Segmentation of axillary lymph nodes in PET/CT scans: First experiments

Diana Lucia Farfan Cabrera, Nicolas Gogin, David Morland, Dimitri Papathanassiou, Nicolas Passat

Segmentation of axillary lymph nodes in PET/CT scans: First experiments

Diana Lucia Farfan Cabrera[*,**], Nicolas Gogin[**], David Morland[*,***], Dimitri Papathanassiou[*,***], Nicolas Passat[*]

[*] Université de Reims Champagne Ardenne, CReSTIC, 51097 Reims, France
[**] General Electric Healthcare, Buc, France
[***] Département de Médecine Nucléaire, Institut Godinot, Reims, France

Abstract. The analysis of axillary lymph nodes is of crucial importance for the staging of breast cancer. As a consequence, an accurate segmentation of the nodes reached by cancer can constitute a precious help for computer-aided diagnosis. However, due to the size of axillary lymph nodes in PET/CT images, and to the low resolution of PET data where their abnomal metabolic hyperactivity may be observed, segmentation remains a challenging task. We investigate the relevance of considering axillary lymph nodes segmentation from PET/CT images, based on Convolutional Neural Networks (CNNs). To this end, our initial working hypotheses were twofold: first, taking advantage of both anatomical information from CT, for detecting the nodes, and from functional information from PET for detecting the inflammatory ones; second, considering region-based attributes extracted from component-tree analysis of PET images in order to enrich the information natively carried by PET, with features that can hardly be inferred by CNNs directly from the images. We describe our first results, and discuss about the validity of these working hypotheses.

1 Introduction

Breast cancer is one of the most common diseases in women and one of the principal causes of death in females. Approximately 1.38 million cases are detected worldwide per year and as a consequence causes 458,000 deaths. This type of cancer develops from breast tissue; lymph nodes near these regions are then among the first structures to be affected. This motivates the involvement of lymph nodes in the usual TNM protocol dedicated to the staging of breast, that relies on three criteria: size of tumor (T); number of lymph nodes reached by cancer (N); and metastasis state (M).

Positron Emission Tomography (PET), generally coupled with X-ray Computed Tomography (CT) is widely used for imaging purpose in cancer, and in particular in the case of breast cancer (Vercher-Conejero et al., 2015; Krammer et al., 2015; Kaseda et al., 2016; Piva et al., 2017). Whereas PET provides information on the high metabolism of cancerous cells, CT provides anatomical information on the structures of interest, with a high spatial resolution. However, it was observed by Groheux et al. (2016) that PET/CT data experiment two major

limitations. First, patients in early stages of cancer may have a very small quantity of cancerous cells. In this context, PET/CT may not easily allow to detect these few cells. Second, as inflammatory cells have a metabolism similar to cancerous cells, the putative presence of such cells may lead to false positives.

Despite these difficulties, PET/CT data constitute an important source of information that may be used for computer-aided diagnosis in the case of breast cancer. The challenging properties of these bimodal images (low resolution of PET, possible presence of false positives) also argue in favour of developing robust lymph node segmentation methods.

However, the literature specifically dedicated to lymph node segmentation is still rather limited. In this context, Deep Learning (DL) has recently emerged as a promising segmentation paradigm (Ehteshami Bejnordi et al., 2017), that seems to outperform other standard machine learning approaches for this specific task. In particular, it was shown by Wang et al. (2017) that Convolutional Neural Networks (CNNs) (Long et al., 2015), already used for PET/CT co-segmentation by Zhong et al. (2018), is a potentially relevant paradigm.

A second, recent approach dedicated to PET/CT analysis consists of considering hierarchical image models in order to emphasize the mixed spatial-spectral information carried by PET data. In this context, some morphological trees, and in particular the component-tree (Salembier et al., 1998), have been involved in segmentation methods, e.g. for interactive PET segmentation (Grossiord et al., 2015), lymphoma lesion segmentation from PET/CT (Grossiord et al., 2017), or coupled PET/enhanced CT co-segmentation (Alvarez Padilla et al., 2018). These methods rely on the hypothesis that hierarchical image models constitute an efficient data-structure for extracting high-level, region-based features that can be hardly computed by other strategies (Machairas et al., 2016; Conze et al., 2017).

