Evidential Segmentation of 3D PET/CT Images

Ling Huang, Su Ruan, Pierre Decazes, Thierry Denoeux

Evidential segmentation of 3D PET/CT images*

Ling Huang[1,2], Su Ruan[2], Pierre Decazes[3], and Thierry Denœux[1,4]

[1] Université de technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France
ling.huang@utc.fr
[2] University of Rouen Normandy, Quantif, LITIS, Rouen, France
[3] CHB Hospital, Rouen, France
[4] Institut universitaire de France, Paris, France

Abstract. Positron Emission Tomography (PET) and Computed Tomography (CT) are two modalities widely used in medical image analysis. Accurately detecting and segmenting lymphomas from these two imaging modalities are critical tasks for cancer staging and radiotherapy planning. However, this task is still challenging due to the complexity of PET/CT images, and the computation cost to process 3D data. In this paper, a segmentation method based on belief functions is proposed to segment lymphomas in 3D PET/CT images. The architecture is composed of a feature extraction module and an evidential segmentation (ES) module. The ES module outputs not only segmentation results (binary maps indicating the presence or absence of lymphoma in each voxel) but also uncertainty maps quantifying the classification uncertainty. The whole model is optimized by minimizing Dice and uncertainty loss functions to increase segmentation accuracy. The method was evaluated on a database of 173 patients with diffuse large b-cell lymphoma. Quantitative and qualitative results show that our method outperforms the state-of-the-art methods.

Keywords: lymphoma segmentation · 3D PET/CT · belief functions · Dempster-Shafer theory · uncertainty quantification · deep learning

1 Introduction

Positron Emission Tomography - Computed Tomography (PET/CT) scanning is an effective imaging tool for lymphoma segmentation with application to clinical diagnosis and radiotherapy planning. The standardized uptake value (SUV) for PET images is widely used to locate and segment lymphomas thanks to its high sensitivity and specificity to the metabolic activity of tumor. CT images are usually used jointly with PET images because of their anatomical feature representation capability.

Although a lot of progress has been made in computer-aided lymphoma segmentation, the segmentation of whole-body lymphomas is still challenging.

* This work was supported by the China Scholarship Council (grant 201808331005). It was carried out in the framework of the Labex MS2T (Reference ANR-11-IDEX-0004-02)

(Fig. 1 shows an example of lymphoma patient. There is great variation in intensity distribution, shape, type and number of lymphomas). The methods can be classified into three main categories: SUV-threshold-based [5], region-growing-based [4] and Convolutional Neural Network (CNN)-based [7] methods. For PET images, it is common to segment lymphomas with a set of fixed SUV thresholds. This method is fast but lacks of flexibility in boundary delineation and requires domain knowledge to locate the region of interest. Region-growing-based methods have been proposed to optimize boundary delineation by taking texture and shape information into account. However, those methods still need clinicians to locate the seeds for region growing [11].

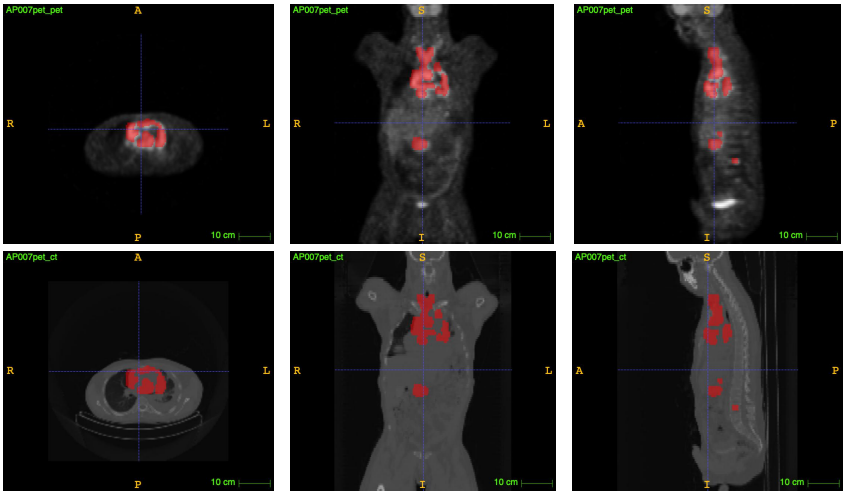

Fig. 1. Examples of patient with lymphomas. The first and second rows show, respectively PET and CT, slices of one patient in axial, sagittal and coronal views. The lymphomas are marked in red.

