Deep PET/CT fusion with Dempster-Shafer theory for lymphoma segmentation*

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

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

Abstract. Lymphoma detection and segmentation from whole-body Positron Emission Tomography/Computed Tomography (PET/CT) volumes are crucial for surgical indication and radiotherapy. Designing automatic segmentation methods capable of effectively exploiting the information from PET and CT as well as resolving their uncertainty remain a challenge. In this paper, we propose an lymphoma segmentation model using an UNet with an evidential PET/CT fusion layer. Single-modality volumes are trained separately to get initial segmentation maps and an evidential fusion layer is proposed to fuse the two pieces of evidence using Dempster-Shafer theory (DST). Moreover, a multi-task loss function is proposed: in addition to the use of the Dice loss for PET and CT segmentation, a loss function based on the concordance between the two segmentation is added to constrain the final segmentation. We evaluate our proposal on a database of polycentric PET/CT volumes of patients treated for lymphoma, delineated by the experts. Our method get accurate segmentation results with Dice score of 0.726, without any user interaction. Quantitative results show that our method is superior to the state-of-the-art methods.

Keywords: PET/CT · multi-modality fusion · lymphoma segmentation · Dempster-Shafer theory · deep learning

1 Introduction

In the clinical diagnosis and radiotherapy planning of lymphoma, PET/CT scanning is an effective imaging tool for tumor segmentation. In PET volumes, the standardized uptake value (SUV) is widely used to locate and segment lymphomas because of its high sensitivity and specificity [11]. Moreover, CT is

* 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)

usually used in combination with PET because of its good representation of anatomical features.

Considering the multiplicity of lymphoma sites (sometimes more than 100) and the wide variation in distribution, shape, type and number of lymphomas, whole-body PET/CT lymphoma segmentation is still challenging although many segmentation methods have been proposed (see a lymphoma patient in Fig. 1). Computer-aided methods for lymphoma segmentation can be classified into three main categories: SUV-threshold-based, region-growing-based and Convolutional-Neural-Network (CNN)-based [7] methods.

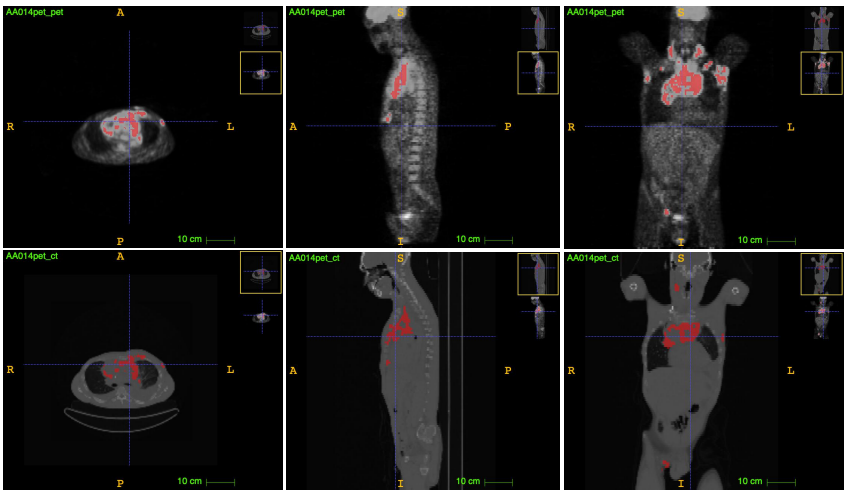


Fig. 1. Examples of PET and CT slices with lymphomas. The first and second row show PET and CT slices of one patient in axial, sagittal and coronal views, respectively.

For PET volumes, it is common to use fixed SUV thresholds to segment lymphoma. This kind of method is fast but it lacks flexibility in boundary delineation and requires clinicians to locate the region of interest. Region-growing-based methods have been proposed to overcome the limitation of boundary delineation of SUV-based methods by taking texture and shape information into account. The common idea is to set the volumes with initializing seeds automatically for segmenting lymphomas [8][4]. However, these methods still require to region of interest manually and are time-consuming.

