Influence of inputs for bone lesion segmentation in longitudinal 18 F-FDG PET/CT imaging studies

Noemie Moreau, Caroline Rousseau, Constance Fourcade, Gianmarco Santini, Ludovic Ferrer, Marie Lacombe, Camille Guillerminet, Mathilde Colombie, Pascal Jezequel, Mario Campone, et al.

2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871081

Influence of inputs for bone lesion segmentation in longitudinal ^{18}F-FDG PET/CT imaging studies

Noémie Moreau[1,2], Caroline Rousseau[3,5], Constance Fourcade[1,2], Gianmarco Santini[2], Ludovic Ferrer[4,5], Marie Lacombe[5], Camille Guillerminet[5], Mathilde Colombié [5], Pascal Jézéquel [4,5], Mario Campone [4,5], Mathieu Rubeaux[2] and Nicolas Normand[1]

***Abstract*— In metastatic breast cancer, bone metastases are prevalent and associated with multiple complications. Assessing their response to treatment is therefore crucial.**

Most deep learning methods segment or detect lesions on a single acquisition while only a few focus on longitudinal studies. In this work, 45 patients with baseline (BL) and follow-up (FU) images recruited in the context of the $EPICURE_{seinmeta}$ study were analyzed. The aim was to determine if a network trained for a particular timepoint can generalize well to another one, and to explore different improvement strategies. Four networks based on the same 3D U-Net framework to segment bone lesions on BL and FU images were trained with different strategies and compared. These four networks were trained 1) only with BL images 2) only with FU images 3) with both BL and FU images 4) only with FU images but with BL images and bone lesion segmentations registered as input channels. With the obtained segmentations, we computed the PET Bone Index (PBI) which assesses the bone metastases burden of patients and we analyzed its potential for treatment response evaluation.

Dice scores of 0.53, 0.55, 0.59 and 0.62 were respectively obtained on FU acquisitions. The under-performance of the first and third networks may be explained by the lower SUV uptake due to treatment response in FU images compared to BL images. The fourth network gives better results than the second network showing that the addition of BL PET images and bone lesion segmentations as prior knowledge has its importance. With an AUC of 0.86, the difference of PBI between two acquisitions could be used to assess treatment response.

***Clinical relevance*— To assess the response to treatment of bone metastases, it is crucial to detect and segment them on several acquisitions from a same patient. We proposed a completely automatic method to detect and segment these metastases on longitudinal ^{18}F-FDG PET/CT images in the context of metastatic breast cancer. We also proposed an automatic PBI to quantitatively assess the evolution of the bone metastases burden of patient and to automatically evaluate their response to treatment.**

I. INTRODUCTION

In metastatic breast cancer, bone metastases are prevalent and associated with multiple complications [1]. Assessing their response to treatment is therefore crucial but if they can be detected on several modalities, they are difficult to segment and therefore to quantitatively assess.

Recently, numerous automatic methods to segment or detect lesions on CT, MRI or PET have been developed [2] but only a few focus on segmenting lesions in longitudinal imaging studies. Lesion segmentation on longitudinal images is important to quantitatively monitor the progression of a disease or assess response to treatment. One of its applications is for Multiple Sclerosis lesions monitoring on brain MR scans and several methods have been proposed to automatically segment new or enlarged (NE) lesions. Denner et al. [3] proposed a U-Net with multiple input channels (one for each acquisition) and multi-task learning with deformable registration between baseline and follow-up to guide the segmentation. Sepahvand, Arnold, and Arbel [4] explored the interest of substraction image between baseline and follow-up as input of U-Net to detect NE lesions. However, to the best of our knowledge, no work to date focuses on longitudinal PET images and bone metastases.

In this study, we investigate if a network trained for a particular timepoint *i.e.* baseline (BL) can generalize well to another one *i.e.* follow-up (FU) and we explore different improvement strategies to segment bone lesions on FU images. A first network trained on BL images was tested on FU images. A second network was trained specifically on FU images and a third on all images (BL and FU). The last network was trained specifically on FU images with all the information available for training *i.e.* the BL PET and bone lesion segmentations.

