Combining Superpixels and Deep Learning Approaches to Segment Active Organs in Metastatic Breast Cancer PET Images *

Constance Fourcade[1,2]   Ludovic Ferrer[3,4] *PhD*   Gianmarco Santini[2] *PhD*

Noémie Moreau[2]   Caroline Rousseau[3,4] *MD PhD*   Marie Lacombe[4] *MD*

Camille Guillerminet[4] PhD   Mathilde Colombié [3,4] *MD*   Mario Campone[3,4] *MD, PhD*

Diana Mateus[1] *PhD*   Mathieu Rubeaux[2] *PhD*

Abstract— Semi-automatic measurements are performed on [18]FDG PET-CT images to monitor the evolution of metastatic sites in the clinical follow-up of metastatic breast cancer patients. Apart from being time-consuming and prone to subjective approximation, semi-automatic tools cannot make the difference between cancerous regions and active organs, presenting a high [18]FDG uptake.

In this work, we combine a deep learning-based approach with a superpixel segmentation method to segment the main active organs (brain, heart, bladder) from full-body PET images. In particular, we integrate a superpixel SLIC algorithm at different levels of a convolutional network. Results are compared with a deep learning segmentation network alone. The methods are cross-validated on full-body PET images of 36 patients and tested on the acquisitions of 24 patients from a different study center, in the context of the ongoing EPICURE_seinmeta study. The similarity between the manually defined organ masks and the results is evaluated with the Dice score. Moreover, the amount of false positives is evaluated through the positive predictive value (PPV).

According to the computed Dice scores, all approaches allow to accurately segment the target organs. However, the networks integrating superpixels are better suited to transfer knowledge across datasets acquired on multiple sites (domain adaptation) and are less likely to segment structures outside of the target organs, according to the PPV.

Hence, combining deep learning with superpixels allows to segment organs presenting a high [18]FDG uptake on PET images without selecting cancerous lesion, and thus improves the precision of the semi-automatic tools monitoring the evolution of breast cancer metastasis.

Clinical relevance— We demonstrate the utility of combining deep learning and superpixel segmentation methods to accurately find the contours of active organs from metastatic breast cancer images, to different dataset distributions.

I. INTRODUCTION

Cancerous lesions in metastatic breast cancer are monitored by whole-body [18]FluoroDeoxyGlucose ([18]FDG) Positron Emission Tomography (PET) and Computed Tomography (CT). In clinical practice, apart from the conventional visual inspection of the lesions, semi-automatic threshold-based measurements are employed [1]. To avoid time-consuming interactions, automatic lesion segmentation algorithms on PET images have been developed [2], [3]. However, these approaches mostly focus on specific body parts, as the lungs or the brain.

One of the challenges of automatic whole-body lesion segmentation on PET images is that areas of abnormal [18]FDG uptake can be obscured by active sites, presenting normal physiological uptake (brain, heart) and excretion (bladder) of the radioactive marker. Most methods proposing to localize and segment these organs use anatomical imaging techniques, such as Magnetic Resonance Imaging (MRI), CT [4], [5], or combine them with PET [6], [7]. However, methods based solely on PET images are less common: while PET imaging is recognized for metastatic activity characterization, it is considered lacking the spatial resolution required for precise organ segmentation. Hence, segmenting directly on PET images is a challenge but removes the dependence on error prone PET/CT registration.

Unlike the methods developed in [4], [8], we did not create bounding boxes around the target organs. Though more challenging, we seek a pixel-wise segmentation of the organ in order to preserve lesions close to the organs of interest for subsequent analysis. Moreover, our approach represents one of the first attempts to perform active organ segmentation using a 3D convolutional neural network (CNN).

The main limitation of the application of deep learning-based methods to medical imaging problems is the lack of expertly annotated data to properly train networks [9]. Thus, designing networks adaptable from an image domain to another is an active research field [10], [11]. As in [12], we rely on superpixel information to favor the domain adaptation from images of a distribution to a different one (e.g. acquired on different sites or with different machines). Since lesions often appear smaller than the normal active organs in the context of metastatic breast cancer, superpixels summarize the important anatomical information while blurring away other changes in image intensities.

In this paper, we propose two approaches to segment PET active organs. They both combine a classical deep segmentation network with a superpixel algorithm, but integrates it at different levels of the CNN: i) in the input images and ii) in the loss function. The deep segmentation network alone is used as a reference.