In recent years, CNNs have achieved great success in computer vision tasks [5][2], as well as in the medical domain. The Fully Connected Network (FCN) model [10] is the first fully convolutional neural network that could be trained end-to-end for pixel-wise classification. UNet [13], a successful modification and extension of FCN, has become the most popular network for medical image segmentation. Based on UNet, many extensions and optimizations have been proposed, such as Deep3D-UNet [18], attention-UNet [12], etc. In [6], Hu et al.

propose a Conv3D-based multi-view PET image fusion strategy for lymphoma segmentation. In this approach, 2D and 3D segmentation are performed separately and the results are fused by a Conv3D layer. In [7], Li et al. propose a whole-body PET/CT lymphoma segmentation method that fuses CT and PET slices by concatenating them before training; the method is based on a two-flow architecture (segmentation flow and reconstruction flow). Using a similar approach, Blanc-Durand et al. propose a CNN-based segmentation network for diffuse large B cell lymphoma segmentation by concatenating PET and CT as two channel inputs [1].

It should be noted that the effective fusion of multi-modality information is of great importance in the medical domain. A single-modality image often does not contain enough information and is often tainted with uncertainty. This is why physicians always use PET/CT volumes together for lymphoma segmentation and radiotherapy. Using CNNs, researchers have mainly adopted probabilistic approaches to data fusion, which can be classified into three strategies [17]: image-level, feature-level and decision-level fusion. However, probabilistic fusion is unable to effectively manage conflicts that occur when the same voxel is labeled with two different labels by CT and PET. Dempster-Shafer theory (DST) [3][14], also known as belief function theory or evidence theory, is a formal framework for information modeling, evidence combination and decision-making with uncertain or imprecise information. Despite the low resolution and contrast of medical images, DST's high ability to describe uncertainty allows us to represent evidence more faithfully than using probabilistic approaches. Researchers from the medical image community have started to actively investigate the use of DST for handling uncertain, imprecision sources of information in different medical tasks, such as medical image retrieval [15], lesion segmentation [9], etc.

In this work, we propose a DST-based PET/CT image fusion model for 3D lymphoma segmentation. To our knowledge, this is the first multi-modality PET/CT volume fusion method using a CNN and DST. The main contributions of this work are (1) a CNN architecture with a DST-based fusion layer that effectively handles uncertainty and conflict when combining PET and CT information; (2) a multi-task loss function making it possible to optimize different segmentation tasks and to increase segmentation accuracy; (3) a 3D segmentation model with end-to-end training for whole-body lymphoma segmentation.

2 Methods

2.1 Network Architecture

Fig. 2 shows the workflow of the multi-modality fusion-based lymphoma segmentation framework. It is composed of two modified encoder-decoder (Unet) modules and an evidential fusion layer. To reduce computation cost, we reduce the number of convolution filters of Unet from $(16, 32, 64, 128, 256)$ to $(8, 16, 32, 64, 128)$ to get a "slim UNet". A fusion layer is constructed based on DST, which will be explained in Section 2.3. Two modality images: PET and CT

are taken as inputs to our framework. We first feed the prepossessed PET volume into UNet1 and the prepossessed CT volume into UNet2 to independently compute their segmentation probability maps seg_{PET} and seg_{CT}. These two 3D maps are then transferred to the fusion layer, which computes an evidential segmentation map seg_F. For training, a multi-task loss function is proposed to minimize the Dice loss and mean square loss between three segmentation maps and masks, as explained in Section 2.4.

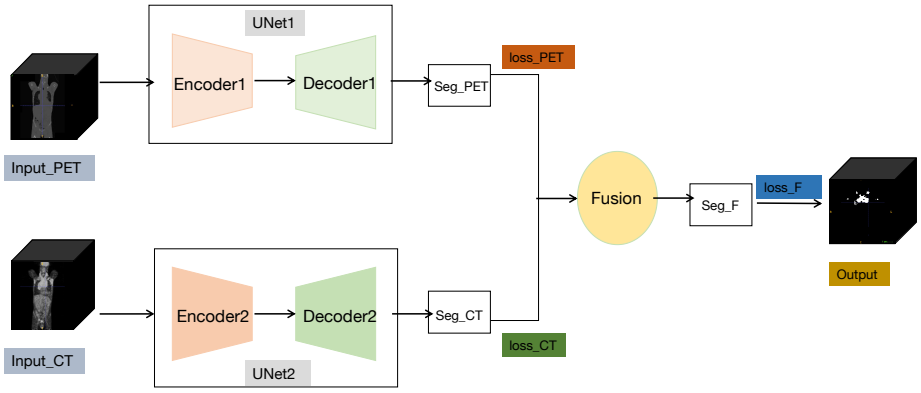


Fig. 2. The global segmentation framework.