Our initial purpose was to consider jointly both approaches, namely CNNs and hierarchical image models, in order to develop a lymph node automatic segmentation method that will take advantage not only of the complementary information carried by PET and CT, but also of high-level features that could not be implicitly discovered from the native images by deep-learning architectures. We present our methodology, and experimental results that shed light on the successes and failures of the proposed approach, and emphasize the next steps and challenges to be tackled in our work.

2 CNN architecture

Our proposed CNN is based on the U-Net architecture (Ronneberger et al., 2015); see Figure 1. The first 3 layers act as an encoder; they perform volume downsampling that correspond to the CNN feature extraction. The last 3 layers act, symetrically, as a decoder; they perform volume up-sampling. During this up-sampling, some skip connections issued from the input PET image provide information for the output volume reconstruction. The loss function is the Dice score whereas the used optimizer is Nadam (Dozat, 2016).

This architecture is designed for considering multiple inputs, by duplicating the encoder part for each of these inputs. Basically, it uses two inputs, namely the PET image and its associated CT image. Indeed, cancer lymph nodes are structures with a globally round shape, which are perceived in PET as high-intensity, compact structures. However, other tissues and organs with round shape can present a high metabolic activity in PET, thus leading to possible false positives segmentation. In order to discriminate both kinds of structures, nuclear medicine

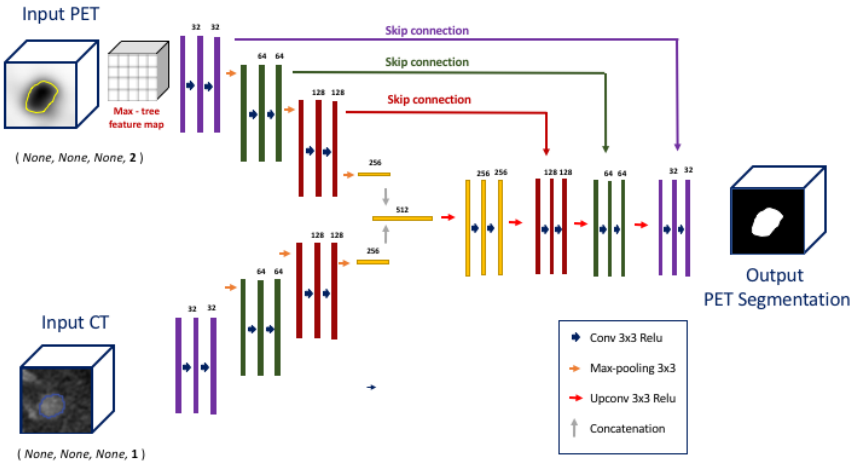

FIG. 1 – *The CNN architecture, illustrated here for two inputs, namely the PET and CT images.*

doctors generally identify cancer lymph nodes by observing both PET and CT data. This is a strategy we chose to reproduce in our CNN, in order to also take into account morphological information in addition to functional one.

Our working hypothesis is that augmenting the information provided as input should improve the segmentation efficiency of the CNN, compared to using the PET data only. This motivates the addition of the CT data. More generally, this also justifies our strategy of computing region-based features from these data, in order to build feature maps involved as supplementary inputs.

3 Data

Our data are composed of 52 PET/CT full-body scans coming from 52 patients with different breast cancer stages. These data may have different resolutions. For instance, 4 of them were obtained with a scanner purchased before 2015, thus resulting in lower resolutions compared to the other images. In order to homogeneize the data, all exams were preprocessed in order to obtain isotropic resolution of 1 mm^3. Since they are acquired during the same exam, the PET and CT volumes were assumed registered, so that one voxel in the CT volume should correspond to the same voxel in the PET volume.