Since the outcome of our networks facilitates the posterior segmentation of lesions, it is essential that resulting segmentation maps do not contain any tumor located out of the target organs. Hence, the evaluation is two-fold: a) the Dice metric compares the results to manual annotations of the normal uptaking sites, and b) the positive predictive value (PPV) highlights the presence of false positive detections.

The approaches integrating superpixel information outperform the one based solely on a 3D CNN, because their introduction within a deep learning pipeline compels the network to focus on voxels presenting *normal* high ^{18}FDG uptake. To the best of our knowledge, this work is the first to perform automatic segmentation of high ^{18}FDG uptake on PET images and to take advantage of superpixel information to enforce networks to focus on the target objects.

II. METHODS

Given a full body ^{18}FDG PET image, we aim to automatically determine which image voxels belong to the active organs (i.e. brain, heart and bladder). The selection of these sites is challenging due to the surrounding tumors presenting a similar intensity, and to the variable tracer uptake amounts across the acquisitions and patients. In this section, we describe the baseline 3D U-Net method and propose two variants including superpixel information.

To segment the target organs, *U-Net* relies on the 3D U-Net branch of the nnU-Net pipeline [13]. This network will be used as a baseline. The objects of interest occupying a small proportion of the full image, the training loss is a balanced combination of a cross-entropy and a Dice loss:

$$\mathcal{L}_{Total} = \mathcal{L}_{CE} + \mathcal{L}_{Dice}, \tag{1}$$

with $\mathcal{L}_{CE}$ as in [14] and:

$$\mathcal{L}_{Dice} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_{i \in I} u_i^k v_i^k}{\sum_{i \in I} u_i^k + \sum_{i \in I} v_i^k} \tag{2}$$

as in [13], having u the output map of the network, v the ground truth segmentation, $i \in I$ the number of pixels in the training batch and $k \in K$ the classes. This loss reduces influence of class imbalance [15].

To reduce the risk of confusing another structure, especially a tumor, with a target organ, we propose to incorporate superpixel information at different levels of the network: as input (*U-Net-SP-Input*) and in the loss function (*U-Net-SP-Loss*). Moreover, superpixels spontaneously tend to smooth differences between images. Hence, using them in the training step of a CNN will increase the adaptability of the network from a dataset distribution to a similar yet different one (domain adaptation).

For the **first approach**, *U-Net-SP-Input*, we trained the reference network *U-Net* on superpixelized images (Fig. 1, right). These superpixelized images were created using the reference Simple Linear Iterative Clustering (SLIC) algorithm [16] in 3D on the whole-body PET images. This method clusters voxels based on their grey level similarity
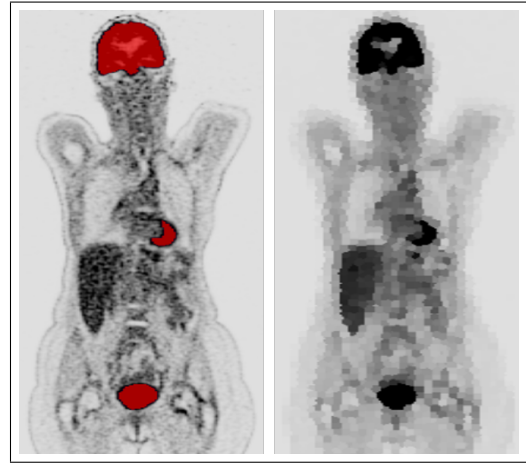

Fig. 1. Left: PET image and ground truth maps of the target organs (red). Right: Superpixelized PET image. Best viewed in color.

and spatial proximity in the image, minimizing the following distance with a k-means algorithm:

$$D = d_{color} + \frac{m}{S} d_{image} \tag{3}$$

where D is the weighted sum of two pixel pairwise Euclidean distances, d_{color} for the color-space and d_{image} for the coordinate-space. S represents the approximate size of the superpixels, while m is introduced to control their compactness. Choosing the appropriate superpixel size and compactness parameters is important to emphasize the normal ^{18}FDG uptake on organs while blurring small lesions and objects.