2.2 Evidential Fusion Strategy

Let $\Omega = \{\omega_1, \omega_2, ..., \omega_K\}$ be a finite set of possible answers to some question, one and only one of which is true. Evidence about that question can be represented by a mapping m from 2^Ω to $[0, 1]$, called a mass function, such that $\sum_{A\subseteq\Omega} m(A) = 1$. For any hypothesis $A \subseteq \Omega$, the quantity $m(A)$ represents a share of a unit mass of belief allocated to the hypothesis that the truth is in A, and which cannot be allocated to any strict subset of A based on the available evidence. Two mass functions m_1 and m_2 representing two independent items of evidence can be combined by Dempster's rule [14] defined as

$$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B\cap C=A} m_1(B)m_2(C), \tag{1}$$

for all $A \subseteq \Omega, A \neq \emptyset$, and $(m_1 \oplus m_2)(\emptyset) = 0$. In Eq. (1), κ represents the degree conflict between m_1 and m_2 defined as

$$\kappa = \sum_{B\cap C=\emptyset} m_1(B)m_2(C). \tag{2}$$

If all focal sets of m are singletons and m is said to be Bayesian, it is equivalent to a probability distribution. In our case, the lymphoma segmentation task can

be considered as a two-class classification task; the set of possibilities is $\Omega = \{0, 1\}$ where 1 and 0 stand, respectively, for presence and absence of lymphoma in a given voxel. Two Bayesian mass functions m_{PET} and m_{CT} can be obtained from two probability segmentation maps maps seg_{PET} and seg_{CT}. They are aggregated by Dempster's rule (1).

The two segmentation results obtained by PET and CT cannot easily classify boundary voxels due to low signal-to-noise ratio and low contrast, resulting in segmentation uncertainty. Moreover, when seg_{PET} and seg_{CT} assign the same voxel to different classes, there is a high conflict between the two segmentation results. It is difficult to solve these problems by classical fusion methods. In this work, the proposed evidential fusion layer allows us to solve these problems thanks to Dempster's rule. To be specific, for a given voxel, if seg_{PET} gives a probability of 0.26 that it belongs to lymphoma, and seg_{CT} gives a probability of 0.85 that it belongs to lymphoma, the two segmentation results are contradictory and it is unreasonable to simply fuse them by linear combination or majority voting. In our evidential fusion layer, we can reassign the probability that the voxel belongs to lymphoma as 0.67 with Eqs. (1)-(2). This rule takes the conflict into consideration and yields more reliable segmentation results.

2.3 Multi-task loss function

Lymphomas segmentation task show sub-optimal performance in CT, which may leads to misleading fusion results in our model. In order to avoid the problem, a multi-task loss function here is proposed to exploit all the available information during training and increase segmentation accuracy. Since lymphomas are visible on both PET and CT, even though they do not have exactly the same shape in these two volumes, they should overlap as much as possible. Here we set PET mask as ground truth for both PET and CT to constrains the overlap rate between PET and CT.