Finally, the obtained segmentations were used to compute the PET Bone Index (PBI) [5] on each patients' acquisition to assess the evolution of their bone metastases burden.

II. DATA AND METHODS

A. Data

Forty-five patients with BL and FU PET/CT images were recruited in the context of the $EPICURE_{seinmeta}$ metastatic breast cancer study (NCT03958136). 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).

Imaging data were acquired at two sites using different imaging systems. BL imaging was acquired before initiation of a new treatment and FU images was acquired mid- or/and

*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]University of Angers, CRCINA, INSERM UMR1232, CNRS-ERL6001, Angers, France
[5]ICO Cancer Center, Nantes - Angers, France

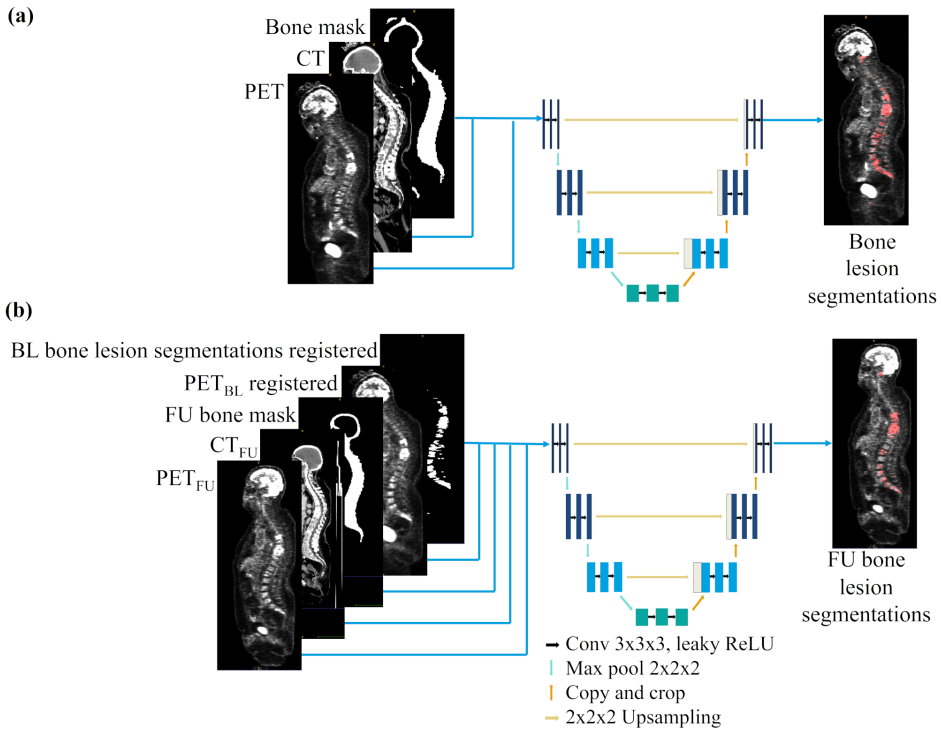


Fig. 1. (a) Segmentation network (*U-Net*$_{BL}$, *U-Net*$_{FU}$,*U-Net*$_{BL\&FU}$) and (b) *U-Net*$_{FU/BL}$

post-treatment. Bone lesions were manually segmented by medical experts. Treatment response at each follow-up time point was evaluated according to the PERCIST criteria by two experts (one from each site).

Both PET and CT were used as most metastatic lesions only appear on PET but anatomical references are mostly present on CT. Moreover, Xu et al. [6] demonstrated that using PET and CT as input channels improves the segmentation compared to CT or PET alone.

B. Bone lesion segmentation

All networks investigated were based on the same 3D U-Net framework called "no new net" (nnU-Net), which achieved state of the art performances [7]. This framework automatically sets a number of hyper-parameters given several information, such as input data features and memory consumption. The original nnU-Net publication [7] gives a detailed list of the configurable parameters.

Networks were trained and validated using 3-fold cross-validation balanced regarding institution, extent of bone metastasis and number of follow-up images.

PET images were converted to SUV_{BW} to normalize the lesion activity to the injected activity and body weight. All images were resampled and normalized by the nnU-Net [7].