In a **second approach**, *U-Net-SP-Loss*, we introduced superpixel information in the optimization of the reference network. The global loss is modified by balancing the cross-entropy loss with a superpixel-inspired regularization term:

$$\mathcal{L}_{Total_{SP}} = \mathcal{L}_{CE} + \mathcal{L}_{SP}, \tag{4}$$

where:

$$\mathcal{L}_{SP}(u, C) = \text{MSE}(\arg\max(u), \Phi(\arg\max(u), C)), \tag{5}$$

with MSE the classical mean square error computation, u the softmax output of the network and C the correspondence map between pixels and superpixels. Φ computes superpixel corrected segmentation maps, forcing all pixels within a superpixel to belong to the same class.

The regularization term, inspired by [17], enforces voxels belonging to the same superpixel to be associated with the same semantic object. Superpixels enhancing patterns in images, in the application at hand, they help networks to learn weights more generalizable to different datasets.

III. EXPERIMENTAL VALIDATION

A. Dataset description

In the context of the ongoing EPICURE$_{seinmeta}$ study (NCT03958136), 16 patients were initially recruited from the Integrated Center for Oncology (ICO) of Nantes (*NICO*), followed by 24 patients from the ICO of Angers (*AICO*). All patients underwent PET/CT imaging for the diagnosis

Center	Number of acquisitions	Age	Imaging systems
NICO	36	60 ± 13	Siemens Biograph mCT 40 Siemens Biograph mCT 64
AICO	24	58 ± 13	Imageurs Philips Vereos GE Discovery 690 GE Discovery IQ

of metastatic breast cancer. From the *NICO*, 36 PET/CT were used, corresponding to one to three acquisitions per patient. PET/CT images were obtained using different imaging systems, depending on the study center, resulting in intensity, resolution and texture variations between images. Two different dual-slice Siemens Biograph PET/CT were used in the *NICO*, while the images of the *AICO* were acquired either on the Philips Vereos, GE Discovery 690 or GE Discovery IQ PET/CT imaging systems (see Table I).

This prognostic 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 obtained.

B. Implementation details

All processed PET images where normalized by the injected radioactivity's concentration and the patient's body weight, giving the SUV according to the standard developed in [18].

The SLIC [16] algorithm was run in 3D, and the free parameters of the method, i.e. the superpixel size and the compactness, were set according to preliminary grid search results to 12 mm $\times$ 12 mm $\times$ 10 mm and 5, respectively, to enhance target organs and to blur small lesions.

We followed the recommendations from [13] regarding the preprocessing steps of the network, namely, cropping the input images to nonzero values, resampling them to the median voxel spacing of the dataset and applying a z-score normalization on each image. This helps to reduce the discrepancy between data domains. Considering the network itself, the architecture is similar to the benchmark U-Net [14] for 3D images, but uses leaky ReLU activation and instance normalization. All deep learning architectures were trained from scratch on a NVIDIA GeForce GTX 1080 Ti GPU. The input patch size ($128 \times 128 \times 128$) and the number of pooling operations (4 in each branch) were computed according to the preprocessed image size. During training, the Adam optimizer was used. An epoch corresponded to an iteration over 250 training batches, each composed of two images. The learning rate was initialized at 3.10^{-4} and reduced by a factor 5 if the loss did not improve by at least 5.10^{-3} within the last 30 epochs. Once the learning rate reached 10^{-6}, the training stopped if the loss did not get better by at least 5.10^{-3} within the last 50 epochs. Random rotations, scaling, elastic deformations, gamma corrections and mirroring were performed on the fly for data augmentation purposes.
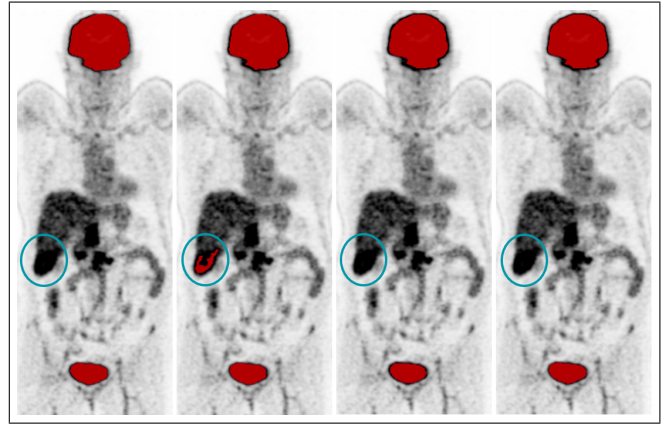

Fig. 2. PET image of a patient from the *AICO* dataset with overlaid masks corresponding to ground truth, *U-Net*, *U-Net-SP-Input* and *U-Net-SP-Loss* segmentations from left to right respectively. Blue circle marks a lesion erroneously segmented by the *U-Net* network. Best viewed in color.