As shown in Fig. 1, we define three loss functions: loss_{CT}, loss_{PET} and loss_{F} measuring the discrepancy between ground truth and, respectively, CT segmentation maps, PET segmentation maps, and the final segmentation output. For loss_{CT} and loss_{PET}, we use Dice loss,

$$\mathsf{loss}_{PET} = 1 - \frac{2 * \sum_{v=1}^{V} S_1^v G^v}{\sum_{v=1}^{V} S_1^v + \sum_{v=1}^{V} G^v}, \tag{3}$$

$$\mathsf{loss}_{CT} = 1 - \frac{2 * \sum_{v=1}^{V} S_2^v G^v}{\sum_{v=1}^{V} S_2^v + \sum_{v=1}^{V} G^v}, \tag{4}$$

where V is the number of voxels of segmentation outputs, S_1 and S_2 are the PET and CT segmentation outputs and G is the ground truth of lymphoma in PET. For loss_{F}, we use mean square loss,

$$\mathsf{loss}_{F} = \sum_{v=1}^{V} (S_f^v - G^v)^2, \tag{5}$$

where S_f is the final segmentation output. The multi-task loss function is defined as

$$\mathsf{loss}_{all} = 0.75 * \mathsf{loss}_{CT} + 0.25 * \mathsf{loss}_{PET} + \mathsf{loss}_{F}. \quad (6)$$

Here we set the weight of $loss_{CT}$ as 0.75 and the weight of $loss_{PET}$ as 0.25 to enable the model learn more from hard example (CT).

2.4 Implementation Details

All methods were implemented in Python using Tensorflow framework 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 Experiments and Analysis

3.1 Dataset and Preprocessing

The experimental dataset consists of 173 labeled cases of real PET/CT volumes, whose labels indicate the ground truth of lymphoma. All PET/CT data were acquired from the Henri Becquerel hospital. The study was approved as a retrospective study by the Henri Becquerel Center Institutional Review Board. All patients' information were de-identified and anonymized prior to analysis. The size of the CT volumes and the corresponding masks vary from $267 \times 512 \times 512$ to $478 \times 512 \times 512$ and their spatial resolution varied from $0.97 \times 0.97 \times 2\ mm^3$ to $1.36 \times 1.37 \times 5\ mm^3$. The size of the PET volumes and the corresponding masks vary from $276 \times 144 \times 144$ to $407 \times 256 \times 256$ and their spatial resolution varied from $5.3 \times 5.3 \times 2\ mm^3$ to $2.73 \times 2.73 \times 3.27\ mm^3$. For preprossessing, we first registered the CT volumes and matched them with PET volumes. Then we resized PET, CT and mask volumes into size $128 \times 256 \times 256$. Both PET and CT volumes were normalized into the standard distribution by the linear standardization method. Two kinds of data augmentation were applied here to enrich the training data: deformation and affine transformation. We used 80% of the data for training, 10% for validation and 10% for testing. The Dice score, Precision and Recall were used to evaluate the segmentation results of lymphoma at the voxel level.

3.2 Results and Discussion

The quantitative results of the experiments are represented in Table 1. UNet is the baseline method. We first tested the segmentation performance when using only PET or CT volume. Then we concatenate PET and CT volumes as inputs of the UNet. This is an input-level fusion method. Finally, we used two independent UNets for PET and CT, respectively. The fusion was carried out at decision level by the proposed evidential fusion layer. Compared with mono-modality input, our proposal shows an obvious advantage. Compared with the concatenate-based fusion method, our proposal has 3%, 4%, 6% increase in Dice score, Precision

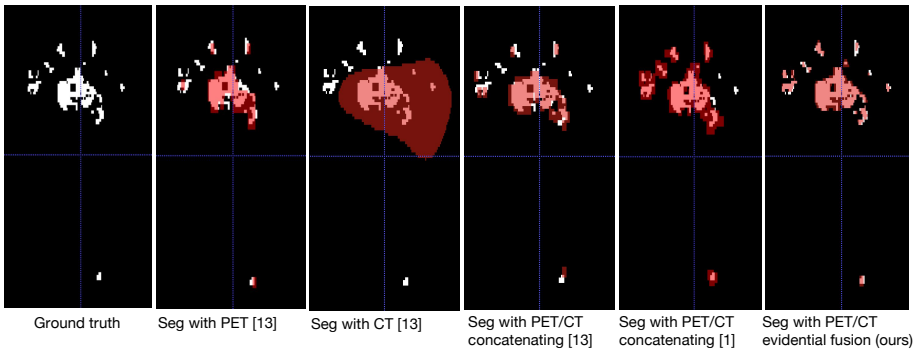


Fig. 3. Comparison of segmentation results of CT, PET, PET/CT concatenation [13], PET/CT concatenation with [1] and PET/CT evidential fusion. The ground truth is marked in white and the segmentation results in red.

and Recall, respectively. Fig. 3 shows the comparison results between the above four input modalities. We can see from the figure that mono-modality input does not allow us to segment small regions, especially for CT. The concatenation of PET and CT solves this problem to some extent [1][13], but the result is still not ideal. The evidential fusion of PET and CT achieves better segmentation result, especially for the small lymphomas located at the top of the images.