Bone segmentations were used during training as they improve bone lesion segmentation results [5]. They were automatically generated using a network presented in [8].

Figure 1 presents the different networks used in this work.

Baseline network (*U-Net*$_{BL}$) For this network, only BL images were used for training. The trained network was then validated on BL images and FU images. The network had 3 inputs channels for PET, CT images and bone mask.

Follow-up network (*U-Net*$_{FU}$) This network was the same as the baseline network, except training was done with FU images instead of BL images. It was then validated on both BL and FU images.

Baseline and follow-up network (*U-Net*$_{BL\&FU}$) This network was the same as the previous two networks but this time both BL and FU images were used for training. The same 3 input channels were used (*i.e.* PET, CT and bone mask). The network was validated on both BL and FU images.

Follow-up network with baseline prior knowledge (*U-Net*$_{FU/BL}$) For this network, only FU images were used for training. Compared to the second network, additional inputs were provided. When performing manual segmentation or assessing treatment response on FU images, experts usually look at both BL and FU images on viewers [9] to determine patients response. Therefore, two new input channels were added: one for the BL PET images and one for BL bone lesion segmentation. Following Krüger et al. [10] and Sepahvand, Arnold, and Arbel [4], baseline and follow-up PETs were rigidly registered using the ANTs pipeline with recommended settings [11]. The registration transformation was then applied to the BL segmentation images. This network therefore had five input channels : the PET, CT, bone mask of FU and the PET, bone lesion mask of BL and was validated on FU images only.

C. PET Bone Index

Inspired by the Bone Scan Index (BSI)[12], we previously developed a method to assess the bone metastases burden of

TABLE I

DICE EVALUATION FOR THE FOUR NETWORKS ON BASELINE AND FOLLOW-UP ACQUISITIONS

Network	*U-Net*$_{BL}$	*U-Net*$_{FU}$	*U-Net*$_{BL\&FU}$	*U-Net*$_{FU/BL}$
Baseline	$\mathbf{0.67 \pm 0.17}$	0.60 ± 0.22	0.65 ± 0.19	$\times$
Follow-up	0.53 ± 0.25	0.55 ± 0.21	0.59 ± 0.21	$\mathbf{0.62 \pm 0.19}$

patients [5]. Using the automatic bone lesion segmentation and a bone mask generated by a network presented in [8], we computed the PBI as the ratio of the bone lesion volume compared to the total bone volume.

PBI changes between baseline and follow-up images were analyzed to assess treatment response. PBI difference between two acquisitions was measured as:

$$\Delta PBI(\%) = \frac{(PBI_{FU} - PBI_{BL})}{PBI_{BL}} \times 100$$

with PBI_{BL}, PBI taken on the baseline acquisition, and PBI_{FU}, PBI taken on the follow-up acquisition.

We binarized the PERCIST responses of each follow-up as responders for patients with Complete Response and Partial Response and non-responders for the other patients. To evaluate the PBI's potential to assess treatment response, we used a Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (AUC) for a binary prediction of treatment response. The optimal cutoff value to differentiate responders and non responders patients was determined using the Youden's J statistic method [13].

III. RESULTS

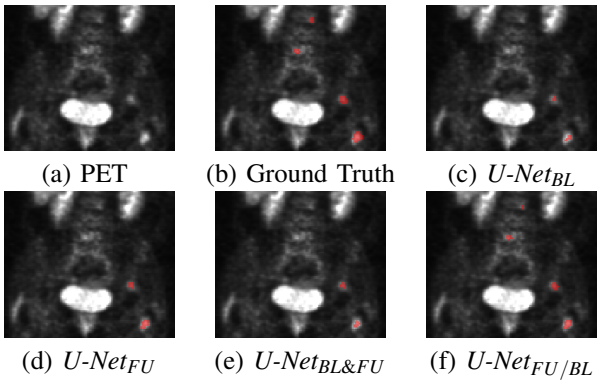
A. Bone lesion segmentation

The Dice score was used to evaluate the degree of overlap between the ground truth and the prediction. As 3-fold cross-validation was used, for each network we had 3 models trained on different parts of the dataset and validation performances were computed for each model and then averaged. To compare the results between networks, the same patients' fold were used to train and validate every network. For networks trained on a single acquisition, validation on the other acquisitions was done with the same patient's split, to avoid using on a patient's acquisition, a model trained on another acquisition of the same patient. This way, we prevented biased results.