After normalization, each exam contains millions of voxels (e.g. $500 \times 500 \times 800$ volumes). This implies a high cost in terms of both memory and computational resources. In order to reduce this cost, thresholding-based lung segmentation was performed in order to extract the axillary region of interest (ROI). Beyond cost reduction, this also allows one to avoid false positives that may be generated within other parts of the body. Figure 2 illustrates this extraction of the axillary ROI.

We trained our model with 320 lymph nodes of CT full body scans (from neck to coccyx) from these 52 patients' images. Initially, we considered to work with both pathological and

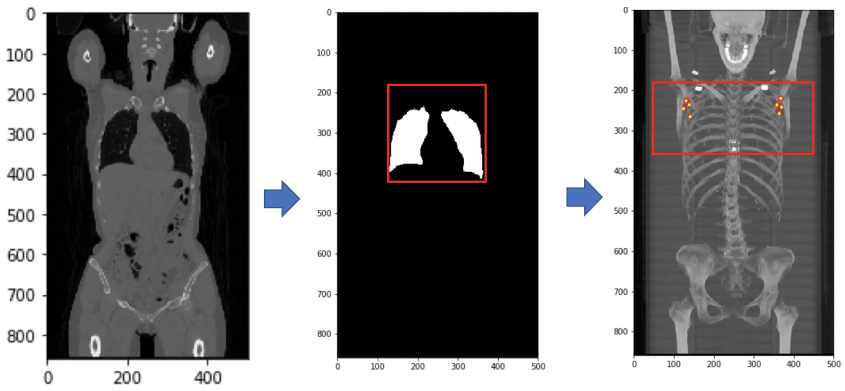

FIG. 2 – *From left to right: original volume; mask of the segmented lungs; extraction of the axilary ROI (coronal view).*

non-pathological lymph nodes. As a consequence, the morphology of these nodes may be different. Indeed, most lymph nodes that contain cells with tumors tend to have a larger volume, and an ellipsoidal morphologically. By contrast, healthy nodes tend to present a "bean" shape, and they are smaller. Figure 3 illustrates the difference between a node that contains tumour cells versus a node that does not.

4 Component-tree-based feature extraction

The component-tree is a lossless hierarchical model dedicated to grey-level images. Basically, a component-tree is a rooted, connected, acyclic graph (i.e. a tree) where each node corresponds to a connected component of a binary level-set of the image. These nodes / connected components are organized with respect to the inclusion relation.

When considering the $\leq$ relation on grey-level values of the image, the component-tree is also called max-tree. In such case, the root of the tree corresponds to the level-set at the lowest value (i.e. 0) where the unique connected component is the whole image support. At the other side of the max-tree, i.e. at the extremities of the branches, the leaves correspond to the flat zones of locally maximal values. In the case of PET images, these regions correspond to high-metabolism areas.

The max-tree can be used as a data-structure allowing to compute features for each node / connected component (Breen and Jones, 1996). More precisely, our purpose is to compute for each relevant node of the max-tree of the PET image, some feature values that gather high-level information that are not directly available to CNNs.

4.1 Node feature extraction

Let us assume that we have computed the max-tree of the considered PET image. For each node, we aim to compute two specific features: (1) the volume; and (2) the compacity

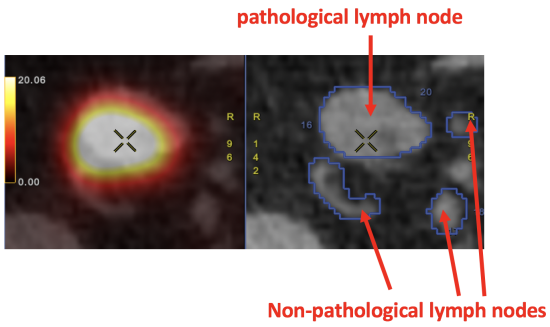

FIG. 3 – *A pathological lymph node next to 3 non-pathological ones (axial view).*

(Grossiord et al., 2015). Indeed, the lesions are of low volume (between 5 and 40 mm^3) and rather compact (round shape).