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

Models	Input Modality	Dice score	Precision	Recall
UNet (mono-input)	PET	0.67±0.02	0.74±0.05	0.65± 0.03
UNet (mono-input)	CT	0.49± 0.05	0.71± 0.02	0.45± 0.02
UNet (concatenate-based fusion)	PET+CT	0.69± 0.01	0.67± 0.01	0.74± 0.05
EUnet (evidential fusion)	PET+CT	**0.72**± 0.04	**0.71**± 0.06	**0.80**± 0.05
Zeng et al. [16]	PET+CT	0.68±0.02	0.68±0.01	0.68±0.01
Hu et al. [6]	PET+CT	0.66±0.03	0.71±0.044	0.67±0.03
Blanc et al.[1]	PET+CT	**0.73**±0.20	**0.75**±0.22	**0.83**±0.17

Since there is no public lymphoma dataset available now, the quantitative comparison with other state-of-the-art methods is difficult, because different datasets and different lymphoma types are segmented. However, we still compare our method with the state-of-the-art in Table 1. In [6], Hu et al. achieve 0.66±0.03 Dice score on a dataset of 109 lymphoma patients with a convolution fusion with 2D slices and 3D volumes. In [1], Blanc-Durand et al. report 0.73±0.20 for Dice score based on the training on 511 lymphoma patients. Though the authors report comparable results with ours, the performance of our model is more stable and needs less training data. We also test the method of [1] with our dataset and achieve 0.64±0.02 in Dice score, where the priory of

our proposal is obvious. A qualitative comparison with our model is shown in Fig. 3. See the fifth image from Fig. 3, their model show sign of slight overfitting but is visually acceptable and our model show visually correct segmentation results. Generally speaking, our model yields better results than state-of-the-art methods with DST-based evidential fusion.

A qualitative comparison is shown in Fig. 4. Here we overlap PET and CT in the same image and paste the segmented mask in this image. Column 1 shows one slice of patient 1 in which a big lymphoma is present. Our model can segment it correctly. Column 2 shows a slice of patient 2 in which multiple lymphomas are present. They are more difficult to segment because some of them are very small. However, our model can still obtain satisfactory results.

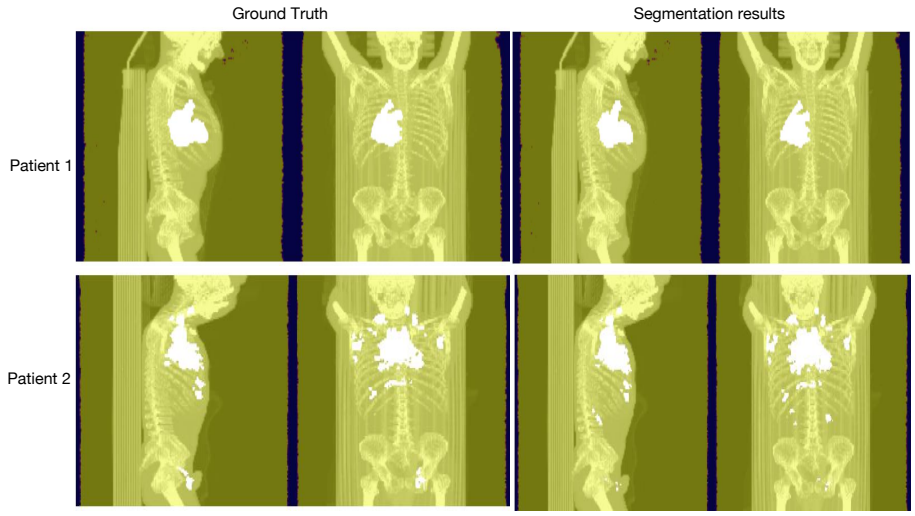


Fig. 4. Illustration of the results of two patients obtained using our EUNet model. From left to right, each column represents the corresponding segmentation results and the ground truth. The lymphomas are showed in white.