CNN-based segmentation methods have recently achieved great success. The UNet architecture [12] has become the most popular medical image segmentation model. Driven by different tasks and datasets, many extended and optimized variants of UNet have been proposed, such as VNet [9], nnUNet [6] and SegResNet [10]. In [7], Li et al. propose a SegResNet-based lymphoma segmentation model with a two-flow architecture (segmentation and reconstruction flows). In [1], Blanc-Durand et al. propose a nnUNet-based lymphoma segmentation network.

Because of low resolution and contrast due to limitations of medical imaging technology, PET/CT image segmentation results are tainted with uncertainty, which greatly limits the segmentation accuracy. Traditional uncertainty measurement methods [8] focus on model rather than information uncertainty to improve the robustness of the model. Belief function (BF) theory [13][3], also

known as Dempster-Shafer theory, is a formal theory for information modeling, evidence combination and decision-making under uncertainty. In this paper, we propose a 3D PET/CT diffuse large bcell lymphoma segmentation model based on BF theory and deep learning. The proposed deep neural network architecture is composed of a UNet module for feature extraction and a BF module for decision with uncertainty quantification. End-to-end learning is achieved by minimizing a two-part loss function allowing us to increase the Dice score while decrease the uncertainty. The model will first be described in Section 2 and experimental results will be reported in Section 3.

2 Methods

2.1 Network Architecture

Fig. 2 shows the global lymphoma segmentation architecture (ES-UNet). It is composed of (1) an encoder-decoder feature extraction module (UNet), and (2) an evidential segmentation (ES) module comprising a distance activation layer, a basic belief assignment layer and a mass fusion layer. Details about the ES module will be given in Section 2.2. Two loss terms are used for optimizing the training process: the *Dice loss*, which quantifies the segmentation accuracy and the *uncertainty loss*, which quantifies the segmentation uncertainty. These loss functions will be described in Section 2.3. A "slim UNet" with $(8, 16, 32, 64, 128)$ convolution filters was implemented to reduce computation cost and avoid overfitting.

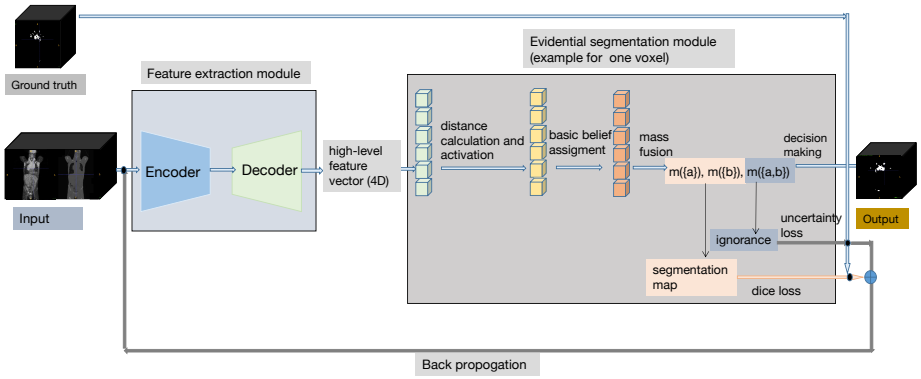

Fig. 2. Global lymphoma segmentation model (ES-UNet).

2.2 Evidential segmentation module

A probabilistic network with a softmax output layer may assign voxels a high probability of belonging to one class while the segmentation uncertainty is ac-

tually very high because, e.g., the voxel is located close to the fuzzy boundary between the tumor region and other tissues. Based on the evidential neural network model introduced in [2] and using an approach similar to that recently described in [14], we propose a BF theory-based ES module to quantify the uncertainty about the class of each voxel by a Dempster-Shafer mass function. The main idea of the ES module is to assign a mass to each of the K classes and to the whole set of classes Ω, based on the distance between the feature vector of each voxel and I prototype centers. For a given voxel x, each prototype p_i is considered as a piece of evidence, the reliability of which decreases with the Euclidean distance d_i between x and p_i. Each prototype p_i is assumed to have a membership degree u_{ik} to each class ω_k with the constraint $\sum_{k=1}^{K} u_{ik} = 1$. The mass function induced by prototype p_i is

$$m_i(\{\omega_k\}) = \alpha_i u_{ik} \exp(-\gamma_i d_i^2), \quad k = 1, \ldots, K \tag{1a}$$

$$m_i(\Omega) = 1 - \alpha_i \exp(-\gamma_i d_i^2), \tag{1b}$$

The network parameters are the prototypes p_i, the coefficients α_i and γ_i, and the membership degrees u_{ik}. They are learnt by minimizing a loss function [2].