The volume (V) is proportional to the number of voxels in the connected component of the processed node.

The compacity (C) is more complex to compute, in particular in a Cartesian grid of dimension 3, where the size of a digital boundary of dimension 2 is not defined in a well-posed way. We chose to approximate the compacity by computing the bounding box of the connected component, and by considering as compact an object with a bounding box close to a cube (i.e. with its height, width and depth globally equal). This measure, although not exact, constitutes a fair approximation of compacity, in particular when dealing with spherical targets.

For each node, the two feature values V and C are computed. For the nodes of relevant volume V (i.e. between 5 and 40 mm^3) we associate the node to its actual compacity. The other, over/undersized, nodes are assigned a negative value that means that the feature / the node is non-relevant.

4.2 Feature map construction

Each node of the max-tree of the PET image is then associated with two feature values V and C. However, CNNs consist of performing convolutions with 3D kernels acting on 3D data. This is not the case of a max-tree, which is organized as a dimension-less graph structure. It is then mandatory to embed the feature values computed at the nodes in a feature map defined the same way as the PET image, i.e. in a 3D Cartesian grid.

A voxel of the PET image belongs, in general, to many nodes of the max-tree. Consequently, for a given voxel, many feature values are available for V and C. In particular, it is required to choose, for each voxel, the "most relevant" feature values between the candidate nodes where it lies.

Then, for any voxel x, we study the evolution of the value V for all the nodes from the node of maximal value containing x, until the root node. We search the first strong gradient on V between two successive nodes. It occurs when the current branch is fused with others, and/or when other regions of different semantics merge with the current region containing x.

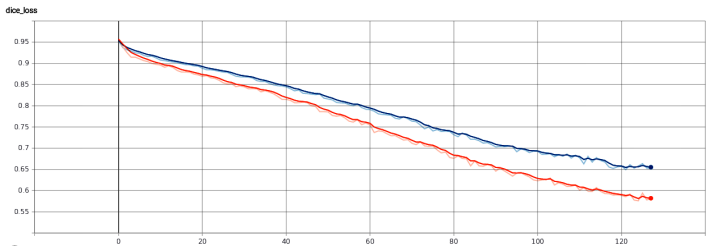

FIG. 4 – *Loss function after 125 epochs. The orange slope corresponds to the training set loss function whereas the blue slope corresponds to the validation set loss function.*

When such gradient is detected, we choose as feature values for x the values V and C of the last node before the gradient occurs.

These feature maps finally computed can then be considered as input of the CNN, for both learning and segmentation.

5 Segmentation results

We trained on 640 patches (320 patches containing lymph nodes and 320 patches containing other structures within the axillary region). The initial learning rate was set to 0.0001 and the number of epochs was 125.

Figure 4 shows the loss function evolution during the training. The loss function value on the validation set converges near 0.6, whereas the loss function value on the training set converges near 0.5.

Figure 5 illustrates segmentation results on 6 different exams from the validation set. One can observe in cases (d–f) that our CNN model is able to distinguish the region that is below the armpits from the rest of the axillary region. However, segmentation results are still not sufficiently accurate when it comes to predicting the contours of the lymph nodes. A possible reason to this fact is that lymph nodes are attached to the lymphatic vessels, and both have the same intensity. It is then difficult, even for medical experts, to correctly delineate the boundary between a lymph node and the lymphatic vessel to which it is connected.

In cases (a–c) we can observe an over-segmentation; larger structures near the heart are segmented in addition to lymph nodes. This over-segmentation influences the Dice score. These segmentation errors can be explained by the fact that the region near the heart tends to be hyperfixating in PET images. In addition, small, round structures around the heart which have a morphology similar to that of the lymph nodes appear to be inflamed or contain cancer cells in several exams. In such cases, using max-tree information could be valuable. Indeed, descriptive features such as lymph node area or lymph node morphology, obtained from the max-tree, could help to discriminate connected components corresponding to such false positives.