4 Conclusion

In this work, a DST-based deep multi-modality medical images fusion strategy has been proposed to deal with the problem of uncertainty and conflict within a deep convolutional neural network. A two-branch segmentation module first processes PET/CT volumes separately. Then the two segmentation maps are fused by the evidential fusion layer. Qualitative and quantitative evaluation results are promising as compared to the baseline and state-of-the-art methods. Future research will aim at improving the network architecture to better exploit the potential of DST to represent and combine uncertain information.

References

1. Blanc-Durand, P., Jégou, S., Kanoun, S., Berriolo-Riedinger, A., Bodet-Milin, C., Kraeber-Bodéré, F., Carlier, T., Le Gouill, S., Casasnovas, R.O., Meignan, M., 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. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
3. Dempster, A.P.: Upper and lower probability inferences based on a sample from a finite univariate population. Biometrika **54**(3-4), 515–528 (1967)
4. Desbordes, P., Petitjean, C., Ruan, S.: 3d automated lymphoma segmentation in pet images based on cellular automata. In: 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2014)
5. Hershey, S., Chaudhuri, S., Ellis, D.P., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., et al.: Cnn architectures for large-scale audio classification. In: 2017 ieee international conference on acoustics, speech and signal processing (icassp). pp. 131–135. IEEE (2017)
6. Hu, H., Shen, L., Zhou, T., Decazes, P., Vera, P., Ruan, S.: Lymphoma segmentation in pet images based on multi-view and conv3d fusion strategy. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 1197–1200. IEEE (2020)
7. Li, H., Jiang, H., Li, S., Wang, M., Wang, Z., Lu, G., Guo, J., Wang, Y.: Densexnet: an end-to-end model for lymphoma segmentation in whole-body pet/ct images. IEEE Access **8**, 8004–8018 (2019)
8. Li, H., Thorstad, W.L., Biehl, K.J., Laforest, R., Su, Y., Shoghi, K.I., Donnelly, E.D., Low, D.A., Lu, W.: A novel pet tumor delineation method based on adaptive region-growing and dual-front active contours. Medical physics **35**(8), 3711–3721 (2008)
9. Lian, C., Ruan, S., Denœux, T., Li, H., Vera, P.: Joint tumor segmentation in pet-ct images using co-clustering and fusion based on belief functions. IEEE Transactions on Image Processing **28**(2), 755–766 (2018)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440. Boston, USA (Jun, 2015)
11. Nestle, U., Kremp, S., Schaefer-Schuler, A., Sebastian-Welsch, C., Hellwig, D., Rübe, C., Kirsch, C.M.: Comparison of different methods for delineation of 18f-fdg pet–positive tissue for target volume definition in radiotherapy of patients with non–small cell lung cancer. Journal of nuclear medicine **46**(8), 1342–1348 (2005)
12. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
13. Ronneberger, O., Fischer, P., Brox, T.n.: Convolutional networks for biomedical image segmentation. In: Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany (Oct, 2015)
14. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)

15. Sundararajan, S.K., Sankaragomathi, B., Priya, D.S.: Deep belief cnn feature representation based content based image retrieval for medical images. Journal of medical systems **43**(6), 1–9 (2019)
16. Zeng, G., Yang, X., Li, J., Yu, L., Heng, P.A., Zheng, G.: 3d u-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3d mr images. In: International workshop on machine learning in medical imaging. pp. 274–282. Springer (2017)
17. Zhou, T., Ruan, S., Canu, S.: A review: Deep learning for medical image segmentation using multi-modality fusion. Array **3**, 100004 (2019)
18. Zhu, W., Huang, Y., Tang, H., Qian, Z., Du, N., Fan, W., Xie, X.: Anatomynet: Deep 3D squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation. bioRxiv p. 392969 (2018)