HHS Public Access

Author manuscript

Neurocomputing. Author manuscript; available in PMC 2021 June 07.

Published in final edited form as:

Neurocomputing. 2020 June 7; 392: 277–295. doi:10.1016/j.neucom.2018.10.099.

Deep Learning for Variational Multimodality Tumor Segmentation in PET/CT

Laquan Li[1,2], **Xiangming Zhao**[1], **Wei Lu**[3], **Shan Tan**[1,a)]

[1]Key Laboratory of Image Processing and Intelligent Control of Ministry of Education of China, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

[2]College of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[3]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA

Abstract

Positron emission tomography/computed tomography (PET/CT) imaging can simultaneously acquire functional metabolic information and anatomical information of the human body. How to rationally fuse the complementary information in PET/CT for accurate tumor segmentation is challenging. In this study, a novel deep learning based variational method was proposed to automatically fuse multimodality information for tumor segmentation in PET/CT. A 3D fully convolutional network (FCN) was first designed and trained to produce a probability map from the CT image. The learnt probability map describes the probability of each CT voxel belonging to the tumor or the background, and roughly distinguishes the tumor from its surrounding soft tissues. A fuzzy variational model was then proposed to incorporate the probability map and the PET intensity image for an accurate multimodality tumor segmentation, where the probability map acted as a membership degree prior. A split Bregman algorithm was used to minimize the variational model. The proposed method was validated on a non-small cell lung cancer dataset with 84 PET/CT images. Experimental results demonstrated that: 1). Only a few training samples were needed for training the designed network to produce the probability map; 2). The proposed method can be applied to small datasets, normally seen in clinic research; 3). The proposed method successfully fused the complementary information in PET/CT, and outperformed two existing deep learning-based multimodality segmentation methods and other multimodality segmentation methods using traditional fusion strategies (without deep learning); 4). The proposed method had a good performance for tumor segmentation, even for those with Fluorodeoxyglucose (FDG) uptake inhomogeneity and blurred tumor edges (two major challenges in PET single modality segmentation) and complex surrounding soft tissues (one major challenge in CT single modality segmentation), and achieved an average dice similarity indexes (DSI) of 0.86 ± 0.05,

a) Author to whom correspondence should be addressed. shantan@hust.edu.cn; Telephone: +86-27-87558572.

sensitivity (SE) of 0.86 ± 0.07, positive predictive value (PPV) of 0.87 ± 0.10, volume error (VE) of 0.16 ± 0.12, and classification error (CE) of 0.30 ± 0.12.

Keywords

Tumor segmentation; PET/CT images; variational method; deep learning; information fusion

1. Introduction

Positron emission tomography (PET) with ^{18}F-fluorodeoxyglucose (FDG) has been widely used in clinical practice. PET imaging extracts the functional and metabolic information of the human body. Compared with other imaging modalities, such as the magnetic resonance imaging (MRI) and computed tomography (CT), PET can be used for an earlier diagnosis of the disease. Because of a faster metabolism and more ^{1S}F-FDG uptake, the diseased areas appear as 'hot' in PET images [1]. The contrast is typically high between malignant tumors and normal tissues in PET (see Fig. 1(a) and (c) for examples). However, the PET scanners usually have a low spatial resolution, which results in blurred tumor edges (Fig. 1(e) and (g)) [2]. Besides, a tumor may present intensity inhomogeneity in PET (Fig. 1(c) or (g)). Both of these make it difficult to accurately delineate a tumor from a single PET image.

CT imaging provides detailed anatomical information of the human body. It has a higher spatial resolution than PET. There are clear boundaries between the tumors and the non-soft tissues in CT images (Fig. 1(f) and (h)). However, the intensity of the tumor is often similar to that of the normal soft tissues in CT. When a tumor invades into the adjacent normal soft tissues, it is very difficult to tell the tumor from the normal tissues in CT. The lung cancer is a typical example of this kind: a lung tumor can attach to the thoracic vertebrae or the chest wall (Fig. 1(b) and (d)). In addition, the background in a CT image can be very complex, which makes tumor segmentation even more difficult. For these reasons, most of traditional methods cannot work well for tumor segmentation using a single CT image.

In recent years, multimodal imaging techniques, such as PET/CT, MRI/CT, MRI/PET, and MRI/SPECT, have been used in clinic to enhance disease diagnosis and treatment assessment [3-5]. These techniques provide different physical or biological aspects for observing a tumor. Fusion strategies have been proposed to utilize the complementary information from multimodality images [6-8]. For tumor segmentation, rational use of the complementary information between different modality images can improve the tumor segmentation accuracy. Over the last decade, some multimodality tumor segmentation methods have been proposed. To identify the biophysical structure, EI Naqa *et al.* proposed a multivalued level set model based variational method to integrate information from multimodality images [9]. A geometric level set contour based multimodality segmentation algorithm was studied in [10], where the Jensen-Rényi divergence was used to evolve the level set contour. In [11], a variety of 3D texture features were extracted to train a combined decision tree (DT) with K-nearest neighbor (KNN) classifier for multimodality image segmentation. Han *et al.* constructed a PET/CT segmentation formulation based on the Markov Random Field optimization problem [12]. Song *et al.* designed an adaptive context

term for the objective function to achieve consistent segmentation results between PET and CT [13]. Ju *et al*. used a random walk method as an initial preprocessor to obtain object seeds, and a graph cut method was then used for lung tumor segmentation on PET/CT images [1]. Bagci *et al*. built an automatic random walk based multimodality image segmentation method [14]. This method combined different modality images using a hyper graph construction. Cui *et al*. proposed a 'topo-poly' graph model for PET/CT lung tumor segmentation. This model combined joint foreground density similarity and topology information between PET and CT [15]. They also proposed a label information propagation model for tumor segmentation in PET/CT [16]. This model correlated the local intensity changes of CT and the spatial–topological information from PET. All these studies mentioned above demonstrated that combining the information of multimodality images could obtain more accurate tumor segmentation result. Actually, there are a lot of inconsistent information between the different modality images due to the different imaging mechanisms. The unreasonable use of these inconsistent information will reduce the segmentation accuracy instead of improving it. The existing multimodality segmentation methods usually obtained a compromise segmentation result for the inconsistent information, which could not make full use of the superior information of each modality image.

The similar intensity distribution between the tumor and its surrounding normal soft tissues and the complex background in CT images always introduce a lot of inconsistent information for the PET and CT images. The traditional (*e.g.*, region based and edge based) segmentation methods are unreliable to separate the tumor from the normal soft tissues. In the past several years, deep learning demonstrated great success in image recognition [17-19], object detection [20], object segmentation [21] and medical image processing [22], such as single-modality CT image segmentation [23-25], multi-channel MRI image segmentation [26-28], standard-dose PET image estimation from low-dose PET/MRI [29], and automated whole-body bone lesion detection from PET/CT images [30]. In the first MICCAI (Medical Image Computing and Computer Assisted Intervention) challenge on PET tumor segmentation, a method based on convolutional neural networks (CNN) won the challenge [31]. The concept of deep learning originated in the research of artificial neural network (ANN). Each deep learning network usually consists of multiple processing layers. It can learn representations of data with multiple levels of abstraction and is very good at discovering intricate structures in high-dimensional data [32].

In view of these good performance of deep learning, we first designed a 3D fully convolutional network (FCN) to produce a probability map in an end-to-end way from the CT image that indicates the probability of each CT voxel belonging to the tumor or the background. For natural images, such a probability map can be considered as a result for object segmentation by its own [33, 34]. It is widely known that the good performance of the end-to-end deep learning method typically depends on a large number of labeled training data. However, acquiring such amount of training data in medical imaging is typically infeasible. As a result, the CT probability map extracted by FCN might not be accurate enough when directly used as a tumor segmentation result. But it removed a great deal of redundant information in the CT image. Fusing the obtained CT probability map with the PET image information would be more conducive to tumor segmentation than directly

fusing the CT image information. We then designed a fuzzy variational model to adaptively combine the CT probability map and the PET image for a more refined tumor segmentation. The variational model consists of a fidelity term constructed using the PET image and a fuzzy membership function, a length measurement term of tumor edges, and a fuzzy membership degree prior term constructed using the CT probability map. The proposed fuzzy variational model not only successfully fused the CT and PET information, but also was specially designed to deal with the blurred tumor edges normally seen in a PET image for more accurate tumor segmentation. Combining the advantages of deep learning and variational segmentation method, the proposed method can accurately delineate the tumors and need not large amount of training samples.

A standard benchmark platform for the evaluation and validation of PET-auto-segmentation methods following the recommendations of the American Association of Physicists in Medicine (AAPM) task group No. 211 [35] was described in [36]. This tool is convenient to evaluate the performance of the PET-auto-segmentation methods. Unfortunately, the current benchmark platform can not evaluate the PET/CT co-segmentation methods. In this study, we validated the performance of the proposed method on a non-small cell lung cancer dataset with 84 PET/CT images. Experimental results showed that the proposed method successfully fused the complementary information between PET and CT. The large probability values of voxels belonging to the tumor obtained from the CT image overcame the impact of tumor intensity inhomogeneity in PET to tumor segmentation. The proposed method could correctly classify the ‘cold’ area (with low intensity) in tumor in PET as foreground (tumor). The small probability values of voxel belonging to the tumor obtained from CT was helpful to correctly classify the ‘hot’ area (normal tissues with high ^{18}F-FDG uptake) in the background of PET as background. On the other hand, the high contrast between tumor and background in PET is helpful to eliminate the influence of the possible error introduced by the fuzzy membership degree prior from the CT probability map. The proposed method worked better for tumor segmentation than deep learning methods over a single modality (either PET or CT). Experimental results also demonstrated that the proposed method worked better than multimodality tumor segmentation methods with traditional fusion strategies (without deep learning) and two recently proposed deep learning-based multimodality segmentation methods using W-net [30] and 3D-UNet [37] respectively.

2. Related Work

2.1. Deep Learning Methods for CT Image Segmentation

Recent advances in deep learning have made a big leap to help solve the difficult problems in CT image segmentation by exploiting hierarchical feature representations learned from data [22]. Deep convolutional neural networks (CNNs), one type of deep learning model, provide enhanced performance in object segmentation. Dou *et al.* [25] proposed a novel deeply supervised network for 3D liver automatic segmentation in CT. Segmentation for pancreas tumor and liver tumor on CT images using deep CNNs were studied in [24] and [38], respectively. Typically, CNNs based segmentation systems do not fully utilize image spatial and configuration information. To better utilize these information, FCN was proposed

by Long *et al.* with the backbone of CNNs, achieving impressive results in multi-object semantic natural image segmentation [39].

FCN based methods such as U-Net [40] and V-Net [41] have been studied very recently for medical image segmentation. Due to the complex intensity and geometry variation in CT [42], additional procedures were often introduced to help enhance these deep learning methods to achieve more accurate and robust results [23, 43]. For example, Hu *et al.* [43] combined surface evolution and deep learning for liver segmentation. Conditional random fields based post-processing was used after the cascaded FCN to automatically segment liver and lesion in CT [23]. In this study, we used FCN to learn the probability map of lung tumors from CT, and integrated the probability map into a fuzzy variational model containing the PET image information to refine the tumor segmentation results.

2.2. Fuzzy Energy Based Variational Models

Fuzzy energy based methods allow the pixels to be classified to several regions simultaneously with probability in [0, 1] [44]. They are flexible and have been widely used for image segmentation. Krinidis and Chatzis proposed a fuzzy energy-based active contour model to detect objects whose boundaries are not necessarily defined by gradient [45]. Shyu *et al.* introduced a global and local fuzzy energy-based active contours for blurred and intensity inhomogeneity image segmentation [46]. Tran *et al.* proposed an active contour segmentation model with a fuzzy energy and a shape prior [47]. Wu *et al.* designed a region-based fuzzy active contour model with kernel metric for a robust and stable image segmentation [48]. This model could detect the boundaries precisely and worked well with images in the presence of noise, outliers and low contrast. Li *et al.* proposed a variational multiphase image segmentation model based on fuzzy membership functions and L_1-norm fidelity [44]. These methods demonstrated that the fuzzy energy-based method had a good performance for blurred image segmentation. In this study, we designed a fuzzy membership function based variational model to handle the blurred tumor edges normally seen in a PET image.

2.3. Lung Tumor Segmentation

Lung cancer is one of the most common malignant tumors, which seriously threaten people's health and life. Accurate tumor segmentation plays a very important part in radiation therapy. Lung tumor segmentation has been extensively studied. As mentioned in Introduction, Cui *et al.* proposed a 'topo-poly' graph model [15] and a label information propagation model [16] for PET/CT lung tumor segmentation. Ju *et al.* proposed a lung tumor segmentation method based on random walk and graph cut [1]. In [49], Jeba *et al.* reviewed on various PET/CT lung tumor segmentation methods proposed before 2017.

In recent two years, some deep-learning-based lung tumor segmentation methods have been proposed. Jiang *et al.* proposed two multiple resolution residually connected network formulations for automatic lung tumor segmentation from CT images [50], Zhong *et al.* proposed a lung tumor segmentation method (3D-UNet+GC) by combining the FCN based semantic segmentation framework (3D-UNet) and Song's graph cut based co-segmentation model [37]. First, two separate deep Unets was constructed to extract the probability maps

from the PET and CT images respectively. Then, the obtained two probability maps were employed in the graph cut based co-segmentation model to produce the final segmentation results. Different from Zhong's method, the method presented in this paper extracted the probability map only from the CT images, and the obtained probability map was used as the membership degree prior in the new designed variational model, which made the proposed method more flexible and stable.