The mass functions induced by the I prototypes are then combined by Dempster's rule [13]

$$m = \bigoplus_{i=1}^{I} m_i. \tag{2}$$

The ES module outputs for each voxel three mass values: two masses corresponding to lymphoma ($\{a\}$) and background ($\{b\}$), and an additional mass corresponding to ignorance (Ω). For the voxels that are easy to classify into lymphoma or background, the mass values $m(\{a\})$ or $m(\{b\})$ are high and the mass $m(\{a, b\})$ is low. A high mass $m(\{a, b\})$ signals a lack of information to make a reliable decision. Thus, some constraints are required to reduce $m(\Omega)$ during training, as will be explained in Section 2.3.

Since the output of the ES module is a mass function with $K + 1$ focal sets while there are K classes, we transform the mass function by distributing a fraction ξ of $m(\Omega)$ to each class, as

$$T_{\xi,k} = m(\{\omega_k\}) + \xi\, m(\Omega) \quad \text{with} \quad 0 \leqslant \xi \leqslant 1. \tag{3}$$

In this paper, $\Omega = \{a, b\}$ and $K = 2$, thus we set $\xi = 0.5$. The crisp output S of module ES module is defined as $S = 1$ if $m(\{a\}) < m(\{b\})$ and $S = 0$ otherwise.

2.3 Loss function based on accuracy and uncertainty for segmentation

In general, a good segmentation system is expected to make few segmentation errors while providing as informative outputs as possible. Since we quantify uncertainty by the "ignorance class" via the evidential network, we propose to minimize a loss function defined as the sum of two terms: a Dice loss loss_d that

measures the discrepancy between the ground truth and segmentation outputs, and an uncertainty loss loss_u that measures the uncertainty of the segmentation outputs. We use the Dice loss instead of the original cross-entropy loss in UNet because the goal of segmentation is to maximize the Dice coefficient. The Dice loss is defined as

$$\mathsf{loss}_d = 1 - \frac{2\sum_{n=1}^{N} S_n G_n}{\sum_{n=1}^{N} S_n + \sum_{n=1}^{N} G_n}, \tag{4}$$

where N is the number of voxels in the image volume, S_n and is the n-th voxel of the segmented output image and G_n is n-th voxel of the ground truth image. The uncertainty loss is defined as

$$\mathsf{loss}_u = \frac{1}{N}\sum_{n=1}^{N}[m_n(\Omega)]^2, \tag{5}$$

where m_n is the mass function computed for voxel n. With the uncertainty loss, the parameters of the model can be further optimized and more precise segmentation results can be obtained. The total loss function is then

$$\mathsf{loss} = \mathsf{loss}_d + \mathsf{loss}_u + \lambda \left\|\alpha\right\|_1, \tag{6}$$

where λ is the regularization coefficient for parameter vector $\alpha = (\alpha_1, \ldots, \alpha_I)$ with the α_i defined in (1). The regularization term allows us to decrease the influence of unimportant prototype centers and avoid overfitting.

3 Experimental results

3.1 Experimental settings

The dataset contains 3D images from 173 patients who were diagnosed with large b-cell lymphoma and underwent PET/CT examination. The study was approved as a retrospective study by the Henri Becquerel Center Institutional Review Board. The lymphomas in mask images were delineated manually by experts and considered as ground truth G. The size and spatial resolution of PET and CT images and the corresponding mask images vary due to different imaging machines and operations, from $267\times512\times512$ to $478\times512\times512$ and from $276\times144\times144$ to $407\times256\times256$, respectively; this makes it difficult to transfer the data into a deep neural model directly. We resized PET, CT and mask images to the same 3D size $256\times256\times128$, and we applied intensity normalization to both PET and CT images from each patient independently by subtracting the mean and dividing by the standard deviation of the body region only. For data augmentation, we applied a random intensity shift between $[-0.1, 0.1]$ of the standard deviation of each channel, as well as a random scaling intensity of the input between scales $[0.9, 1.1]$.