C. Evaluation metrics

The resulting segmentation maps of the three approaches were compared to ground truth masks containing the target organs (Fig. 2, left), manually delineated with the Keosys Viewer [19] on natural PET images.

Results were first assessed computing the Dice metric between the segmented and reference rasks, considering all three organs of interest together. The CNNs ability to segment target organs rather than other anatomical structures, i.e. to avoid false positive results, was assessed using the PPV. Indeed, this metric points out the ratio of true positive (TP) pixels among all the positive predictions, characterised here by the sum of TP and false positive (FP):

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

D. Intra-domain experiment

To counteract the limited size of our dataset and maintain a reasonable computational time, all deep learning segmentation networks were initially validated on the *NICO* dataset using a 3-fold cross validation.

In that intra-domain experiment, the three CNNs present similarly high resulting Dice scores: 0.97 ± 0.01, 0.96 ± 0.01 and 0.96 ± 0.01 for the *U-Net*, *U-Net-SP-Input* and *U-Net-SP-Loss* approaches respectively. Hence, the three approaches accurately segment the target organs.

E. Domain adaptation experiments

Over a second phase, the networks trained on the *NICO* dataset were tested on the *AICO* images to evaluate their adaptability and generalization abilities. While slightly lower in the intra-domain case, the Dice scores obtained with the three approaches still remain high, with 0.92 ± 0.07, 0.92 ± 0.05 and 0.93 ± 0.05 for the *U-Net*, *U-Net-SP-Input* and *U-Net-SP-Loss* networks respectively.

Since we aim to remove the normal high ^{18}FDG uptake organs before segmenting lesions, we evaluated the methods capability to avoid the surrounding anatomical structures.
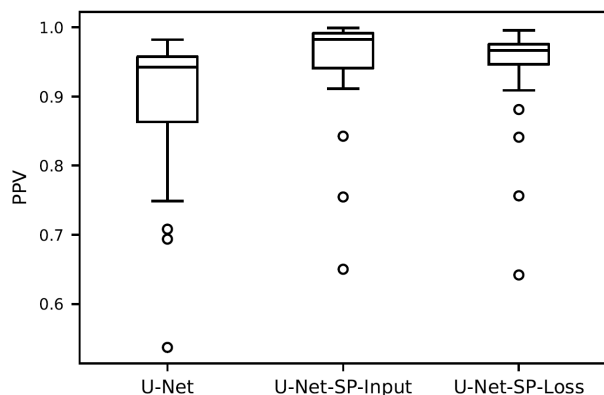
Fig. 3. Boxplot of the Positive Predictive Value (PPV) of the segmentations predicted by the *U-Net*, *U-Net-SP-Input* and *U-Net-SP-Loss* networks on the AICO dataset.

Considering first a qualitative approach, the Fig. 2 shows that networks using superpixel information are better suited to segment only the target organs. Quantitatively, we computed the PPV for the *AICO* dataset. As illustrated in Fig. 3, networks integrating superpixels information produce less FP than the reference *U-Net* network.

IV. DISCUSSION AND CONCLUSION

As shown by the obtained Dice scores, when the train and test datasets come from the same data distribution, the three approaches present similar results. In contrast, the superpixel regularization smooths variations between images acquired with different acquisition systems, improving segmentation performance on test data coming from a center unseen during training, showing the possibility of better generalization effect on unseen data. Moreover, according to the PPV measure, the integration of superpixel information in the training of a segmentation network reduces the amount of FP when using this network on unseen images.

From a computational point of view, the training time of the *U-Net-SP-Input* approach resulted to be faster (õne day) than the one of *U-Net-SP-Loss* (t̃wo days). Indeed, superpixel computation is performed only once in the preprocessing step of the former network, while the latter requires superpixel computation during each epoch. However, the integration of the superpixel computation inside the training loop of *U-Net-SP-Loss* makes the process end-to-end, and potentially leaves room to integrate an automatic superpixel hyperparameters learning.