3. Methods

The proposed method was constructed based on a deep learning method and a novel fuzzy variational model. Firstly, a 3D FCN with the V-Net architecture was designed and trained on CT images to learn a probability map describing the probability of each CT voxel belonging to the tumor or the background. Generally, FCN needs a large amount (thousands or even millions) of labeled training data to get accurate tumor segmentation. However, acquiring that amount of training data in medical imaging is typically infeasible. The probability map obtained from the CT image with limited training samples was not enough to accurately delineate the tumor. Taking into account the flexibility of variational model in fusing different image information, we then designed a fuzzy variational model to integrate the obtained probability map and the image intensity information of PET to refine the tumor segmentation. The final segmentation result was acquired by minimizing there proposed variational model using a split Bregman algorithm, and the probability map was also used as the initialization of the fuzzy membership function in the variational model.

3.1. 3D FCN for CT Image Processing

3.1.1. Network Architecture—CT imaging has low contrast, and the tumor intensity may be similar to that of the surrounding normal tissues. It is difficult to delineate a tumor from the complex background of the CT image using traditional methods. Take the lung cancer as an example. In lung cancer, a tumor may invade into the chest wall, the mediastinal structure or the diaphragm [13]. When this happens, there is barely a clear boundary between the tumor and its surrounding normal tissues, and most traditional methods especially those over single imaging modality would fail to delineate the tumor.

By extracting higher-dimensional and more abstract features, deep learning methods have achieved state-of-the-art performance in many applications. In our study, we designed a 3D deep learning network based on FCN to produce a probability map, in order to extract CT information for our multimodality tumor segmentation method.

As shown in Fig. 2, the network includes a down-sampling path and an up-sampling path. The down-sampling path can be divided into three stages. Each stage includes two or three convolutional blocks. For each convolutional block, kernels of size 3×3×3 voxels and the stride of 1 were employed to perform the convolution operation. In order to keep the size unchanged, appropriate padding operation were also performed at the same time. The down-sampling process was completed by convolution with kernels of size 2×2×2 and the stride of 2. Along with the down-sampling process, the number of channels increased two or four times, and the size of the features map decreased by half. Correspondingly, the structure of the up-sampling path was almost symmetrical with the down-sampling path. The up-

sampling path also consisted of three convolutional blocks, where the size of kernels was 3×3×3 with stride 1. During the convolution process, the size of the feature maps and the number of channels also remained unchanged. The size of the feature maps increased due to the de-convolutional operation with kernels of size 2×2×2 and a stride of 2. The number of channels kept the same as the corresponding down-sampling path. At the end of the network, convolution with kernels size 1×1×1 was performed, which led to the number of channels of the feature maps be 2 (foreground and background). The probability maps were output by calculating the softmax function. Moreover, rectified linear units were applied in this network and dropout was also employed to prevent the network from overfitting. The foreground probability map would be used as a membership degree prior of the tumor and initial membership function in the next step of the proposed method.

3.1.2. Implementation Details—In our study, the V-Net style network has been improved so that it can accept inputs of any size. Data augmentations, including adding Gaussian noise and rotation, were performed to increase the number of samples. Considering the GPU memory limitation and the various size of training sample, the batch size was set to 1. For the gradient descent optimization algorithms, the Adam optimizer with a stable learning rate of 0.001 was used. The rate of dropout was 0.2.

The network was implemented using Python and the Tensorflow library. It was trained on Nvidia GTX 1080 Ti GPU about 350 epochs and took about 3 hours. The number of epoch was a hyperparameter in the network, which was adjusted based on the learning curve. The hyperparameter was determined when the loss remained stable.

3.2. A Novel Fuzzy Variational Model for Tumor Segmentation

3.2.1. The Model Description—The variational model is powerful and convenient to consider prior information in various image-related application frameworks. In this study, we integrated the probability map obtained by FCN from CT and the image intensity information of PET into a novel variational model for multimodality tumor segmentation. Note that PET imaging has a low spatial resolution, and tumor edges in a PET image are usually blurred, which is one of the major challenges for tumor segmentation in PET. Fuzzy energy based methods work well with images with weak, blurred or even discontinuous boundaries [44-48]. In this study, we designed a novel variational model based on the fuzzy membership function.

The energy functional of the proposed model was defined as:

$$F\ c_1, c_2, u = F_{fidelity}\ c_1, c_2, u + \alpha F_{length}\ u + \beta F_{prior}\ u\,, \tag{1}$$

where α, β are fixed positive parameters; c_1 and c_2 are average uptake prototypes of the tumor area and the background area in a PET image, respectively; $u \in [0,1]$ is a fuzzy membership function, $u(x, y, z)$ and $1-u(x, y, z)$ represent the membership value that the voxel at point (x, y, z) belongs to the tumor and the background, respectively.

$F_{fidelity}$ is the data fidelity term based on the PET image and the fuzzy membership function u. Let $\Omega \subset R$ be the image domain, and $I: \Omega \rightarrow R$ be a PET image. The fidelity term can be expressed as:

$$F_{fidelity}(c_1, c_2, u) = \int_{\Omega} I(x,y,z) - c_1)^2 u(x,y,z)^2 dxdydz + \int_{\Omega} I(x,y,z) - c_2)^2 (1-u(x,y,z))^2 dxdydz. \quad (2)$$

F_{length} is the total variational (TV) regularization term for the fuzzy membership function u, which can be written as:

$$F_{length}(u) = \int_{\Omega} |\nabla u(x,y,z)| dxdydz, \quad (3)$$

and controls the length of the boundaries of the segmented areas. This term is important to the proposed model since PET images are usually seriously corrupted by noise.

F_{prior} is a membership degree prior term. It integrated the probability map function P obtained by FCN from the corresponding CT image into the energy functional:

$$F_{prior}(u) = \int_{\Omega} \left(\frac{1-P(x,y,z)}{P(x,y,z)} u(x,y,z) + \frac{P(x,y,z)}{1-P(x,y,z)} (1-u(x,y,z)) \right) dxdydz. \quad (4)$$

Let $w_1 = (1-P(x,y,z))/P(x,y,z)$ and $w_2 = P(x,y,z)/(1-P(x,y,z))$. If the probability prior $P(x, y, z)$ is large, which indicates that the voxel has a higher probability of belonging to the tumor in the CT image, w_2 would be much bigger than w_1. This makes $u(x, y, z)$ get a larger value in minimizing the energy functional Eq. (1). In contrast, a small probability prior leads to a smaller membership value. In other words, the prior term leads the variational model to automatically achieve balanced tumor segmentation result between the PET and CT images. Especially, when $P(x, y, z) \rightarrow 0$ or $P(x, y, z) \rightarrow 1$, w_1 or w_2 has a large value. In this case, the prior term plays a large role in acquiring the membership function u.

Substituting Eqs. (2) (3) and (4) into Eq. (1), the proposed model could be rewritten as:

$$F(c_1, c_2, u) = \int_{\Omega} I(x,y,z) - c_1)^2 u(x,y,z)^2 dxdydz + \int_{\Omega} I(x,y,z) - c_2)^2 (1-u(x,y,z))^2 dxdydz + \alpha \int_{\Omega} |\nabla u(x,y,z)| dxdydz + \beta \int_{\Omega} \left(\frac{1-P(x,y,z)}{P(x,y,z)} u(x,y,z) + \frac{P(x,y,z)}{1-P(x,y,z)} (1-u(x,y,z)) \right) dxdydz. \quad (5)$$

3.2.2. Energy Minimization—Before minimizing the energy functional, we up-sampled the PET image to have the same voxel dimensionality as the CT image. This was to obtain a

one-to-one voxel correspondence between the PET image and the probability map P. In our study, a cubic B-spline interpolator method was used for the up-sampling.

In the process of energy minimization, we firstly fixed the fuzzy membership function u and minimized the energy functional (5) with respect to c_1 and c_2. Taking the derivative of Eq. (5) with respect to c_1 and c_2 respectively, it was easy to get the updating equations of c_1 and c_2:

$$c_1 = \frac{\int_\Omega (u(x,y,z))^2 I(x,y,z)dxdydz}{\int_\Omega (u(x,y,z))^2 dxdydz}, \tag{6}$$

$$c_2 = \frac{\int_\Omega (1-u(x,y,z))^2 I(x,y,z)dxdydz}{\int_\Omega (1-u(x,y,z))^2 dxdydz}. \tag{7}$$

Then we fixed c_1 and c_2 and minimized the energy functional (5) with respect to u using a split Bregman algorithm. The algorithm details were described in the next section.

3.2.3. A Split Bregman Algorithm for u—The energy functional (5) is convex with respect to the fuzzy membership function u. After each update of c_1 and c_2, we used a split Bregman method [51] to find the global optimal solution u^* of the energy functional.

Note that minimizing the functional (5) with respect to u is equivalent to optimizing the following convex energy functional:

$$\min_{u \in [0,1]} F\,u = \int_\Omega r_1 u^2 + r_2 u \; dxdydz + \alpha \int_\Omega \mid \nabla u \mid dxdydz, \tag{8}$$

where

$$r_1 = I\,x,y,z - c_1{}^2 + I\,x,y,z - c_2{}^2, \tag{9}$$

and

$$r_2 = \beta\left[\frac{1-P\,x,y,z}{P\,x,y,z} - \frac{P\,x,y,z}{1-P\,x,y,z}\right] - 2\,I\,x,y,z - c_2{}^2. \tag{10}$$

We introduced an auxiliary variable d and the equality constraint $d = \nabla u$. The optimization problem in Eq. (8) became as:

$$u^*, d^* = \arg\min_{u \in [0,1],\, d} F\,u = \int_\Omega r_1 u^2 + r_2 u \; dxdydz + \alpha \int_\Omega \mid d \mid dxdydz + \frac{\lambda}{2}\int_\Omega \|d - \nabla u\|^2 \, dxdydz, \tag{11}$$

where λ is a positive penalty parameter and (u^*, d^*) denotes the global optimal solution. We performed this minimization by splitting the L^1 and L^2 components of functional (11) with

respect to u and d separately. Two elegant iterations in the split Bregman method can be obtained:

$$u^{n+1}, d^{n+1} = \underset{u \in [0,1], d}{\arg\min} \left\{ \int_\Omega r_1 u^2 + r_2 u \, dxdydz + \alpha \int_\Omega | d | \, dxdydz + \frac{\lambda}{2} \int_\Omega \| d - \nabla u - b^n \|^2 \, dxdydz \right\}, \tag{12}$$

and

$$b^{n+1} = b^n + \nabla u^{n+1} - d^{n+1}. \tag{13}$$

We needed to perform two sub-optimization problems for Eq. (12):

PI:

$$u^{n+1} = \underset{u \in [0,1], d}{\arg\min} \left\{ \int_\Omega r_1 u^2 + r_2 u \, dxdydz + \frac{\lambda}{2} \int_\Omega \| d^n - \nabla u - b^n \|^2 \, dxdydz \right\}, \tag{14}$$

P2:

$$d^{n+1} = \underset{d}{\arg\min} \left\{ \alpha \int_\Omega | d | \, dxdydz + \frac{\lambda}{2} \int_\Omega \| d - \nabla u^{n+1} - b^n \|^2 \, dxdydz \right\}. \tag{15}$$

For P1, the Euler-Lagrange equation of the optimization problem with respect to u is

$$-2r_1 + \lambda \Delta \, u^{n+1} = r_2 + \lambda \nabla_x \, d^n - b^n + \lambda \nabla_y \, d^n - b^n + \lambda \nabla_z \, d^n - b^n. \tag{16}$$

The central difference was used for the Laplace operator (Δ) and the backward difference was used for the gradient operator. With the Gauss-Seidel formula, the solution at the $(n+1)^{th}$ iteration was given as

$$u^{n+1}_{i,j,k} = -\frac{r_{2,i,j,k}}{2r_{1,i,j,k} + 6\lambda} - \frac{\lambda}{2r_{1,i,j,k} + 6\lambda} \Phi^n_{i,j,k} - u^n_{i-1,j,k} - u^n_{i+1,j,k} - u^n_{i,j-1,k} - u^n_{i,j+1,k} - u^n_{i,j,k-1} - u^n_{i,j,k+1}, \tag{17}$$

where

$$\Phi^n_{i,j,k} = d^{n,x}_{i,j,k} - d^{n,x}_{i-1,j,k} - b^{n,x}_{i,j,k} + b^{n,x}_{i-1,j,k} + d^{n,y}_{i,j,k} - d^{n,y}_{i,j-1,k} - b^{n,y}_{i,j,k} + b^{n,y}_{i,j,k-1} + d^{n,z}_{i,j,k} - d^{n,z}_{i,j,k-1} - b^{n,z}_{i,j,k} + b^{n,z}_{i,j,k-1}. \tag{18}$$

To be restricted within the solution space, the solution u was truncated as

$$u^{n+1} = \max \min u^{n+1}, 1, 0 \ . \tag{19}$$

For P2, since the function $f(x) = |x|$ is differentiable in $R\backslash\{0\}$, the optimization problem (15) with respect to d at the $(n+1)^{th}$ iteration is:

$$d^{n+1} = shrink\left(b^n + \nabla u^{n+1}, \frac{\alpha}{\lambda}\right), \tag{20}$$

where the shrinkage operator was given as:

$$shrink\ x, y = \begin{cases} \frac{x}{|x|}\max\ |x| - y, 0\ , x \neq 0 \\ 0, \qquad\qquad x = 0 \end{cases}. \tag{21}$$

The split Bregman algorithm of the proposed method was summarized in Algorithm 1. After the split Bregman iteration, the tumor region and background region were given as $\{(x, y, z) \in \Omega | u^*(x, y, z) \geq 0.5\}$ and $\{(x,y,z) \in \Omega | u^*(x,y,z) < 0.5\}$, respectively.