We randomly selected 80% of the data for training, 10% for validation and 10% for testing. Dice score, sensitivity, specificity, precision and F1 score were

used to evaluate the segmentation performance. We first computed the five indices for each test patient and then averaged these indices over the patients. During training, PET and CT images were concatenated as a two-channel input. The number of prototypes was set to 20, taking class number and computation cost into consideration. The prototype vectors and membership degrees were initialized randomly using uniform distributions, while the parameters α and γ were initialized, respectively, at 0.5 and 0.01. The learning rate was set to 10^{-3} during training and the model was trained with 50 epochs using the Adam optimization algorithm. The regularization coefficient λ in (6) was set to 10^{-5}. All methods were implemented in Python with a PyTorch-based, medical image framework MONAI and were trained and tested on a desktop with a 2.20GHz Intel(R) Xeon(R) CPU E5-2698 v4 and a Tesla V100-SXM2 graphics card with 32 GB GPU memory.

3.2 Results and discussion

The quantitative results are shown in Table 1. Our model outperforms the baseline model UNet as well as the other state-of-the-art methods. In particular, our model outperforms the best model SegResNet by, respectively, 1.9%, 2.4%, 1.4% in Dice score, Sensitivity and F1 score. It should be noted that the state-of-the-art models were trained with 100 epochs on our dataset because they are slower to converge during training. Fig. 3 displays the learning curves of the training loss and validation Dice score for UNet and ES-UNet, showing the advantage of ES-UNet in terms not only of segmentation accuracy, but also of convergence speed. Fig. 4 shows the segmentation and uncertainty maps at different steps during the training of ES-UNet. Our model quantifies the uncertainty of ambiguous pixels instead of classifying them unambiguously into a single class. The uncertainty decreases during the learning process thanks to the minimization of the uncertainty loss term.

Table 1. Performance comparison with the baseline methods on the test set.

Models	Dice score	Sensitivity	Specificity	Precision	F1 score
ES-UNet (our model)	**0.830**	0.923	0.908	0.912	**0.915**
UNet [12]	0.769	0.798	**0.963**	0.890	0.833
nnUNet [6]	0.702	**0.950**	0.499	0.758	0.807
VNet [9]	0.802	0.882	0.904	0.916	0.909
SegResNet [10]	0.811	0.899	0.942	**0.925**	0.901

Fig. 5 shows an example of segmentation results obtained by ES-UNet. Our model can locate and segment most of the lymphomas. The segmentation results were found credible and were confirmed by experts.

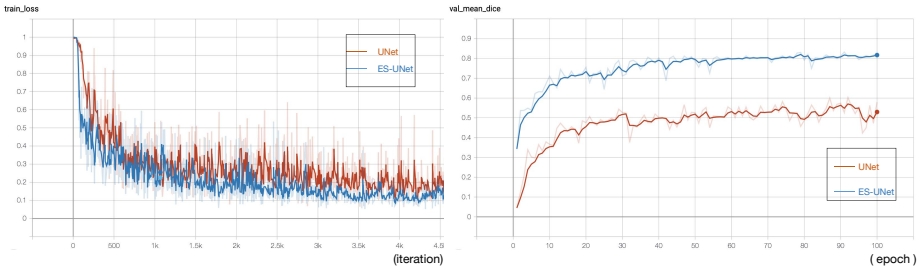

Fig. 3. Training process visualization: training loss (left) and validation Dice score (right).

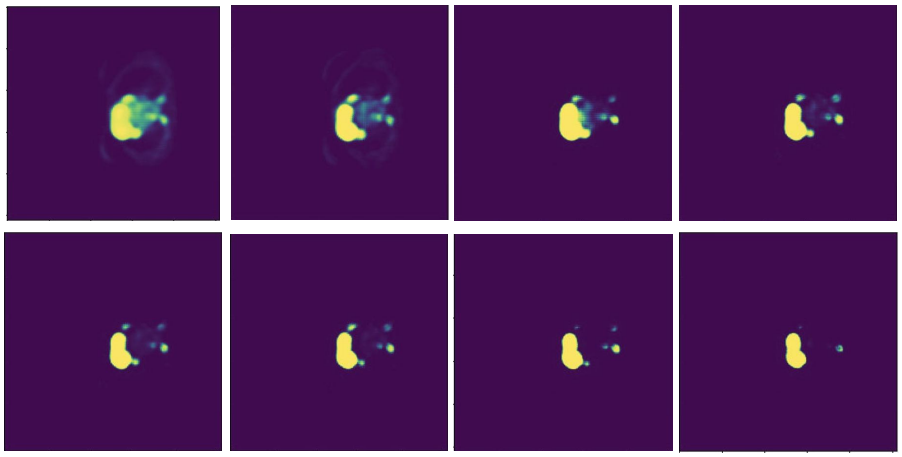

Fig. 4. Uncertainty maps obtained during training, corresponding to different training steps for the same image. For one map, the pixels classified to background, lymphoma, ignorance are marked in purple, yellow and iridescent, respectively.

