to measure the predictive power of the PBI.

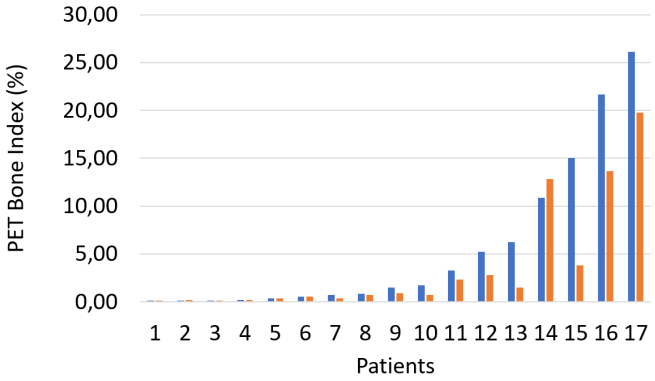


Fig. 5. Comparison of ground truth (blue) vs automatic (orange) PET Bone index (PBI) per patient sorted according to their ground truth PBI. A good agreement is achieved except for a few cases like patient 15 illustrated on the middle row of Fig. 4.

References

[1] Carol E. DeSantis et al. "Breast cancer statistics, 2019". In: *CA: A Cancer Journal for Clinicians* 69.6 (2019), pp. 438–451.

[2] Sang Kyu Yang, Nariya Cho, and Woo Kyung Moon. "The role of PET/CT for evaluating breast cancer". In: *Korean journal of radiology* 8.5 (2007), pp. 429–437.

[3] Brent Foster et al. "A review on segmentation of positron emission tomography images". In: *Computers in Biology and Medicine* 50 (2014), pp. 76–96.

[4] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42 (2017), pp. 60–88.

[5] R. Coleman and R. Rubens. "The clinical course of bone metastases from breast cancer". In: *British Journal of Cancer* 55 (1987), pp. 61–66.

[6] Hakan Demirci et al. "Uveal metastasis from breast cancer in 264 patients". In: *American Journal of Ophthalmology* 136.2 (2003), pp. 264–271.

[7] M. Imbriaco et al. "A new parameter for measuring metastatic bone involvement by prostate cancer: The bone scan index". In: *Clinical Cancer Research* 4.7 (1998), pp. 1765–1772.

[8] Ai Idota et al. "Bone Scan Index predicts skeletal-related events in patients with metastatic breast cancer". In: *SpringerPlus* (2016).

[9] M. Colombié et al. "Évaluation d'une méthode de quantification de la masse métastatique osseuse par mesure automatisée du Bone Scan Index, dans le suivi thérapeutique des cancers du sein". In: *Médecine Nucléaire* 4101.6 (2013), pp. 233–273.

[10] Chun K. Kim et al. "Standardized Uptake Values of FDG: Body Surface Area Correction is Preferable to Body Weight Correction". In: *Journal of Nuclear Medicine* (1994).

[11] Fabian Isensee et al. "Automated Design of Deep Learning Methods for Biomedical Image Segmentation". In: *arXiv preprint* arXiv:1904.08128 (2020).

[12] Fabian Isensee et al. "nnU-Net: Breaking the Spell on Successful Medical Image Segmentation". In: *arXiv preprint* arXiv:1904.08128 (2019).

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 9351. Springer (2015).

[14] Nicholas Heller et al. "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 Challenge". In: *arXiv preprint* (2019).

[15] *Keosys Medical Imaging Viewer*. `https://www.keosys.com/read-system/`. Accessed: 2020-01-28.

[16] S. Banik, R. Rangayyan, and G. Boag. *Landmarking and Segmentation of 3D CT Images*. Morgan & Claypool, 2009.