The above optimization process was performed in Matlab. It took about 15 seconds to segment one tumor on a computer with 3.40 GHz CPU and 8GB RAM.

Algorithm 1

Initialization: $u^0 = P$ and $\varepsilon = 1\times10^{-6}$
while $\left\|u^{n+1} - u^n\right\| > \varepsilon$
  Compute c_1 and c_2 as Eqs. (6) and (7);
  Compute r_1 and r_2 as Eqs. (9) and (10);
  Update u^{n+1} using Eqs. (17) and (19);
  Update d^{n+1} using Eq. (20);
  Update b^{n+1} using Eq. (13);
End
Output: u^*

4. Experimental Settings

4.1. Datasets

The performance of the proposed method was validated on a clinical dataset with 84 non-small cell lung cancer patients. 48 randomly selected CT images were used to train the FCN, and the remaining 36 PET/CT images were used as the test dataset.

All 3D PET/CT images were obtained on a PET/CT scanner (Reveal HD, CTI, Knoxville, TN, USA). Each pair of PET/CT images has been registered through the hardware in the PET/CT scanner. The transaxial resolution of the PET scanner increased from 4.6 mm (full-width at half-maximum, FWHM) at 1 cm from the center to 7.8 mm at a radius of 20 cm. The axial resolution increased from 3.5 mm at the center to 7.8 mm at a radius of 20 cm [52, 53]. The PET images were reconstructed using the iterative Ordered Subset Expectation Maximization (OSEM) algorithm with 8 subsets and 2 iterations. The dimensionality of each reconstructed PET slice was 128×128, with a voxel size 4.69×4.69×3.27 mm^3. For the CT images, the dimensionality of each reconstructed slice was 512×512, with a voxel size

$0.98\times0.98\times3.27\ mm^3$. The tumor volumes in the training and test datasets range from 3.82 ml to 313.69 ml and 1.59 ml to 526.77 ml respectively.

The ground truth for each pair of the PET/CT images in the datasets was manually delineated by an experienced radiation oncologist. As with the existing lung tumor segmentation methods [15, 16, 37, 50], a rectangular ROI was manually defined for each PET/CT images, which enclosed the whole tumor. The proposed method and all comparison algorithms were implemented in the ROI.

4.2. Comparison Algorithms

To determine the contribution of multimodality information fusion to tumor segmentation, we compared the proposed method with the FCN on CT images (FCN_CT) — thresholding the probability map by 0.5, and the fuzzy variational model — the proposed variational model on PET image without the membership degree prior term from CT (FVM_PET), which was defined as:

$$F\left(c_1, c_2, u\right) = \int_{\Omega} I\left(x, y, z\right) - c_1{}^2 u\left(x, y, z\right)^2 dxdydz + \int_{\Omega} I\left(x, y, z\right) - c_2{}^2 1 - u\left(x, y, z\right)^2 dxdydz + \alpha \int_{\Omega} \left| \nabla u\left(x, y, z\right) \right| dxdydz. \tag{22}$$

We also compared the proposed method with two existing deep learning-based multimodality segmentation method — W-net [30] and 3D-UNet+GC [37], and two traditional variational multimodality tumor segmentation methods (without deep learning) which fused the PET and CT image information with traditional image fusion strategies. The W-Net architecture was constructed by cascading two V-Nets. This network can extract features from both PET and CT images, and all these features were used in unison to delineate the tumors. As mentioned in Section 2.3, the 3D-UNet+GC method combined the fully convolutional networks (3D-UNet) and Song's graph cut based co-segmentation model [13]. Traditionally in multimodality segmentation, one of the most widely used fusion strategy considered the PET and CT images as two channels of a hyper-image (*e.g.*, the methods in [9] and [14]). Another widely used strategy fused the PET and CT information by constraining the consistency of the segmentation results of the two modalities (*e.g.*, the method proposed by the Song *et al.* [13]). The two variational multimodality segmentation methods used for comparison in this study were constructed based on the fuzzy clustering theory and the two widely used fusion strategy respectively. The model of the first multimodality method (FVM_CO_1) with the hyper-image strategy was formulated as:

$$
\begin{aligned}
&F\left(u\left(x,y,z\right),c_{1\_PET},c_{1\_CT},c_{2\_PET},c_{2\_CT}\right)\\
&=\int_{\Omega}\eta\left(I_{PET}\left(x,y,z\right)-c_{1\_PET}\right)^{2}+\left(1-\eta\right)\left(I_{CT}\left(x,y,z\right)-c_{1\_CT}\right)^{2}u\left(x,y,z\right)^{2}dxdydz\\
&\quad+\int_{\Omega}\xi\left(I_{PET}\left(x,y,z\right)-c_{2\_PET}\right)^{2}+\left(1-\xi\right)\left(I_{CT}\left(x,y,z\right)-c_{2\_CT}\right)^{2}\left(1-u\left(x,y,z\right)\right)^{2}dxdydz\\
&\quad+\alpha\int_{\Omega}\left|\nabla u\left(x,y,z\right)\right|\\
&dxdydz,
\end{aligned}
\tag{23}
$$

where η and ξ are two positive parameters assigning the weight of PET and CT, I_{PET} and I_{CT} denote the PET image and CT image respectively, c_{1_PET} (c_{1_CT}) and c_{2_PET} (c_{2_CT}) are average prototypes of the tumor area and the background area of PET (CT) image, $u(x, y, z)$ denotes a fuzzy membership function (like the proposed method). This model is exactly a bi-modality extension of the classical Active Contours without edges model (AC) [54] and also a convex approximation of the El Naqa's multimodality segmentation model [9]. The model of the second multimodality method method (FVM_CO_2) with the another fusion strategy was formulated as

$$
\begin{aligned}
&F\left(u_1\left(x,y,z\right),u_2\left(x,y,z\right),c_{1\_PET},c_{1\_CT},c_{2\_PET},c_{2\_CT}\right)\\
&=\zeta\int_{\Omega}I_{PET}\left(x,y,z\right)-c_{1\_PET}\right)^{2}u_1\left(x,y,z\right)^{2}dxdydz\\
&+\zeta\int_{\Omega}I_{PET}\left(x,y,z\right)-c_{2\_PET}\right)^{2}\left(1-u_1\left(x,y,z\right)^{2}dxdydz+\alpha_1\int_{\Omega}\left|\nabla u_1\left(x,y,z\right)\right|dxdydz\\
&+\left(1-\zeta\right)\int_{\Omega}I_{CT}\left(x,y,z\right)-c_{1\_CT}\right)^{2}u_2\left(x,y,z\right)^{2}dxdydz\\
&+\left(1-\zeta\right)\int_{\Omega}I_{CT}\left(x,y,z\right)-c_{2\_CT}\right)^{2}\left(1-u_2\left(x,y,z\right)^{2}dxdydz+\alpha_2\int_{\Omega}\left|\nabla u_2(x,y,z)\right|dxdydz\\
&+u\int_{\Omega}\left(u_1(x,y,z)-u_2(x,y,z)\right)^{2}dxdydz,
\end{aligned}
\tag{24}
$$

where ζ, μ, α_1 and α_2 are positive parameters and ζ balances the weight of PET and CT, $u_1(x, y, z)$ and $u_2(x, y, z)$ are two membership functions corresponding to the PET and CT images respectively. This method obtained two different segmentation results for each pair of PET/CT images — one for PET (denoted as FVM_CO_2_PET) and another for CT (denoted as FVM_CO_2_CT). The last term penalized the segmentation difference between PET and CT, which fused the PET and CT information and achieved consistent tumor segmentation results between the two modalities. This model is an approximated variational version of Song's PET/CT co-segmentation graph model [13] (please see Eqs. (2)-(6) in [13]).

Besides, we compared the proposed method with some classical segmentation methods on PET images. They are thresholding algorithms — Otsu automatic thresholding [55], 42% and 50% of the maximum standardized uptake value (SUV_{max}, 42% Threshold and 50% Threshold); clustering algorithm — Fuzzy C-means Clustering (FCM); variational methods — MS segmentation method (MS) [56], Active Contours without edges (AC) [54] and

Geodesic Active Contours (GAC) [57]; and a graph theory algorithm — Graph Cuts (GC) [58].

4.3. Parameter Setting

All PET images were normalized linearly to [0, 255] in the proposed method. We first chose a pair of PET/CT image to adjust the parameters. The obtained parameters were used for the remaining images.

In our experiments, the parameters were set as: $\alpha = 10^3$, $\beta = 0.1$. $\lambda = 500$ (see Eqs. (1) and (11)). We added a small constant 1×10^{-6} to the denominator of the fuzzy membership prior term to avoid the zero denominator.

In FVM_CO_1 (Eq. (23)), we set η=0.7 and ξ=0.5 to assign a larger weight to PET in the background areas. In FVM_CO_2 (Eq. (24)), we set ζ=0.6. A larger weight was assigned to PET images to overcome the influences of complex background of CT images.

4.4. Evaluation Metrics

The segmentation accuracy was evaluated by computing the dice similarity index (DSI), sensitivity (SE), positive predictive value (PPV), classification error (CE) and volume error (VE).

DSI [2, 59-61], SE and PPV [31, 35, 36] measures the similarity (spatial overlap) between the segmented tumor volume V_A and the reference volume V_R:

$$DSI(V_A, V_R) = \frac{2 \mid V_A \cap V_R \mid}{\mid V_A \mid + \mid V_R \mid}, \tag{25}$$

$$SE = \frac{\mid V_A \cap V_R \mid}{\mid V_R \mid}, \tag{26}$$

$$PPV = \frac{\mid V_A \cap V_R \mid}{\mid V_A \mid}. \tag{27}$$

$|X|$ denotes the size of the set X. All these three metrics range from 0 (no spatial overlap) to 1 (perfect segmentation). As recommended by the TG211 [35] and the first MICCAI challenge on PET tumor segmentation [31] we also calculated the accuracy score by the mean of SE and PPV:

$$Score = 0.5 \times SE + 0.5 \times PPV\ . \tag{28}$$

VE and CE measure the volume difference and spatial location bias between the segmented tumor volume and the reference volume respectively [2, 59, 62]:

$$VE(V_A, V_R) = \frac{abs(|V_A| - |V_R|)}{|V_R|}, \tag{29}$$

$$CE(V_A, V_R) = \frac{|V_{FP}| + |V_{FN}|}{|V_R|}. \tag{30}$$

where $|V_{FP}|$ denotes the number of false positive errors, $|V_{FN}|$ represents the number of false negative errors. The smaller VE and CE mean the better segmentation result.

5. Experimental Results

5.1. Performance Validation of Deep Learning on CT Tumor Probability Map Extraction

In this section, we show the performance of deep learning (FCN) on CT tumor probability map extraction. In Fig. 3, we display a learning curve of FCN (see Fig. 3(i)) and four CT slices coming from four different patient data in the first row. Figure 3(a) shows a simple case where the lung tumor is isolated from the normal tissues and organs. The other three complex cases are exhibited in Figs. 3(c), (e) and (g). These three tumors were attached to the chest wall. There are no clear boundaries between the tumors and the normal tissues. The corresponding probability maps obtained by FCN are shown in the second row. Although these probability maps could not accurately describe the tumors (*e.g.* some tumor voxels corresponding to the upper-left corner in Fig. 3(c) had small probability values; and some background voxels corresponding to the upper-right corner in Fig. 3(e) had large probability values), they could surprisingly distinguish the tumor from its surrounding normal tissues with similar intensity distribution as the tumor (see Figs. 3(b, d, f, h)). This indicated that the probability maps extracted by FCN could effectively describe and tell the tumor area and background area, for both the simple case (*e.g.*, that shown in Fig. 3(a)) and the complex cases (e.g., those shown in Fig. 3(c), (e) and (g)). This was important for the proposed method to further fuse the CT and PET information for accurate tumor segmentation.

5.2. Contribution of Information Fusion to Tumor Segmentation

In Figs. 4 and 5, we show the segmentation results of the proposed method, FCN_CT and FVM_PET methods on two pairs of PET/CT images. In Fig. 4, there are necrosis tissues in the center of the tumor. Image intensities of the necrosis tissues are very low in the PET image (Fig. 4(a)). They are closer to the background intensities. Without the membership degree prior extracted from the CT image, FVM_PET method (Eq. (22)) failed to segment these necrosis tissues with low intensities (Fig. 4(e)). In the CT image (Fig. 4(b)), the tumor intensity is homogeneous. However, the tumor was attached to the chest wall. As we can see, the tumor and the chest wall have similar intensity in CT. In this case, the segmented tumor of the FCN_CT method has a small amount of false positive voxels on the chest wall (see the upper-right in Fig. 4(f)). Thanks to the use of the probability map from CT and the image intensity information from PET, the proposed method efficiently segmented the necrosis tissues, and the segmented tumor area did not include those false positive voxels (Figs. 4(c)-(d)). Figure 5 also shows the outstanding performance of the proposed method. All these

demonstrated that the proposed method could effectively fuse the complementary information of PET and CT, and successfully improved the tumor segmentation accuracy.