On the other hand, our model only relies on 360 examples of lymph nodes. In order to improve the results, it will be mandatory to increase our training set. Data augmentation techniques may contribute to such enrichment of the pool of information available for learning purpose. Nevertheless, additional real examples are also of crucial importance in order to enrich the variability of the training set.

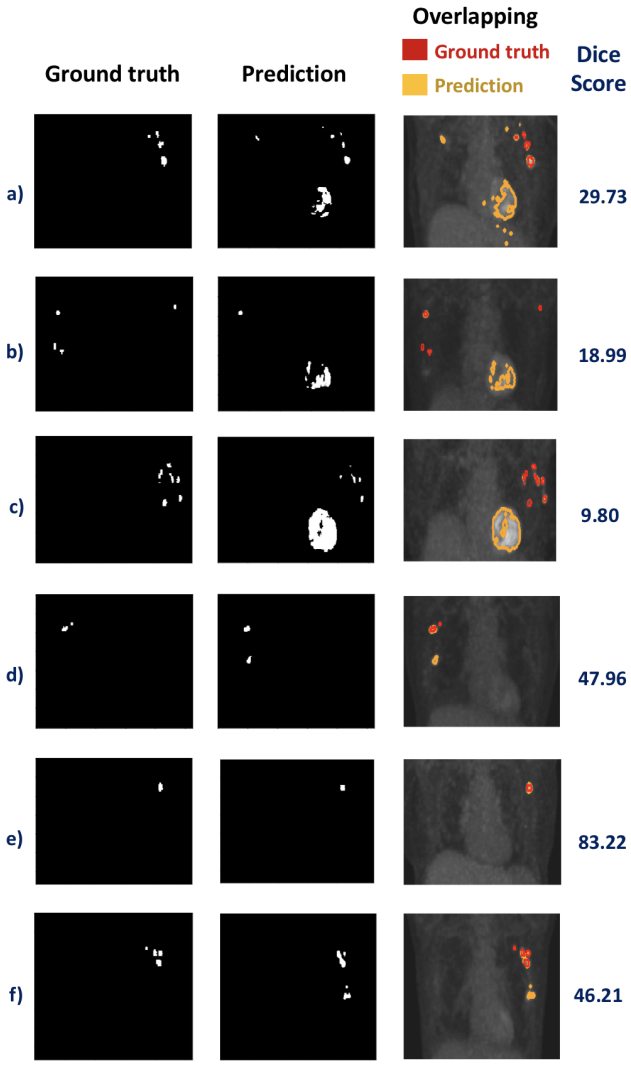

FIG. 5 – *Predictions results on 6 different patients (maximum intensity projection, coronal view, axillary region). First column: ground-truth. Second column: lymph node segmentation results. Third column: comparison of ground-truth contours (in red) and segmentation contours (in yellow). Fourth column: Dice score (over 100) of each segmentation result.*

6 Discussion

6.1 Lymph node clusters

In rare cases, we can observe that patients may contain several lymph nodes visually connected together, due to their small size and close relative positions, see Figure 6. In such cases,

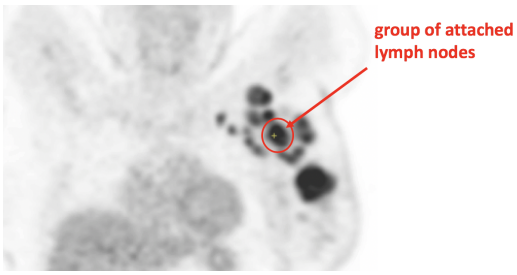

FIG. 6 – *Lymph nodes, forming a cluster that visually appears as a large, single node. PET image, coronal view.*

medical experts consider these clusters as single nodes, since it is impossible for them to discriminate the lymph nodes composition within the clusters. In particular, it is then impossible to determine exactly how many lymph nodes are connected.