Results are presented in Table I, the *U-Net*$_{FU/BL}$ network performed better than any other networks on FU images. Figure 2 shows segmentation results on FU images from one patient with the different networks and Figure 3 shows segmentation results on different acquisitions from the same patient obtained with the two best networks for each acquisition.

B. PET Bone Index

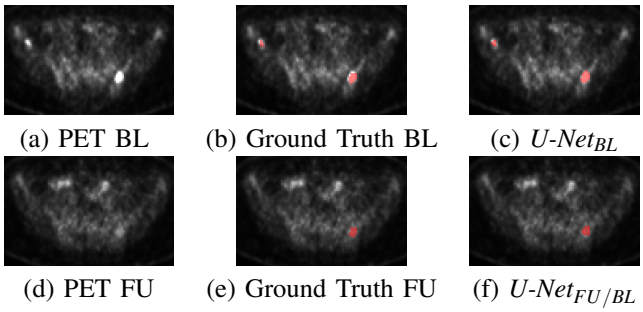
According to the ROC analysis, the PBI obtained an AUC of 0.86 for a binary prediction of treatment response (responders vs non-responders). The optimal cutoff value to classify patients was -11% (sensitivity 75%, specificity



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

(d) *U-Net*$_{FU}$ (e) *U-Net*$_{BL\&FU}$ (f) *U-Net*$_{FU/BL}$

Fig. 2. Segmentation examples on FU images from one patient with multiple lesions with the different networks. Due to the patient's response to treatment, some lesions have low contrast and are more difficult to segment. Zoom on the abdomen: kidneys, spine and bladder are visible.



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

(d) PET FU (e) Ground Truth FU (f) *U-Net*$_{FU/BL}$

Fig. 3. Segmentation examples on different acquisitions from the same patient. Zoom on the pelvic region.

85%) meaning that a patient with a $\Delta PBI(\%) \leq -11\%$ is considered responder (lesion volume diminution) while a patient with a $\Delta PBI(\%) > -11\%$ is considered non-responder (lesion volume augmentation or stability). Figure 4 shows an exemple of the PBI automatically computed on two acquisitions of the same patient. This patient had a PBI augmentation of 244% and, according to the optimal cutoff, is non-responder which is in agreement with the PERCIST evaluation.

IV. DISCUSSION AND CONCLUSION

This work compared different methods to segment bone lesions on follow-up images from patients with metastatic breast cancer.

As presented in Table I, the *U-Net*$_{FU/BL}$ network performed better than any other networks on FU images. It is interesting to notice the poor results given by the *U-Net*$_{BL\&FU}$ network and the failure of the *U-Net*$_{BL}$ network to generalize on FU images, which may be explained by the lower SUV uptake due to treatment response in FU images. As shown in Figure 2, these two networks miss two small lesions with low SUV uptake and only segment the high SUV uptake part of the two other lesions. Moreover, Figure 3 shows the difference in SUV uptake between the Baseline and Follow-up acquisitions due to treatment response which complexify the segmentation task. The *U-Net*$_{FU/BL}$ network gave better results than the *U-Net*$_{FU}$ showing the importance of BL PET and segmentation as prior knowledge. At the

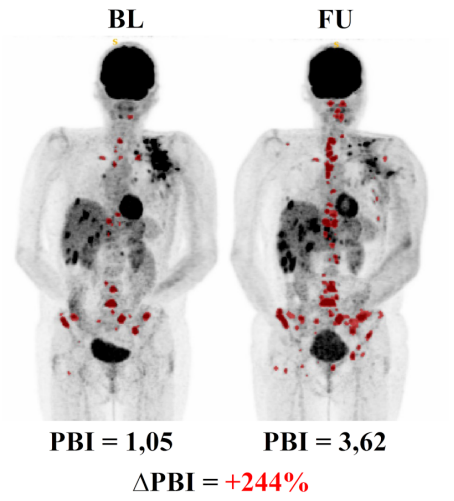


Fig. 4. Automatic segmentations overlaid on the maximum intensity projections of PET images and PBI assessments for one patient with Progressive Disease ($\Delta PBI = +244\%$). On the left the baseline acquisition and on the right the follow-up acquisition. BL for baseline and FU for follow-up.