5.3. Comparison with FCN on CT and Fuzzy Variational Model Solely on PET

In Figs. 6 and 7, we compared the overall performance of the proposed method with FCN on CT and the fuzzy variational model on PET (Eq. (22)). Figure 6 shows the result comparison for all 36 PET/CT images according to the DSI, SE and VE values. The means and standard deviations of DSIs, SEs, PPVs, VEs and CEs are shown in Fig. 7, and the detailed numerical values are listed in Table 1. As we can see, the proposed method (DSI=0.86, SE=0.86, PPV=0.87, VE=0.16 and CE=0.30) was superior to the other two methods which only used the PET or CT image information. The proposed method had an overall improvement and a consistent segmentation performance for all tested images. The major reason for the inferior segmentation performance of the FVM_PET method (DSI=0.82, SE=0.82, PPV=0.85, VE=0.25 and CE=0.38) is that it could not handle well with the intensity inhomogeneity in PET images (one major challenge in PET single modality segmentation). The FVM_PET method was not able to correctly segment the tumor areas with low intensities. The FCN_CT method (DSI=0.76, SE=0.82, PPV=0.74, VE=0.28 and CE=0.53) had the worst segmentation performance. A possible reason is that the FCN was trained with only a small amount of training samples in our study (Note that it is normal in PET/CT imaging that a large number of training samples are seldom available). FCN_CT method still had the leaking problem due to the low contrast of intensity between the tumor and the surrounding normal tissues (one major challenge in CT single modality segmentation). Combining both information of PET and CT, the proposed method could handle well the intensity inhomogeneity of PET by using the probability map extracted from CT, and overcome the leaking problem in CT-only methods by using the relatively less complex background in PET.

5.4. Comparison with Other Multimodality Tumor Segmentation Methods

Figure 8 shows the average performance comparison among the proposed method, two multimodality tumor segmentation methods (Eqs. (23) and (24)) with traditional fusion strategies and two deep learning-based multimodality segmentation methods — W-net [30] and 3D-UNet+GC [37]. The proposed method was significantly superior to the FVM_CO_1 and FVM_CO_2 methods on both segmentation accuracy and stability. A possible reason is that the proposed method used the ‘clean’ CT information (the probability map) for the multimodality segmentation while the other two methods used the original but complex CT image. This again demonstrated the high ability of the designed method in learning to describe the tumor area and the background area using deep learning. The proposed method also had higher segmentation accuracy and better stability than the W-net and the 3D-UNet +GC methods. It shows that combining the advantages of the deep learning and the specially designed variational segmentation method could more exactly delineate the tumors from the PET/CT images.

Visual comparison of the segmentation results on three pairs of PET/CT images are shown in Fig. 9. The first and second columns show that when the background areas have similar intensities with the tumor areas in the CT image, the FVM_CO_1 (Eq. (23)) and

FVM_CO_2 (Eq. (24)) methods wrongly segmented some of these areas as tumor, although we assigned a larger weight to PET. The third column shows that the FVM_CO_1 and FVM_CO_2 methods wrongly segmented the tumor areas with low intensities in PET as background, although the tumor in CT image has homogeneous intensities. A main reason for the failure of these two traditional multimodality tumor segmentation methods is that these two methods could not adaptively combine the image information of PET and CT. The third and the last two rows in Fig. 9 show that the proposed method, W-net and 3D-UNet +GC methods could correctly segment these normal tissues in CT and the tumor areas with low intensities in PET, and the proposed method obtained more accurate segmentation results (Figs. 9(g)-(i)) than the W-net method (Figs. 9(s)-(u)) and the 3D-UNet+GC method (Figs. 9(v)-(x)).

5.5. Comparison with Several Traditional PET Segmentation Methods

In Fig. 10 and Table 1, we compared the average segmentation performance of the proposed method with several traditional single-modality PET segmentation methods. The mean and standard deviation of DSIs, SEs, PPVs, VEs and CEs of all the segmentation results of each method are listed in Table 1. A direct visual comparison is shown in Fig. 10. Combining the information from both PET and CT, the proposed method was superior to all the traditional segmentation methods conducted solely on the PET images on both accuracy and robustness. The FCM method, although not as good as the proposed method, performed the best in these traditional segmentation methods. A possible reason is that, similar with the proposed fuzzy variational model, the fuzzy characteristic of the FCM method was helpful to segment the blurred PET tumors. It seemed that both combining the information from different modalities and handling the fuzzy characteristic of tumors in PET could improve the tumor segmentation accuracy.

Illustrative results of the proposed and the eight compared methods on two lung cancer data are shown in Figs. 11 and 12, respectively. In the case of Fig. 11, there is a normal organ area with high intensity values as similar as the tumor area in the ROI (Fig. 11(a), the upper-right corner). Except for the 50% threshold method, all the other methods wrongly segmented this organ area as tumor more or less. But this normal organ area did not affect the proposed method much (Fig. 11(d)). This is mainly because FCN provided a good tumor probability map in which this normal area with high intensities had small probability values (Fig. 11(c)), and the proposed fuzzy variational model effectively used the probability information. In Fig. 12, the tumor shows intensity inhomogeneity and there are necrosis tissues. All the traditional methods missed these necrosis tissues and the tumor areas with low intensities, while the proposed method did not (Fig. 12(d)).

6. Discussion

PET/CT images have been widely used in clinical practice. The PET imaging has a high contrast but a low spatial resolution. Tumor edges are usually blurred and tumor areas may have intensity inhomogeneity in a PET image. By contrast, CT imaging has higher spatial resolution but low contrast between the tumor and surrounding normal soft tissues. Combining the advantages of the two modalities together can improve the tumor

segmentation accuracy. However, the multimodality tumor segmentation using PET/CT is challenging since the information from the PET and CT can be either complementary or contradictory. In this study, we proposed a novel method to fuse the PET/CT information for tumor segmentation.

In a CT image, the tumors present various shapes and sizes and have similar intensity with its surrounding normal tissues. It is difficult for traditional methods to delineate the tumor over the CT image. In the proposed method, the state-of-the-art deep learning method was used to cope with the complex CT images. It is generally known that, the good performance of deep learning relies on the learning of a large amount of samples with exact labels. However, in medical image processing field, especially for PET/CT images, either acquiring that amount of training data or obtaining their exact labels are generally difficult or even infeasible. Instead, we designed a deep learning network to extract the tumor probability map from the CT image and proposed a fuzzy variational model to incorporate the obtained tumor probability map and the PET intensity image for accurate multimodality tumor segmentation. This deep learning network was fed with only a small amount of CT samples for learning in our study. Although the probability map could not accurately describe the tumors, it effectively distinguished the tumor from the complex background in CT (Fig. 3). This was important for us to further fuse the CT and PET image information for more accurate tumor segmentation. The high performance of the proposed method indicated the success of the designed network at describing the tumor and the background in CT. The proposed multimodality tumor segmentation method was effective in small PET/CT datasets and had higher tumor segmentation accuracy than the existing deep learning-based multimodality segmentation methods (W-net [30] and 3D-UNet+GC [37]) (see Figs. 8 and 9).

In the proposed fuzzy variational model (Eq. (5)), the data term was designed based on the image intensity information of PET and a fuzzy membership function. The tumor probability map obtained by deep learning network from CT acted as a membership degree prior. With the prior information, the proposed variational model could handle well with the intensity inhomogeneity in a PET image (Figs. 4 and 12). The fuzzy membership function based data term coped well with the blurred tumor edges of the PET image. Because of the advantage of the high contrast, a PET image usually has a relatively less complex background, which is helpful to remove the false positive prior information introduced from the CT image due to its complex background (Figs. 4 and 5). Moreover, the proposed variational model is convex with respect to the fuzzy membership function u. We can easily and quickly get the global optimal solution of u.

The traditional multimodality tumor segmentation methods could not adaptively combine the advantages from PET and CT. They usually balanced the contradictory information between PET and CT using a compromised tumor segmentation result. These methods may still suffer from the complex background and the low contrast in CT (Fig. 9). Differently, the proposed method used the cleaner CT version, *i.e.*, the probability map, rather than the original CT image. Using this new strategy, the proposed method made full use of the complementary information between PET and CT, and avoided the interference of the

complex background of CT. The proposed method had more accurate tumor segmentation results than these traditional multimodality tumor segmentation methods.

In a PET image, some normal organ areas near the tumor may have high intensity values. These normal organ areas were often incorrectly segmented as tumor by traditional PET segmentation methods (Fig. 11). With the constraint of a prior tumor probability map from CT, the proposed method could correctly classify those areas as background.

It is well known that different PET/CT scanners have different acquisition and reconstruction properties. The PET and CT images obtained from different scanners may have different spatial resolution, noise properties and voxel size, especially for PET. However, the proposed variational model could deal with the effects of the limited spatial resolution of PET on tumor segmentation thanks to the introduction of the fuzzy membership function. On the other hand, the regularization term about the fuzzy membership function could restrain the influence of noise on tumor segmentation. Therefore, the proposed method was robust to the spatial reconstruction and noise properties of the PET images. As mentioned in Section 3.2.2, the PET images need to be up-sampled to have the same voxel dimensionality as the CT image. Image intensity errors could be introduced during the up-sampling process, and the bigger the voxel size of the original PET images, the greater the error. These errors might affect the tumor segmentation accuracy. But because of the combination of the CT image information, those errors have only minor effects on tumor segmentation accuracy. Overall, the proposed method is robust to the PET and CT images obtained from different PET/CT scanners. Very possibly, the FCN trained on the present dataset is able to perform well on PET/CT images from a different scanner, which will be further verified in our future work.

Although the proposed method worked well with a small number of training samples, increasing the training samples will contribute to extracting more precise tumor probability map from CT and improving the tumor segmentation accuracy. In this work, the deep learning network only needed to be trained on the CT images. In clinical practice, the amount of CT images is much larger than PET/CT. Possibly we can increase the training sample using the single-modality CT images with the same tumor type – like transfer learning – to further improve the segmentation performance, which we will leave to our future study.

7. Conclusions

In this study, we proposed a deep learning based variational method for tumor segmentation in PET/CT images. In this method, a 3D deep learning network was firstly constructed and trained to learn a tumor probability map from the CT image. The probability map and the PET intensity image was then integrated into a novel fuzzy variational model for more accurate tumor delineation. Jointly using the deep learning network and the variational method, the proposed method could accurately segment the tumor areas and did not need a large amount of training samples. Experimental results demonstrated that the proposed method made full use of the complementary information from PET and CT, and had a good performance for tumor segmentation, even for those with intensity inhomogeneity and

blurred tumor edges in the PET image and low contrast and complex backgrounds in the CT image.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China, under Grant Nos. 61375018 and 61672253. Wei Lu was supported in part by the NIH/NCI Grant No. R01 CA172638 and the NIH/NCI Cancer Center Support Grant P30 CA008748.

Biography

Laquan Li is a teacher in the College of Science, Chongqing University of Posts and Telecommunications, China. She received her PhD degree from the School of Automation, Huazhong University of Science and Technology, China. Her research interests include medical image processing and analysis, and variational method for inverse problem.

Xiangming Zhao is a graduate student at the School of Automation, Huazhong University of Science and Technology, China. His research interests include image processing and analysis, machine learning and deep learning. He obtained his Bachelor degree in automation in 2016.

Wei Lu is an associate attending physicist in the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, US. His research interests include PET/CT for cancer diagnosis and response evaluation, image guided radiation therapy (IGRT), 4D-CT for tumor motion compensation, and medical image analysis. He received his PhD in biological engineering from the University of Missouri in 2003.

Shan Tan is a professor with the School of Automation, Huazhong University of Science and Technology, China. His research interests include biomedical image reconstruction and analysis, pattern recognition, and inverse problem in image processing. He obtained his PhD degree in pattern recognition and intelligent system from Xidian University in 2007.

References

[1]. Ju W, Xiang D, Zhang B, Wang L, Kopriva I, Chen X, Random walk and graph cut for co-segmentation of lung tumor on PET-CT images, IEEE Transactions on Image Processing, 24 (2015) 5854–5867. [PubMed: 26462198]

[2]. Li L, Wang J, Lu W, Tan S, Simultaneous tumor segmentation, image restoration, and blur kernel estimation in PET using multiple regularizations, Computer Vision and Image Understanding, 155 (2017) 173–194. [PubMed: 28603407]

[3]. James AP, Dasarathy BV, Medical image fusion: A survey of the state of the art, Information Fusion, 19 (2013)4–19.

[4]. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, D'Souza WD, Suntharalingam M, Lu W, Spatial-temporal [^{18}F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy, International Journal of Radiation Oncology Biology Physics, 85 (2013) 1375.

[5]. Zhang H, Tan S, Chen W, Kligerman S, Kim G, D'Souza WD, Suntharalingam M, Lu W, Modeling Pathologic Response of Esophageal Cancer to Chemoradiotherapy Using Spatial-Temporal 18F-FDG PET Features, Clinical Parameters, and Demographics, International Journal of Radiation Oncology Biology Physics, 88 (2014) 195–203.

[6]. Zhu Z, Chai Y, Yin H, Li Y, Liu Z, A novel dictionary learning approach for multi-modality medical image fusion, Neurocomputing, 214 (2016) 471–482.

[7]. Du J, Li W, Lu K, Xiao B, An overview of multi-modal medical image fusion, Neurocomputing, 215 (2016) 3–20.