Considering these clusters in our training set, and assimilating them as single nodes, may lead to introduce inconsistency into the training process. Indeed these clusters are significantly different from single nodes, both in terms of size and morphology. These cases are then challenging to predict, since they appear as outliers (both in training and testing), compared to usual lymph nodes.

6.2 Registration issues between PET and CT volumes

Both CT and PET images are computed, slice by slice, during a same acquisition. However, the time required for both is quite different, from a few seconds for the CT, to several minutes for the PET. In such conditions, two kinds of movements can alter the PET image acquisition (whereas CT remains robust to such issues). First, breathing has an impact on the position of the axillary lymph nodes, which are located close to the lungs. An acquisition of several minutes will be altered by these physiological artifacts, leading to partial volume effects and/or displacement of the putative position of the nodes in the image. On the other hand, the patient is asked to have his/her arms stretched upwards. This position can rapidly become uncomfortable, leading the patient to slightly move. In such case, the spatial coherence of the lymph nodes between the beginning and the end of the PET image acquisition may be altered.

As a consequence, a same lymph node in the CT image may be located a few millimeters away (and blurred) in the PET image, as illustrated in Figure 7.

In our initial approach, the ground-truths for the lymph nodes contours were defined from the CT image, taking advantage of the accurate morphological information provided by these images. Unfortunately, the movement-based registration issues make this strategy non-valid, since we have no sufficient guarantee of spatial coherence between the lymph node position in the CT and the PET.

A potential solution to tackle this issue may consist of considering multimodal, non-rigid registration procedures. However, such approaches are designed to compute 3D continuous mappings between two images, whereas in our case, the spatial incoherence is related to complex 3D + time movements that cannot be modeled by such kinds of transformations. As a conclusion, it will then be mandatory to define further ground-truth directly from the PET im-

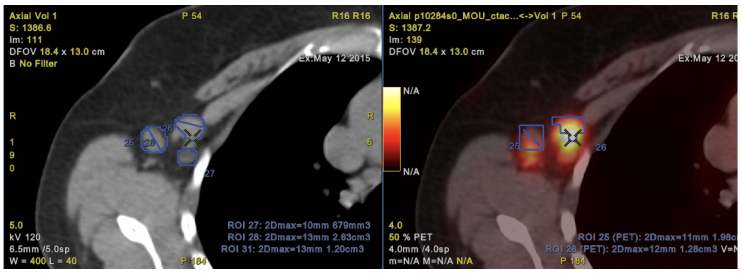

FIG. 7 – *Left: contouring of 2 lymph nodes in CT axial view. Right: Contouring of the same 2 lymph nodes in PET axial view. (See text.)*

ages (and no longer from CT ones). This will induce another difficulty, that lies in the fact that such functional images do not provide crisp frontiers between the nodes and their neighbourhood.

7 Perspective works

Segmentation of lymph nodes in PET/CT images is challenging, especially because these structures are small, sparse and located near the lungs. In our future work, we aim at segmenting tumors using two types of Neural Network architectures using different types of inputs: (1) only gray-level information from PET; (2) only gray-level information from PET/CT; (3) gray-level information from the PET/CT combined with max tree descriptors.

It is important to keep in mind that PET and CT modalities can both bring information, which are indeed complementary. As a consequence, during training using both PET and CT information, it will be important to provide patches in PET and CT, whereas ground-truth mask will correspond to the PET contours only.

In addition, it will be also mandatory to increase both the precision of the segmentation, but also to determine a set of ground-truths sufficiently large, involving at least 1,000 different lymph nodes, in order to reach a sufficient amount of data for further avoiding over-fitting.

Acknowledgements The research leading to these results has been supported by the French *Association Nationale Recherche Technologie* (ANRT).

References

Alvarez Padilla, F., B. Romaniuk, B. Naegel, S. Servagi-Vernat, D. Morland, D. Papathanassiou, and N. Passat (2018). Hierarchical forest attributes for multimodal tumor segmentation on FDG-PET/contrast-enhanced CT. In *ISBI, Procs.*, pp. 163–167.