end, the *U-Net*$_{FU/BL}$ network achieves a dice score of 0.62 which is a good result given the difficulty of the task due to lesions number, size and low contrast. Another difficulty and limitation, in this work, was the lack of labelled data. It was partially corrected by using data-augmentation during training but we can assume that using a larger dataset would lead to improve the performances.

With the obtained segmentations, we computed the PBI which assess the bone metastases burden of patients and we analyzed its potential for treatment response evaluation. With an AUC of 0.86, ΔPBI could be used to assesses treatment response as done for the patient in Figure 4. Moreover, the PBI, as the BSI for prostate cancer [12] could be used to predict Progression Free Survival (PFS) and Overall Survival (OS).

Our work showed that a network trained only on one acquisition can fail to segment lesions on a new acquisition due to intensity differences between acquisitions especially on PET where treatment response influences the SUV uptake of lesions. To improve longitudinal segmentation, we proposed a new network that use the PET and segmentation of a previous acquisition to segment a new one as usually done by medical experts. This method proved to be the most effective compared to other methods and gave comparable results with the network used to segment baseline acquisitions. Moreover, we showed that the PBI could be used to automatically assess the treatment response of patients with bone metastases.

REFERENCES

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

[2] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42 (2017), pp. 60–88. DOI: 10.1016/j.media.2017.07.005.

[3] S. Denner et al. *Spatio-temporal Learning from Longitudinal Data for Multiple Sclerosis Lesion Segmentation*. 2020. arXiv: 2004.03675 [eess.IV].

[4] N. M. Sepahvand, D. L. Arnold, and T. Arbel. "CNN Detection of New and Enlarging Multiple Sclerosis Lesions from Longitudinal Mri Using Subtraction Images". In: 2020, pp. 127–130. DOI: 10.1109/ISBI45749.2020.9098554.

[5] Noémie Moreau et al. "Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer". In: *Engineering in Medicine and Biology Conference* (2020). DOI: 10.1109/EMBC44109.2020.9175904.

[6] Lina Xu et al. "Automated Whole-Body Bone Lesion Detection for Multiple Myeloma on 68Ga-Pentixafor PET/CT Imaging Using Deep Learning Methods". In: *Contrast Media and Molecular Imaging* (2018). DOI: 10.1155/2018/2391925.

[7] Fabian Isensee et al. "Automated Design of Deep Learning Methods for Biomedical Image Segmentation". In: *arXiv preprint* arXiv:1904.08128 (2020). DOI: 10.1038/s41592-020-01008-z.

[8] Noémie Moreau et al. "Comparison between threshold-based and deep learning-based bone segmentation on whole-body CT images". In: *Medical Imaging 2021: Computer-Aided Diagnosis*. Vol. 11597. International Society for Optics and Photonics. SPIE, 2021, pp. 661–667. DOI: 10.1117/12.2580892.

[9] *Keosys Medical Imaging Viewer*. https://www.keosys.com/read-system/. Accessed: 2020-08-18.

[10] J. Krüger et al. "Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks". In: *NeuroImage: Clinical* 28 (2020), p. 102445. DOI: 10.1016/j.nicl.2020.102445.

[11] Brian B. Avants, Nick Tustison, Gang Song, et al. "Advanced normalization tools (ANTS)". In: *Insight j* 2.365 (2009), pp. 1–35. DOI: https://doi.org/10.1007/978-3-319-14678-2_1.

[12] 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.

[13] William J Youden. "Index for rating diagnostic tests". In: *Cancer* 3.1 (1950). Publisher: Wiley Online Library, pp. 32–35. DOI: https://doi.org/10.1002/1097-0142.