[8]. Bhatnagar G, Wu QMJ, Liu Z, A new contrast based multimodal medical image fusion framework, Neurocomputing, 157 (2015) 143–152.

[9]. El Naqa I, Yang D, Apte A, Khullar D, Mutic S, Zheng J, Bradley JD, Grigsby P, Deasy JO, Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning, Medical Physics, 34 (2007) 4738–4749. [PubMed: 18196801]

[10]. Markel D, Zaidi H, El Naqa I, Novel multimodality segmentation using level sets and Jensen-Renyi divergence, Medical Physics, 40 (2013) 121908. [PubMed: 24320519]

[11]. Markel D, Caldwell C, Alasti H, Soliman H, Ung Y, Lee J, Sun A, Automatic segmentation of lung carcinoma using 3D texture featmes in 18-FDG PET/CT, International Journal of Molecular Imaging, 2013(ID: 980769) (2013).

[12]. Han D, Bayouth J, Song Q, Taurani A, Sonka M, Buatti J, Wu X, Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method, Biennial International Conference on Information Processing in Medical Imaging, (Springer Berlin Heidelberg, Kloster Irsee, Germany, 2011), pp. 245–256.

[13]. Song Q, Bai J, Han D, Bhatia S, Sun W, Rockey W, Bayouth JE, Buatti JM, Wu X, Optimal co-segmentation of tumor in PET-CT images with context information, IEEE Transactions on Medical Imaging, 32 (2013) 1685–1697. [PubMed: 23693127]

[14]. Bagci U, Udupa JK, Mendhiratta N, Foster B, Xu Z, Yao J, Chen X, Mollura DJ, Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images, Medical Image Analysis, 17 (2013) 929–945. [PubMed: 23837967]

[15]. Cui H, Wang X, Zhou J, Eberl S, Yin Y, Feng D, Fulham M, Topology polymorphism graph for lung tumor segmentation in PET-CT images, Physics in Medicine & Biology, 60 (2015) 4893–4914. [PubMed: 26056866]

[16]. Cui H, Wang X, Lin W, Zhou J, Eberl S, Feng D, Fulham M, Primary lung tumor segmentation from PET-CT volumes with spatial–topological constraint, International Journal of Computer Assisted Radiology and Surgery, 11 (2016) 19–29. [PubMed: 26133651]

[17]. Krizhevsky A, Sutskever I, Hinton GE, ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012), pp. 1097–1105.

[18]. Simonyan K, Zissennan A, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv preprint arXiv: 1409. 1556, (2014).

[19]. He K, Zhang X, Ren S, Sun J, Deep Residual Learning for Image Recognition, IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016), pp. 770–778.

[20]. Ren S, He K, Girshick R, Sun J, Faster R-CNN: towards real-time object detection with region proposal networks, Advances in Neural Information Processing Systems, Montreal, Canada, 2015), pp. 91–99.

[21]. He K, Gkioxari G, Dollár P, Girshick R, Mask R-CNN, IEEE International Conference on Computer Vision, Venice, Italy, 2017), pp. 2380–7504.

[22]. Shen D, G Wu HI Suk, Deep Learning in Medical Image Analysis, Annual Review of Biomedical Engineering, 19 (2017) 221–248.

[23]. Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, Rempfler M, Armbruster M, Hofmann F, D'Anastasi M, Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields, International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 2016), pp. 415–423.

[24]. Roth HR, Farag A, Lu L, Turkbey EB, Summers RM, Deep convolutional networks for pancreas segmentation in CT imaging, SPIE Medical Imaging: Image Processing, Orlando, Florida, United States, 2015), pp. 476–484.

[25]. Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA, 3D Deeply Supervised Network for Automatic Liver Segmentation from CT Volumes, International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 2016), pp. 149–157.

[26]. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Medical image analysis, 36 (2017) 61–78. [PubMed: 27865153]

[27]. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y, A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Medical image analysis, 43 (2018) 98–111. [PubMed: 29040911]

[28]. Havaei M, Davy A, Wardefarley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H, Brain tumor segmentation with Deep Neural Networks, Medical Image Analysis, 35 (2015) 18–31.

[29]. Xiang L, Qiao Y, Nie D, An L, Wang Q, Shen D, Deep Auto-context Convolutional Neural Networks for Standard-Dose PET Image Estimation from Low-Dose PET/MRI, Neurocomputing, 267 (2017) 406–416. [PubMed: 29217875]

[30]. Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, Piraud M, Buck A, Shi K, Menze BH, Automated Whole-Body Bone Lesion Detection for Multiple Myeloma on 68Ga-Pentixafor PET/CT Imaging Using Deep Learning Methods, Contrast Media & Molecular Imaging, 2018 (2018).

[31]. Hatt M, Laurent B, Ouahabi A, Fayad H, Tan S, Li L, Lu W, Jaouen V, Tauber C, Czakon J, The first MICCAI challenge on PET tumor segmentation, Medical Image Analysis, 44 (2018) 177–195. [PubMed: 29268169]

[32]. Lecun Y, Bengio Y, Hinton G, Deep learning, Nature, 521 (2015) 436–444. [PubMed: 26017442]

[33]. Long J, Shelhamer E, Darrell T, Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition2015), pp. 3431–3440.

[34]. Badrinarayanan V, Kendall A, Cipolla R, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation, IEEE Transactions on Pattern Analysis & Machine Intelligence, 39 (2017) 2481–2495.

[35]. Hatt M, Lee JA, Schmidtlein CR, Naqa IE, Caldwell C, De BE, Lu W, Das S, Geets X, Gregoire V, Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211, Medical Physics, 44 (2017) e1.

[36]. Berthon B, Spezi E, Galavis P, Shepherd T, Apte A, Hatt M, Fayad H, Bemardi ED, Soffientini CD, Schmidtlein CR, Toward a standard for the evaluation of PET - Auto - Segmentation methods following the recommendations of AAPM task group No. 211: Requirements and implementation, Medical Physics, 44 (2017) 4098–4111. [PubMed: 28474819]

[37]. Zhong Z, Kim Y, Zhou L, Plichta K, Allen B, Buatti J, Wu X, 3D fully convolutional networks for co-segmentation of tumors on PET-CT images, IEEE 15th International Symposium on Biomedical Imaging, 2018.

[38]. Li W, Jia F, Hu Q, Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks, Journal of Computer & Communications, 03 (2017) 146–151.

[39]. Shelhamer E, Long J, Darrell T, Fully convolutional networks for semantic segmentation, IEEE Transactions on Pattern Analysis & Machine Intelligence, 39 (2017) 640–651. [PubMed: 27244717]

[40]. Ronneberger O, Fischer P, Brox T, U-Net: Convolutional Networks for Biomedical Image Segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015), pp. 234–241.

[41]. Milletari F, Navab N, Ahmadi SA, V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, Fourth International Conference on 3D Vision, Stanford, CA, USA, 2016), pp. 565–571.

[42]. Litjens G, Kooi T, Bejnordi BE, Aaa S, Ciompi F, Ghafoorian M, Jawm VDL, Van GB, Sánchez CI, A survey on deep learning in medical image analysis, Medical Image Analysis, 42 (2017) 60–88. [PubMed: 28778026]

[43]. Hu P, Wu F, Peng J, Liang P, Kong D, Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution, Physics in Medicine & Biology, 61 (2016) 8676–8698. [PubMed: 27880735]

[44]. Li F, Osher S, Qin J, Yan M, A Multiphase Image Segmentation Based on Fuzzy Membership Functions and L_1-Norm Fidelity, Journal of Scientific Computing, 69 (2016) 82–106.

[45]. Krinidis S, Chatzis V, Fuzzy Energy-Based Active Contours, IEEE Transactions on Image Processing, 18 (2009) 2747–2755. [PubMed: 19696000]

[46]. Shyu KK, Pham VT, Tran TT, Lee PL, Global and local fuzzy energy-based active contours for image segmentation, Nonlinear Dynamics, 67 (2011) 1559–1578.

[47]. Tran TT, Pham VT, Shyu KK, Image segmentation using fuzzy energy-based active contour with shape prior, Journal of Visual Communication & Image Representation, 25 (2014) 1732–1745.

[48]. Wu Y, Ma W, Gong M, Li H, Jiao L, Novel fuzzy active contour model with kernel metric for image segmentation, Applied Soft Computing, 34 (2015) 301–311.

[49]. Jeba JA, Devi DSN, Various Segmentation techniques for lung tumor therapy planning on fused PET/CT images - Survey, International Journal of Advanced Research Trends in Engineering and Technology, 4 (2017) 266–271.

[50]. Jiang J, Hu YC, Liu CJ, Halpenny D, Hellmann MD, Deasy JO, Mageras G, Veeraraghavan H, Multiple Resolution Residually Connected Feature Streams For Automatic Lung Tumor Segmentation From CT Images, IEEE Transactions on Medical Imaging, doi: 10.1109/TMI.2018.2857800 (2018).

[51]. Goldstein T, Osher S, The split Bregman method for L1-regularized problems, SIAM Journal on Imaging Sciences, 2 (2009) 323–343.

[52]. Mawlawi O, Wendt III R, Wong W-H, Principles of PET/CT, : Clinical PET, (Springer, New York, NY, 2004), pp. 44–61.

[53]. Humm JL, Rosenfeld A, Del Guerra A, From PET detectors to PET scanners, European Journal of Nuclear Medicine and Molecular Imaging, 30 (2003) 1574–1597. [PubMed: 14579100]

[54]. Chan TF, Vese LA, Active contours without edges, IEEE Transactions on Image Processing, 10 (2001) 266–277. [PubMed: 18249617]

[55]. Otsu N, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics, 9 (1979) 62–66.

[56]. Mumford D, Shah J, Optimal approximations by piecewise smooth functions and associated variational problems, Communications on Pure and Applied Mathematics, 42 (1989) 577–685.

[57]. Caselles V, Kimmel R, Sapiro G, Geodesic active contours, International Journal of Computer Vision, 22 (1997) 61–79.

[58]. Boykov Y, Funka-Lea G, Graph cuts and efficient ND image segmentation, International Journal of Computer Vision, 70 (2006) 109–131.

[59]. Dewalle-Vignion A, Betrouni N, Lopes R, Huglo D, Stute S, Vermandel M, A new method for volume segmentation of PET images, based on possibility theory, IEEE Transactions on Medical Imaging, 30 (2011) 409–423. [PubMed: 20952337]

[60]. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, Wells WM, Jolesz FA, Kikinis R, Statistical validation of image segmentation quality based on a spatial overlap index, Academic Radiology, 11 (2004) 178–189. [PubMed: 14974593]

[61]. Tan S, Li L, Choi W, Kang MK, D'Souza WD, Lu W, Adaptive region-growing with maximum curvature strategy for tumor segmentation in ^{18}F-FDG PET, Physics in Medicine & Biology, 62 (2017) 5383–5402. [PubMed: 28604372]

[62]. Hatt M, Cheze le Rest C, Turzo A, Roux C, Visvikis D, A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET, IEEE Transactions on Medical Imaging, 28 (2009) 881–893. [PubMed: 19150782]

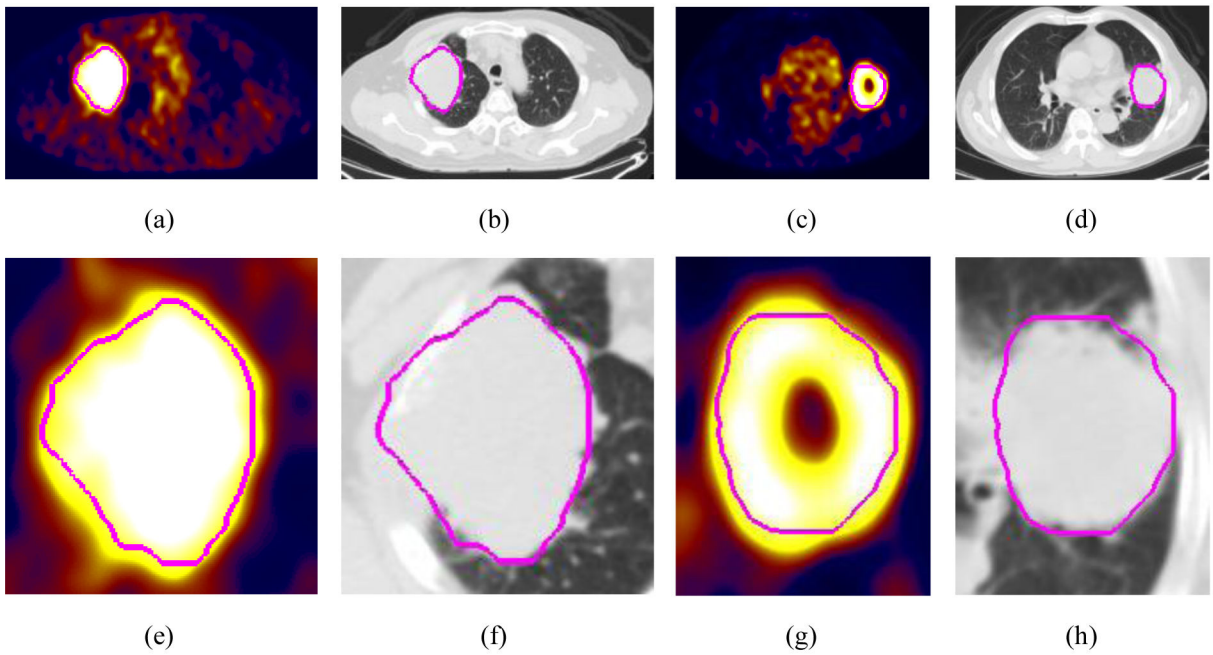


Figure 1.