As a conclusion, we have integrated superpixel information at several level of a deep segmentation network to segment high ^{18}FDG uptake organs on PET images. The strength of this method is the addition of superpixels, making the network focus on the organs of interest, and more prone to segment data that has not been seen by the network.

In a future work, we will evaluate the performances of both superpixel-based approaches on a more significant dataset. We will consider using the most promising network in a lesion segmentation pipeline to reduce the perturbations caused by normal active organs on PET images. Our method is an important first step towards a lesion segmentation algorithm in the context of metastatic breast cancer, in order to avoid the confusion between high ^{18}FDG uptake organs and lesions, especially when they are contiguous.

REFERENCES

[1] D. Goulon, H. Necib, B. Henaff, C. Rousseau, T. Carlier, and F. Kraeber-Bodere, "Quantitative Evaluation of Therapeutic Response by FDG PET-CT in Metastatic Breast Cancer," *Frontiers in Medicine*, vol. 3, 2016.

[2] T. Henry, C. Meyer, V. Chevance, V. Roblot, E. Blanchet, V. Arnould, G. Grimon, M. Chekroun, F. Parent, A. Seferian, R. Jovan, S. Bulifon, X. Jais, D. Montani, M. Humbert, P. Chaumet-Riffaud, V. Lebon, E. Durand, and F. Besson, "Automated PET segmentation for lung tumors: Can deep learning reach optimized expert-based performance?," *Journal of Nuclear Medicine*, vol. 59, no. 1, pp. 322–322, 2018.

[3] P. Blanc-Durand, A. Van Der Gucht, N. Schaefer, E. Itti, and J. O. Prior, "Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study," *PLoS ONE*, vol. 13, no. 4, pp. 1–11, 2018.

[4] Bob D. de Vos, Jelmer M. Wolterink, Pim A. de Jong, Max A. Viergever, and Ivana Išgum, "2D image classification for 3D anatomy localization: employing deep convolutional neural networks," *Medical Imaging 2016: Image Processing*, vol. 9784, pp. 97841Y, 2016.

[5] X. Lu, D. Xu, and D. Liu, "Robust 3D Organ Localization with Dual Learning Architectures and Fusion Xiaoguang," *Deep Learning and Data Labeling for Medical Applications*, pp. 12–20, 2016.

[6] L. Bi, J. Kim, D. Feng, and M. Fulham, "Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies," *MICCAI*, vol. 8673 LNCS, no. PART 1, pp. 569–576, 2014.

[7] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, "Adaptive Supervoxel Patch-based Region Classification in Whole-Body PET-CT," *MICCAI*, 2015.

[8] S. Afshari, A. BenTaieb, and G. Hamarneh, "Automatic localization of normal active organs in 3D PET scans," *Computerized Medical Imaging and Graphics*, vol. 70, pp. 111–118, 2018.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis.," *Medical Image Analysis*, vol. 42, no. December 2017, pp. 60–88, 2017.

[10] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magnetic Resonance in Medicine*, vol. 80, no. 3, pp. 1189–1205, 2018.

[11] V. Cheplygina, M. de Bruijne, and J. P.W. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Medical Image Analysis*, vol. 54, pp. 280–296, may 2019.

[12] Y. Zhang, P. David, H. Foroosh, and B. Gong, "A Curriculum Domain Adaptation Approach to the Semantic Segmentation of Urban Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[13] F. Isensee, J. Petersen, S. A. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnU-Net: Breaking the Spell on Successful Medical Image Segmentation," *arXiv preprint arXiv:1904.08128*, vol. 1, pp. 1–8, 2019.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI*, vol. 9351, pp. 234–241, 2015.

[15] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Int. Conf. on 3D Vision*, pp. 565–571, 2016.

[16] R. Achanta, A. Shaji, K. Smith, and A. Lucchi, "SLIC Superpixels," *EPFL Technical Report*, vol. 149300, no. June, pp. 1–15, 2010.

[17] M. Papadomanolaki, M. Vakalopoulou, and K. Karantzalos, "A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks," *Remote Sensing*, vol. 11, no. 6, pp. 684, 2019.

[18] C. K. Kim, N. C. Gupta, B. Chandramouli, and A. Alavi, "Standardized uptake values of FDG: Body surface area correction is preferable to body weight correction," *Journal of Nuclear Medicine*, vol. 35, no. 1, pp. 164–167, 1994.

[19] Keosys Medical Imaging, "https://www.keosys.com/read-system," .