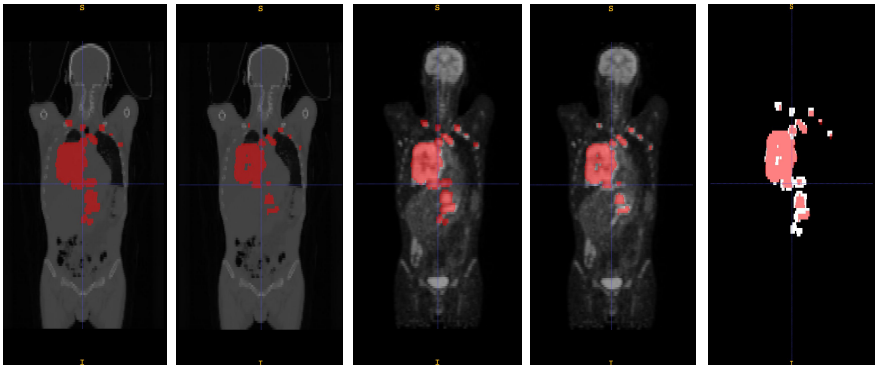

Fig. 5. Segmentation results of ES-UNet. From left to right: ground truth and segmented lymphomas in overlapped CT modality, ground truth and segmented lymphomas overlapped in PET modality, difference map between the ground truth, and segmented lymphomas.

4 Conclusion

An evidential segmentation framework (ES-UNet) for segmentation of lymphomas from 3D PET/CT with uncertainty quantification has been introduced. The proposed architecture is based on the concatenation of a UNet and an evidential segmentation layer, making it possible to compute output mass functions for each voxel. The training is performed by minimizing a two-part loss function composed of a Dice loss and an uncertainty loss, with the effect of increasing the Dice score while decreasing the uncertainty. Qualitative and quantitative evaluations show promising results when compared to the baseline model UNet as well as the state-of-the-art methods. While we only concatenated PET and CT as a two-channel input in this work, future research will tackle multi-modality medical image fusion with BF theory by considering PET and CT images separately. Moreover, the sensitivity of the results with respect to the number of prototypes and the initial parameters will be studied in greater detail.

References

1. Blanc-Durand, P., Jégou, S., Kanoun, S., et al.: Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. European Journal of Nuclear Medicine and Molecular Imaging pp. 1–9 (2020)
2. Denœux, T.: A neural network classifier based on Dempster-Shafer theory. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans **30**(2), 131–150 (2000)
3. Denœux, T., Dubois, D., Prade, H.: Representations of uncertainty in artificial intelligence: Beyond probability and possibility. In: Marquis, P., Papini, O., Prade, H. (eds.) A Guided Tour of Artificial Intelligence Research, vol. 1, chap. 4, pp. 119–150. Springer Verlag (2020)

4. Hu, H., Decazes, P., Vera, P., Li, H., Ruan, S.: Detection and segmentation of lymphomas in 3D PET images via clustering with entropy-based optimization strategy. International journal of computer assisted radiology and surgery **14**(10), 1715–1724 (2019)
5. Ilyas, H., Mikhaeel, N.G., Dunn, J.T., Rahman, F., Møller, H., Smith, D., Barrington, S.F.: Defining the optimal method for measuring baseline metabolic tumour volume in diffuse large b cell lymphoma. European journal of nuclear medicine and molecular imaging **45**(7), 1142–1154 (2018)
6. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
7. Li, H., Jiang, H., Li, S., et al.: DenseX-net: an end-to-end model for lymphoma segmentation in whole-body PET/CT images. IEEE Access **8**, 8004–8018 (2019)
8. Mehta, R., Christinck, T., Nair, T., et al.: Propagating uncertainty across cascaded medical imaging tasks for improved deep learning inference. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures, pp. 23–32. Springer (2019)
9. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision. pp. 565–571. IEEE (2016)
10. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: International MICCAI Brain lesion Workshop. pp. 311–320. Springer (2018)
11. Onoma, D., Ruan, S., Thureau, S., et al.: Segmentation of heterogeneous or small FDG PET positive tissue based on a 3d-locally adaptive random walk algorithm. Computerized Medical Imaging and Graphics **38**(8), 753–763 (2014)
12. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
13. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
14. Tong, Z., Xu, P., Denœux, T.: Evidential fully convolutional network for semantic segmentation. Applied Intelligence (2021), `https://doi.org/10.1007/s10489-021-02327-0`