PET/CT image pairs. (a) the PET image, and (b) its corresponding CT image, for one patient with lung cancer; (c) the PET image, and (d) its corresponding CT image, for another patient with lung cancer; (e)-(h) zoomed views of the regions of interest corresponding to the images in the first row. The pink contours are the tumor boundaries.

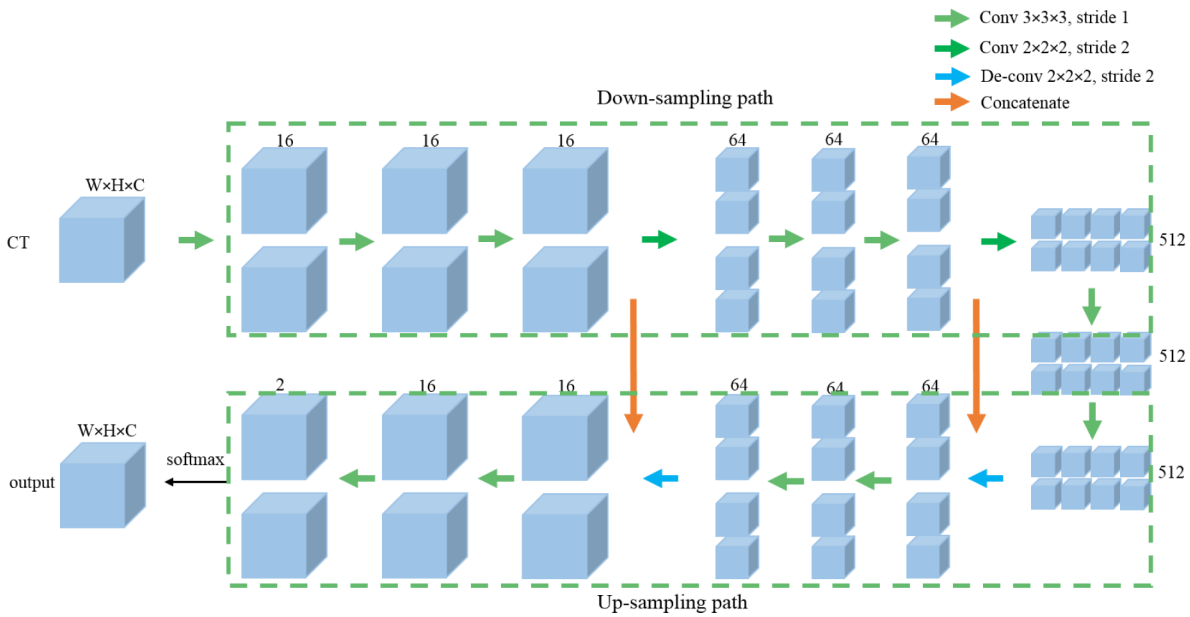


Figure 2.

The 3D fully convolutional network to extract the tumor probability map from CT images. W×H×C denotes the dimensionality of the input CT image and the output probability map.

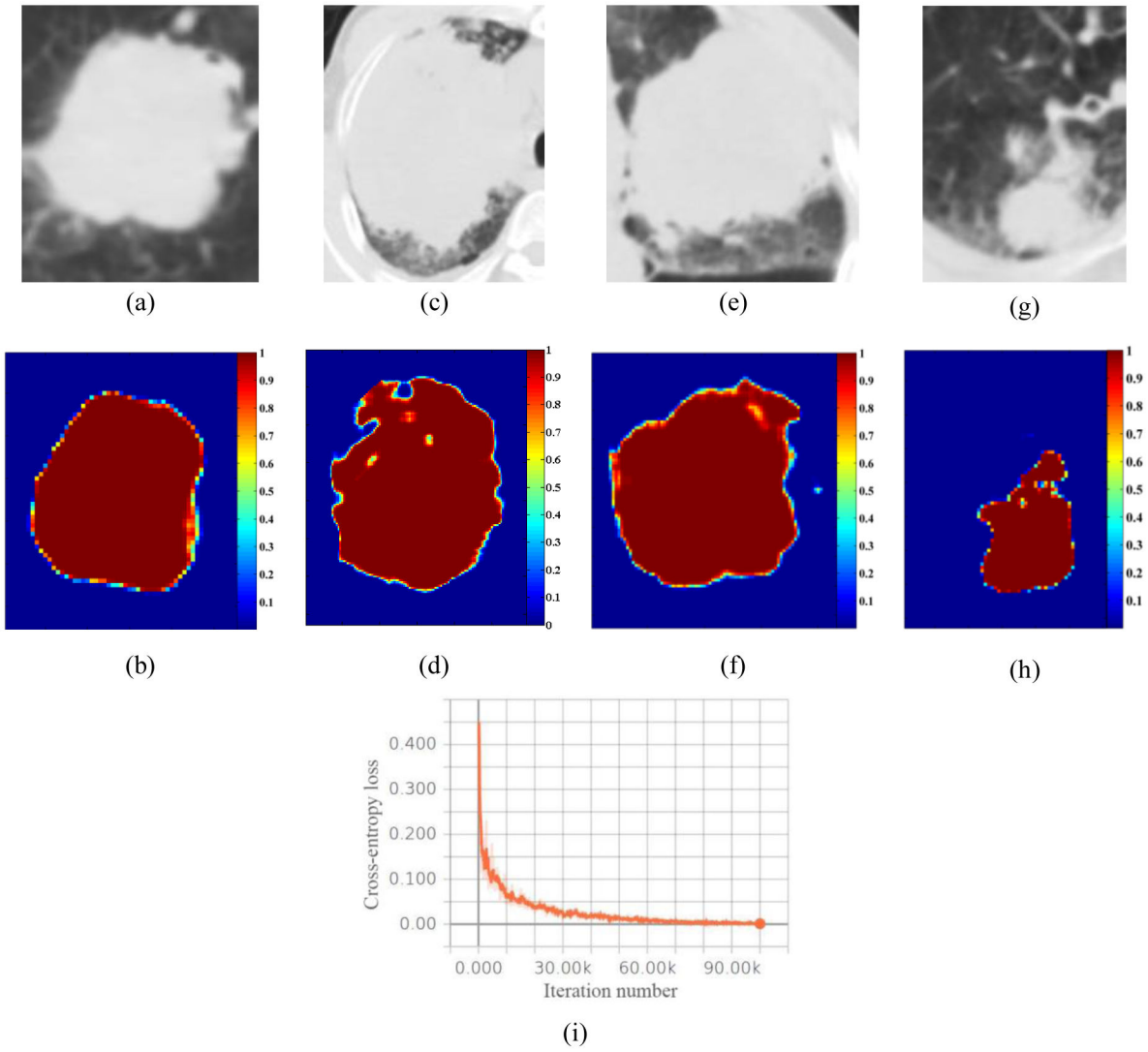


Figure 3.

Three CT images and the corresponding probability maps obtained by FCN. (a), (c), (e) and (g): CT images; (b), (d), (f) and (h): probability maps; (i) a learning curve of FCN.

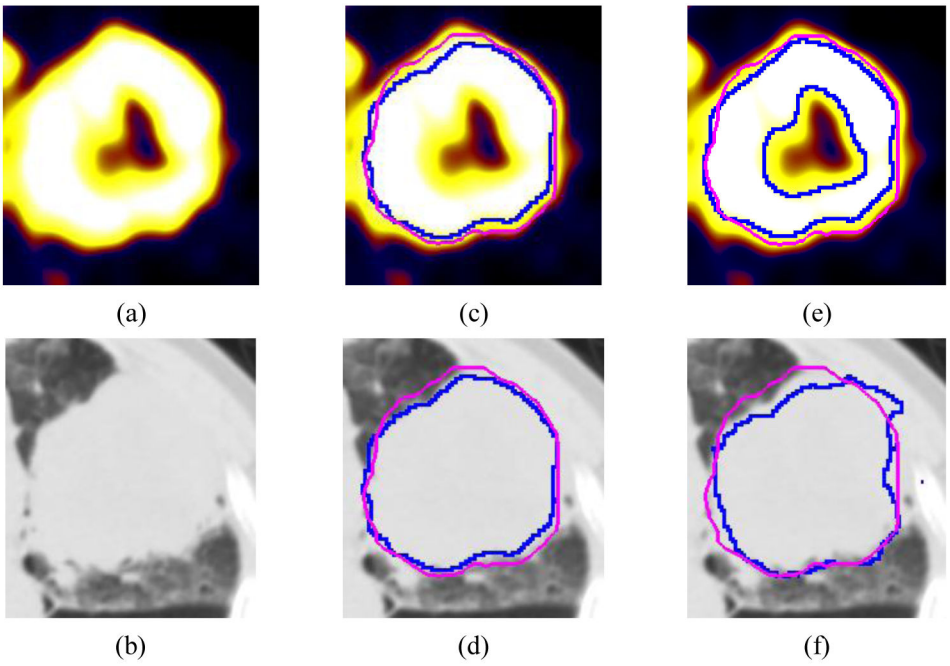


Figure 4.

Compare (axial view) the segmentation results of the proposed method with those obtained by FCN on CT and the fuzzy variational method on PET (Eq. (22)), for one patient. The pink contour is the manual contour and the blue curves are the contours of the segmented tumors. (a) the PET image, (b) the CT image, (c)-(d) the proposed method (DSI=0.91, SE=0.86, PPV=0.97), superimposed on PET (c) and CT (d) images, (e) FVM_PET (DSI=0.80, SE=0.69, PPV=0.95), (f) FCN_CT (DSI=0.83, SE=0.73, PPV=0.95).

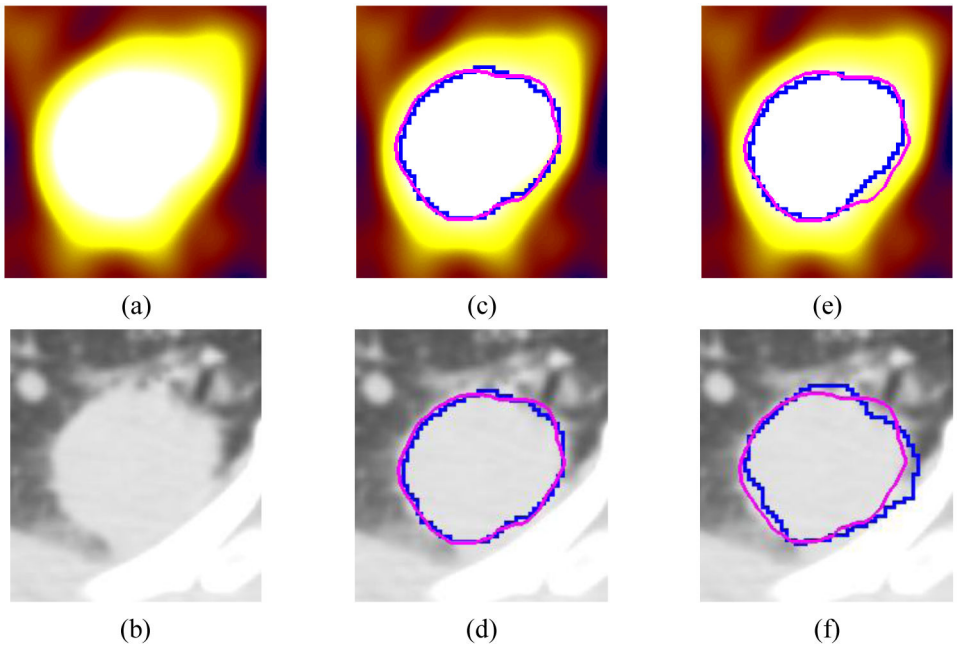


Figure 5.

Compare (axial view) the segmentation results of the proposed method with those obtained by FCN on CT and the fuzzy variational method on PET (Eq. (22)), for another patient. The pink contour is the manual contour and the blue curves are the contours of the segmented tumors. (a) the PET image, (b) the CT image, (c)-(d) results of the proposed method (DSI=0.91, SE=0.85, PPV=0.98), superimposed on PET (c) and CT (d) images, (e) FVM_PET (DSI=0.88, SE=0.79, PPV=0.99), (f) FCN_CT (DSI=0.87, SE=0.84, PPV=0.90).