Breen, E. J. and R. Jones (1996). Attribute openings, thinnings, and granulometries. *Computer Vision and Image Understanding 64*, 377–389.

Conze, P.-H., V. Noblet, F. Rousseau, F. Heitz, V. de Blasi, R. Memeo, and P. Pessaux (2017). Scale-adaptive supervoxel-based random forests for liver tumor segmentation in dynamic

contrast-enhanced CT scans. *International Journal for Computer Assisted Radiology and Surgery 12*, 223–233.

Dozat, T. (2016). Incorporating Nesterov momentum into Adam. In *ICLR Workshop, Procs.*, pp. 2013–2016.

Ehteshami Bejnordi, B., M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. van der Laak, and the CAMELYON16 Consortium (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Journal of the American Medical Association 318*, 2199–2210.

Groheux, D., A. Cochet, O. Humbert, J.-L. Alberini, E. Hindié, and D. Mankoff (2016). 18F-FDG PET/CT for staging and restaging of breast cancer. *Journal of Nuclear Medicine 57(Supplement 1)*, 17S–26S.

Grossiord, É., H. Talbot, N. Passat, M. Meignan, and L. Najman (2017). Automated 3D lymphoma lesion segmentation from PET/CT characteristics. In *ISBI, Procs.*, pp. 174–178.

Grossiord, É., H. Talbot, N. Passat, M. Meignan, P. Terve, and L. Najman (2015). Hierarchies and shape-space for PET image segmentation. In *ISBI, Procs.*, pp. 1118–1121.

Kaseda, K., K. Watanabe, K. Asakura, A. Kazama, and Y. Ozawa (2016). Identification of false-negative and false-positive diagnoses of lymph node metastases in non-small cell lung cancer patients staged by integrated 18F-FDG-positron emission tomography/computed tomography: A retrospective cohort study. *Thoracic Cancer 7*, 473–480.

Krammer, J., A. Schnitzer, C. Kaiser, K. Buesing, E. Sperk, J. Brade, S. Wasgindt, M. Suetterlin, S. Schoenberg, E. Sutton, and K. Wasser (2015). 18 F-FDG PET/CT for initial staging in breast cancer patients–Is there a relevant impact on treatment planning compared to conventional staging modalities? *European Radiology 25*, 2460–2469.

Long, J., E. Shelhamer, and T. Darrell (2015). Fully convolutional networks for semantic segmentation. In *CVPR, Procs.*, pp. 3431–3440.

Machairas, V., T. Baldeweck, T. Walter, and E. Decencière (2016). New general features based on superpixels for image segmentation learning. In *ISBI, Procs.*, pp. 1409–1413.

Piva, R., F. Ticconi, V. Ceriani, F. Scalorbi, F. Fiz, S. Capitanio, M. Bauckneht, G. Cittadini, G. Sambuceti, and S. Morbelli (2017). Comparative diagnostic accuracy of 18F-FDG PET/CT for breast cancer recurrence. *Breast Cancer: Targets and Therapy 9*, 461.

Ronneberger, O., P. Fischer, and T. Brox (2015). U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI, Procs. III*, pp. 234–241.

Salembier, P., A. Oliveras, and L. Garrido (1998). Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing 7*, 555–570.

Vercher-Conejero, J., L. Pelegrí-Martinez, D. Lopez-Azna, and M. del Puig Cózar-Santiago (2015). Positron emission tomography in breast cancer. *Diagnostics 5*, 61–83.

Wang, H., Z. Zhou, Y. Li, Z. Chen, P. Lu, W. Wang, W. Liu, and L. Yu (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. *EJNMMI Research 7*, 11.

Zhong, Z., Y. Kim, L. Zhou, K. Plichta, B. Allen, J. Buatti, and X. Wu (2018). 3D fully convolutional networks for co-segmentation of tumors on PET-CT images. In *ISBI, Procs.*, pp. 228–231.