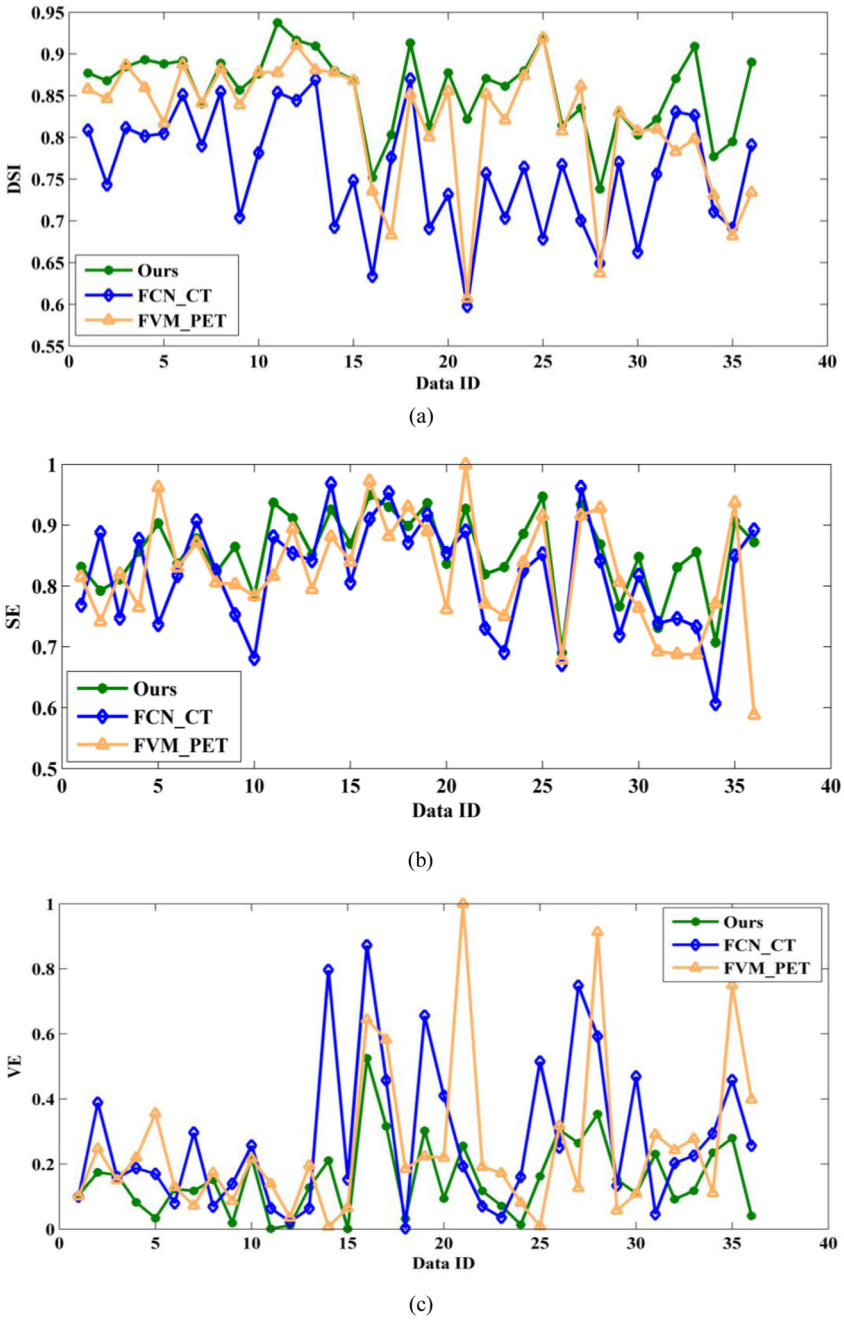


Figure 6.

Compare the quantitative results of the proposed method (green), FCN on CT (blue) and the fuzzy variational model on PET (orange) (Eq. (22)) according to DSI (a), SE (b) and VE (c) on 36 PET/CT images.

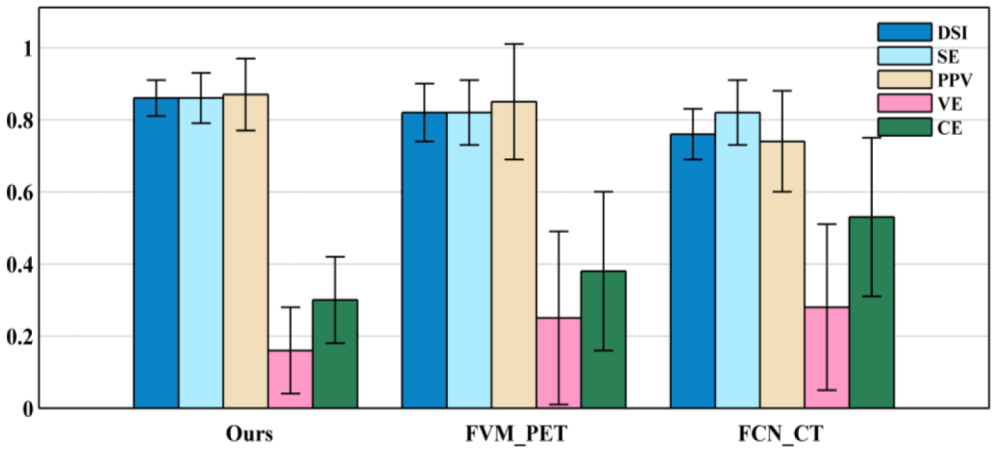


Figure 7.

Compare the average performance of the proposed method, FCN on CT and the fuzzy variational model on PET (Eq. (22)) using the mean and standard deviation of DSI, SE, PPV, VE and CE of the 36 PET/CT images.

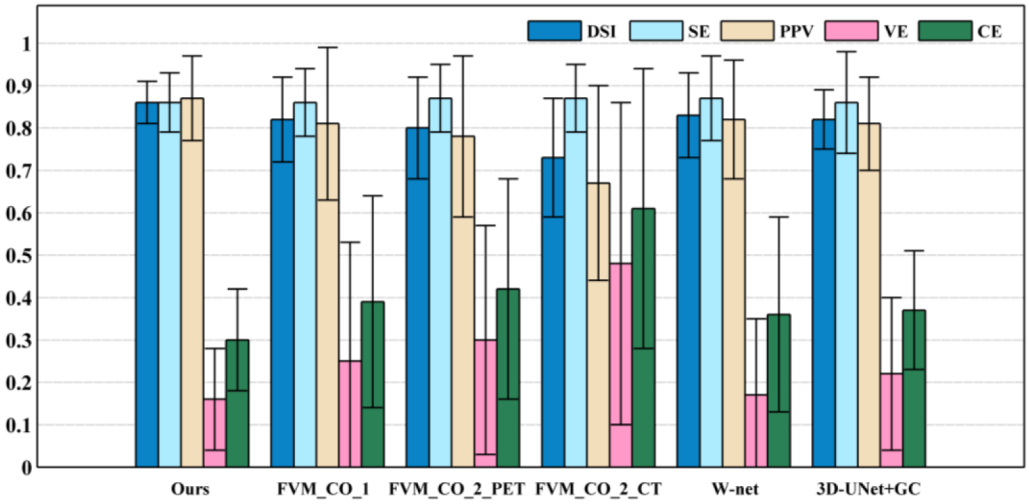
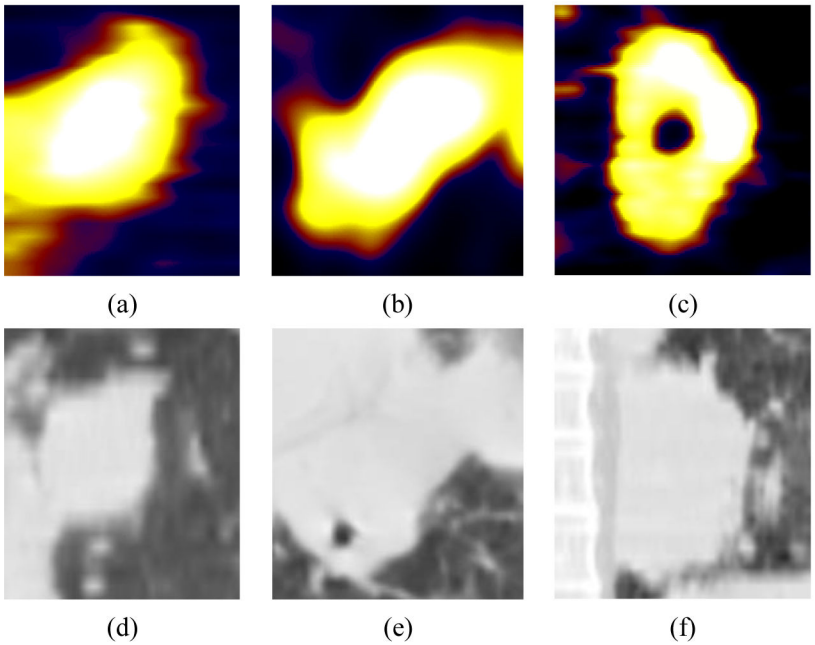


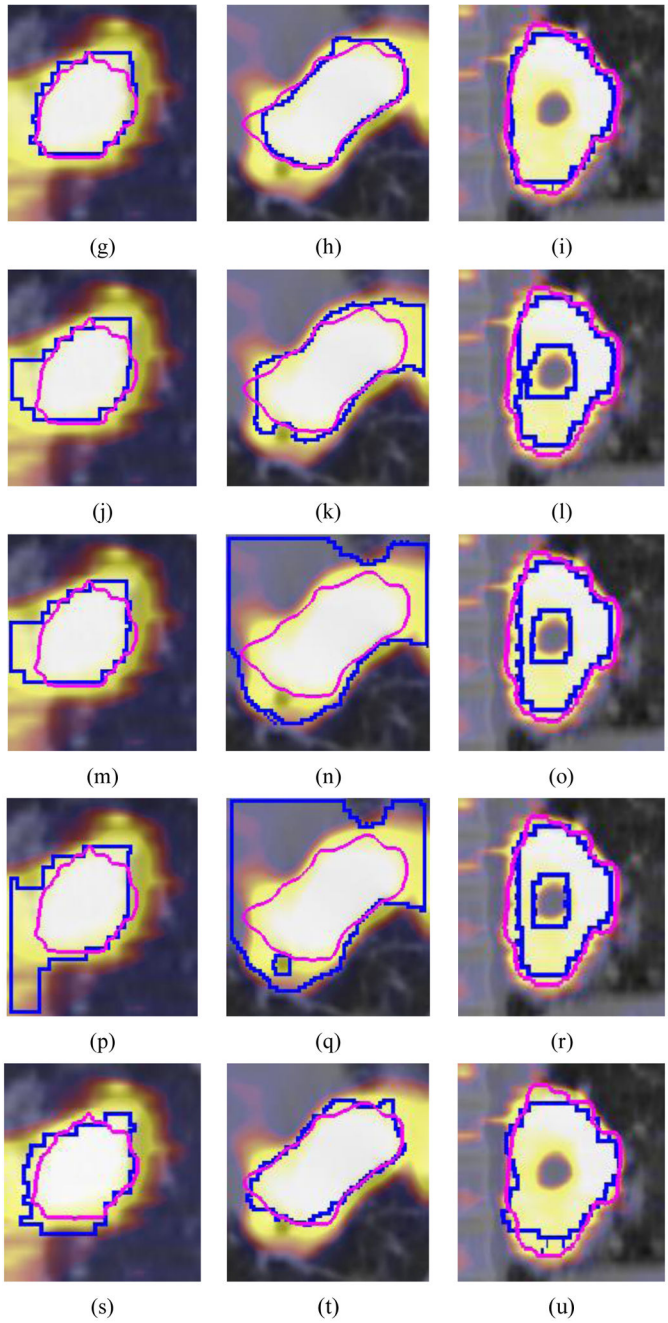
Figure 8.

Compare the average performance of the proposed method with two traditional multimodality segmentation methods (Eqs. (23) and (24)) and two deep learning-based multimodality segmentation methods — W-net [30] and 3D-UNet+GC [37], according to the mean and standard deviation of DSI, SE, PPV, CE and YE of the 36 PET/CT images.



(a) (b) (c)

(d) (e) (f)



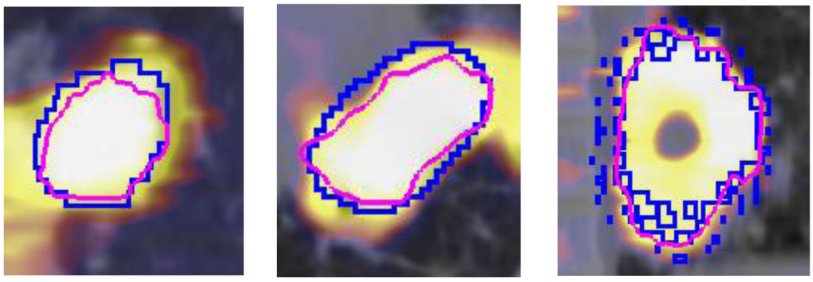
(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

(p) (q) (r)

(s) (t) (u)



(v) (w) (x)

Figure 9.

Compare the segmentation results of the proposed method with two traditional multimodality segmentation methods and two deep learning-based multimodality segmentation methods (W-net [30] and 3D-UNet+GC [37]) on three pairs of PET/CT images. The pink contours are the manual contours and the blue curves are the contours of the segmented tumors. (a)-(c) PET, (d)-(f) CT, (g)-(i) the proposed method, (j)-(l) FVM_CO_1 (Eq. (23)), (m)-(o) FVM_CO_2_PET (Eq. (24)), (p)-(r) FVM_CO_2_CT (Eq. (24)), (s)-(u) W-net, (v)-(x) 3D-UNet+GC.

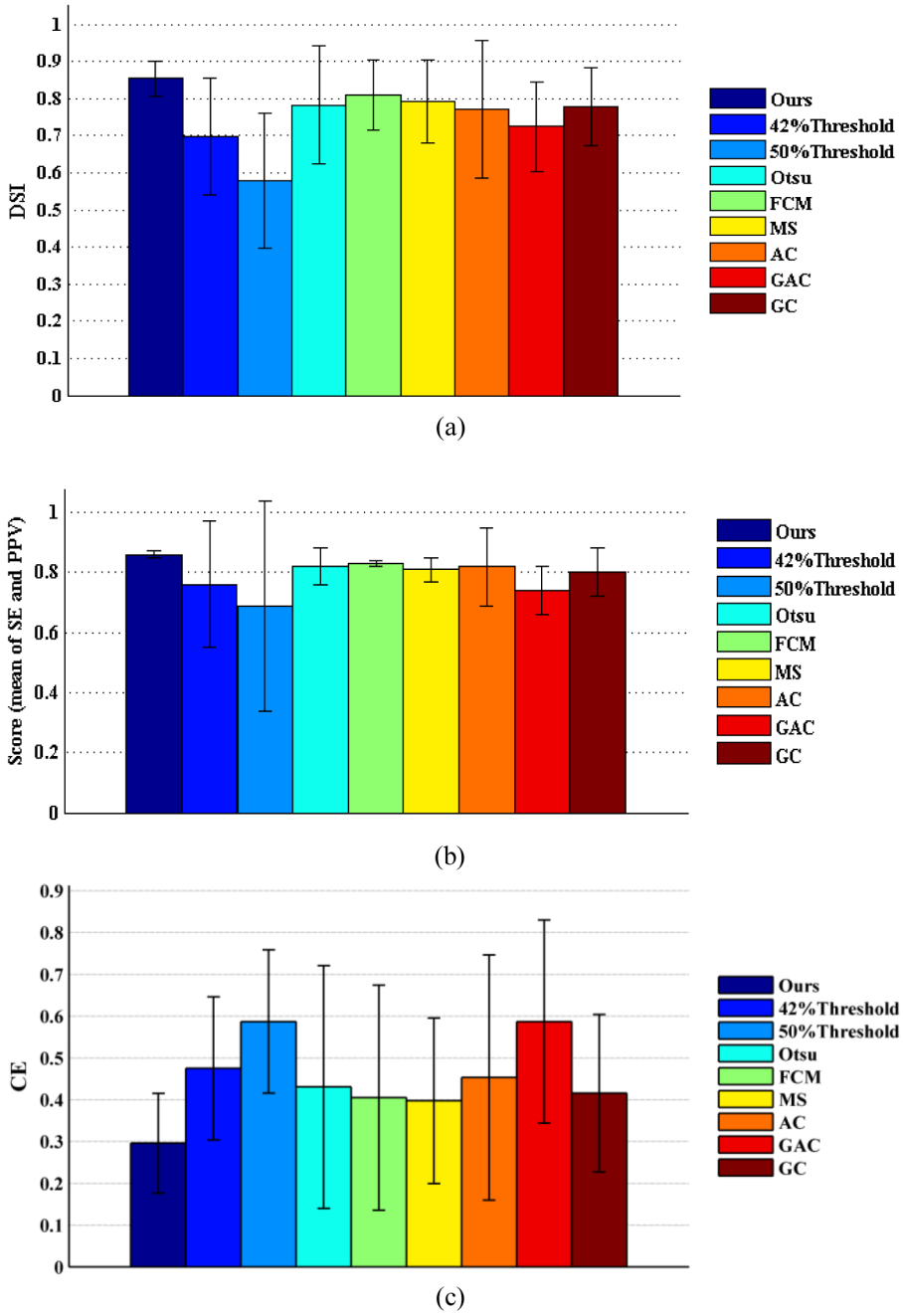


Figure 10.

Compare the average segmentation performance ot the proposed method with several traditional segmentation methods on PET. (a) DSI, (b) Score (mean of SE and PPV), (c) CE, averaged over 36 lung cancer data.

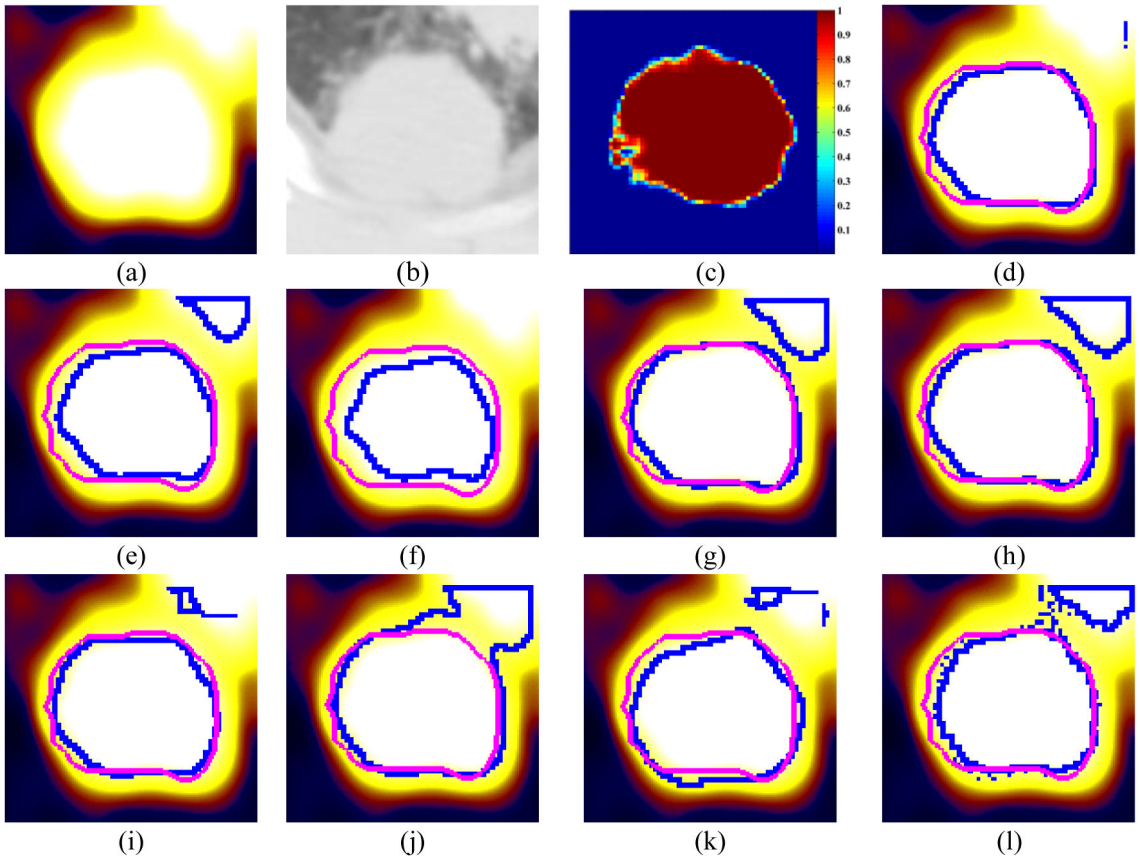


Figure 11.

Visual comparison of segmentation performance for one patient. The pink curve denotes the manual contour, the blur curves are the contours of the segmented tumor of different methods. (a) PET, (b) CT, (c) the probability map from CT, (d) the proposed method, (e) 42% Threshold, (f) 50% Threshold, (g) Otsu, (h) FCM, (i) MS, (j) AC, (k) GAC, (l) GC.

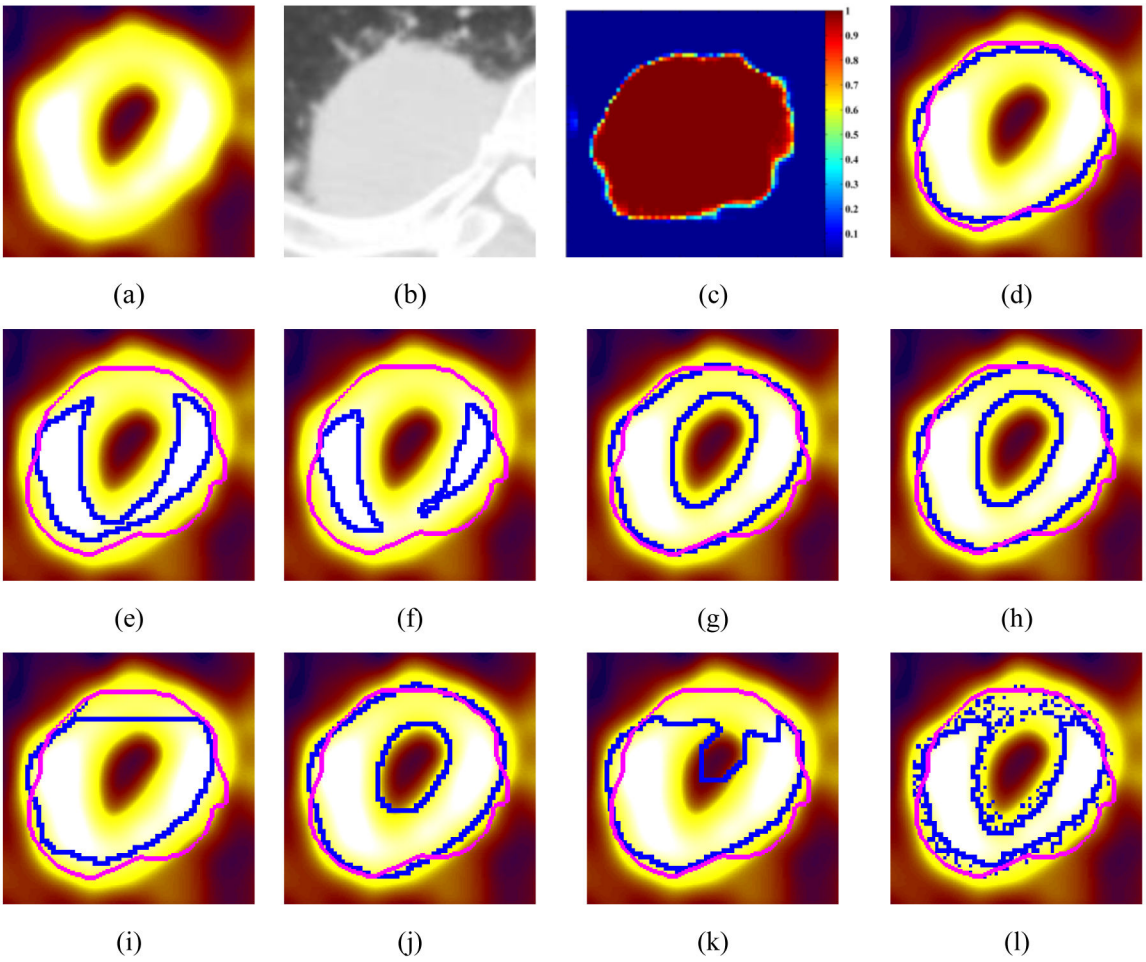


Figure 12.

Visual comparison of segmentation performance for another patient. The pink curve denotes the manual contour, the blur curves are the contours of the segmented tumor of different methods. (a) PET, (b) CT, (c) the probability map from CT, (d) the proposed method, (e) 42% Threshold, (f) 50% Threshold, (g) Otsu, (h) FCM, (i) MS, (j) AC, (k) GAC, (l) GC.

Table 1.

The mean and standard deviation of DSI, CE and VE of the segmentation results of different segmentation algorithms.

Methods	DSI (mean ± SD)	SE (mean ± SD)	PPV (mean ± SD)	Score	VE (mean ± SD)	CE (mean ± SD)
Ours	**0.86 ± 0.05**	0.86 ± 0.07	0.87 ± 0.10	**0.87**	**0.16 ± 0.12**	**0.30 ± 0.12**
FVM_PET	0.82 ± 0.08	0.82 ± 0.09	0.85 ± 0.16	0.84	0.25 ± 0.24	0.38 ± 0.22
FCN_CT	0.76 ± 0.07	0.82 ± 0.09	0.74 ± 0.14	0.78	0.28 ± 0.23	0.53 ± 0.22
FVM_CO_1	0.82 ± 0.10	0.86 ± 0.08	0.81 ± 0.18	0.84	0.25 ± 0.28	0.39 ± 0.25
FVM_CO_2_PET	0.80 ± 0.12	0.87 ± 0.08	0.78 ± 0.19	0.83	0.30 ± 0.27	0.42 ± 0.26
FVM_CO_2_CT	0.73 ± 0.14	0.87 ± 0.08	0.67 ± 0.23	0.77	0.48 ± 0.38	0.61 ± 0.33
W-net	0.83 ± 0.10	0.87 ± 0.10	0.82 ± 0.14	0.84	0.17 ± 0.18	0.36 ± 0.23
3D-UNet+GC	0.82 ± 0.07	0.86 ± 0.12	0.81 ± 0.11	0.84	0.22 ± 0.18	0.37 ± 0.14
42% Threshold	0.70 ± 0.16	0.61 ± 0.17	0.90 ± 0.18	0.76	0.38 ± 0.21	0.47 ± 0.17
50% Threshold	0.58 ± 0.18	0.45 ± 0.18	**0.94 ± 0.17**	0.69	0.54 ± 0.20	0.59 ± 0.17
Otsu	0.78 ± 0.16	0.86 ± 0.09	0.78 ± 0.24	0.82	0.32 ± 0.34	0.43 ± 0.29
FCM	0.81 ± 0.09	0.84 ± 0.09	0.82 ± 0.19	0.83	0.29 ± 0.30	0.41 ± 0.27
MS	0.79 ± 0.11	0.78 ± 0.16	0.84 ± 0.12	0.81	0.21 ± 0.19	0.40 ± 0.20
AC	0.77 ± 0.19	**0.91 ± 0.07**	0.73 ± 0.24	0.82	0.34 ± 0.35	0.45 ± 0.29
GAC	0.73 ± 0.12	0.80 ± 0.17	0.68 ± 0.11	0.74	0.27 ± 0.20	0.59 ± 0.24
GC	0.78 ± 0.10	0.75 ± 0.16	0.86 ± 0.14	0.80	0.26 ± 0.19	0.42 ± 0.19