

A Segmentation Algorithm for Quantitative Analysis of Heterogeneous Tumors of the Cervix with ^{18}F -FDG PET/CT

Wei Mu, Zhe Chen, Wei Shen, Feng Yang, Ying Liang, Ruwei Dai, Ning Wu* and Jie Tian*, *Fellow, IEEE*

Abstract—As positron-emission tomography (PET) images have low spatial resolution and much noise, accurate image segmentation is one of the most challenging issues in tumor quantification. Tumors of the uterine cervix present a particular challenge because of urine activity in the adjacent bladder. Here, we propose and validate an automatic segmentation method adapted to cervical tumors. Our proposed methodology combined the gradient field information of both the filtered PET image and the level set function (LSF) into a level set framework by constructing a new evolution equation. Furthermore, we also constructed a new hyper-image to recognize a rough tumor region (RTR) using the fuzzy c-means (FCM) algorithm according to the tissue specificity as defined by both PET (uptake) and CT (attenuation) to provide the initial zero level set (ZLS), which could make the segmentation process fully automatic. The proposed method was verified based on simulation and clinical studies. For simulation studies, seven different phantoms, representing tumors with homogenous/heterogeneous low/high uptake patterns and different volumes, were simulated with five different noise levels. Twenty-seven cervical cancer patients at different stages were enrolled for clinical evaluation of the method. Dice similarity coefficients (DSC) and Hausdorff distance (HD) were used to evaluate the accuracy of the segmentation method,

while a Bland-Altman analysis of the mean standardized uptake value (SUV_{mean}) and metabolic tumor volume (MTV) was used to evaluate the accuracy of the quantification. Using this method, the DSCs and HDs of the homogenous and heterogeneous phantoms under clinical noise level were $93.39 \pm 1.09\%$ and $6.02 \pm 1.09\text{mm}$, $93.59 \pm 1.63\%$ and $8.92 \pm 2.57\text{mm}$, respectively. The DSCs and HDs in patients measured $91.80 \pm 2.46\%$ and $7.79 \pm 2.18\text{mm}$. Through Bland-Altman analysis, the SUV_{mean} and the MTV using our method showed high correlation with the clinical gold standard. The results of both simulation and clinical studies demonstrated the accuracy, effectiveness, and robustness of the proposed method. Further assessment of the quantitative indices indicates the feasibility of this algorithm in accurate quantitative analysis of cervical tumors in clinical practice.

Index Terms—Cervical tumor segmentation; PET/CT Images; improved level set method; Fuzzy-C-Means (FCM)

I. INTRODUCTION

Cervical cancer was the third most common cancer and the fourth most common cause of cancer death in women worldwide in 2012[1]. Accurate diagnosis of cervical cancer is very important. Positron-emission tomography/computed tomography (PET/CT) has played an important role in oncology because of its high specificity and sensitivity. As a non-invasive functional imaging modality, PET can reflect the metabolic characteristics of the tumors at the molecular level, especially the quantitative analysis characteristic can provide diagnostic information and assist in planning therapy.

One of the most widely used indices for uptake quantification in PET images is the standardized uptake value (SUV), which quantifies the metabolism of tumors by measuring the ^{18}F -FDG uptake within the tumor region and normalizing to the dosage of injected ^{18}F -FDG and a scale factor representing the patient's size (such as body weight or body surface area). Another important index is the metabolic tumor volume (MTV), which has been proved to be an independent prognostic factor for tumor recurrence in patients with cervical cancer[2]. Recently, another series of semi-quantitative indices, textural parameters, focusing on describing the relationships between the intensity of the voxels and their spatial position within the region of interest (ROI)[3],

Manuscript received December 23, 2014; revised April 9, 2015 and May 8, 2015; accepted May 9, 2015. This paper is supported by the National Basic Research Program of China (973 Program) under Grant 2011CB707700, the National Natural Science Foundation of China under Grant No. 81227901, 61231004, the Chinese Academy of Sciences Fellowship for Young International Scientists under Grant 2013Y1GB0005, the National High Technology Research and Development Program of China (863 Program) under 2012AA021105, the “Guangdong Province-Chinese Academy of Sciences” comprehensive strategic cooperation program under 2010A090100032 and 2012B090400039, the NSFC-NIH Biomedical collaborative research program under 81261120414, the Beijing Natural Science Foundation under Grant No. 4132080, the Fundamental Research Funds for the Central Universities under Grant No. 2013JBZ014, the National Basic Research Program of China under Grant No. 61301002 and No. 61302025, the Chinese Academy of Sciences Visiting Professorship for Senior International Scientists under Grant 2010T2G36, and the Chinese Academy of Sciences Key deployment program under Grant No. KGZD-EW-T03.

W. Mu, Z. Chen, W. Shen, R. Dai, *J. Tian are with the Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China. (e-mail: tian@ieec.org).

F. Yang is with School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044.

Y. Liang, *N. Wu is with Cancer Institute and Hospital, Chinese Academy of Medical Sciences, 100021

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieec.org.

have been shown to reflect the heterogeneity of tumors, and can be used to predict therapy response[4] and patient outcome[5]. Unfortunately, although some of these indices are robust to segmentation, some are greatly dependent on the accurate delineation of the target volumes[6]. In addition, the definition of the tumor boundary plays a vital role in radiation therapy, which is a common treatment in oncology. On one hand, the boundary should be kept as compact as possible to minimize damage to the nearby healthy tissues. On the other hand, the boundary must include the entire extent of the tumors[7]. In clinical practice, the widely used segmentation method is manually delineation by physicians. However this method is time consuming and subjective, and the introduced inter-observer variance causes difficulty in comparing the quantitative indices acquired in different institutions. Therefore, automatic and accurate segmentation is desirable.

Currently, the segmentation methods for PET images are divided into four main categories. Firstly, due to the nature of PET images (i.e. low spatial resolution but high contrast), threshold-based segmentation techniques are proposed, and become the most widely used methods mainly because of their simplicity of implementation and use. These methods are usually divided into two main groups: fixed thresholding and adaptive thresholding. For fixed thresholding, the common used thresholds are $SUV = 2.5$ and 40% of the SUV_{max} (the maximum SUV of the ROI) in clinical use[8]. For adaptive thresholding, many techniques try to find a more optimal threshold by minimizing the intra-class variance (i.e. Otsu method[9]), or applying an iterative thresholding method based on calibrated threshold-volume curves at varying S/B ratio acquired through phantoms[10-12]. Nevertheless, finding an optimal threshold is still a challenging task for threshold-based segmentation techniques. Secondly, taking advantage of the fuzzy boundaries of uptake regions in PET images, a series of methods based on fuzzy set theory have been developed, such as Fuzzy-C-Means (FCM)[13], Fuzzy Locally Adaptive Bayesian (FLAB)[14, 15], Fuzzy Connectedness (FC), and Fuzzy Hidden Markov models (FHMM)[16, 17]. These methods are quite suitable for fuzzy boundaries, and usually perform well on the segmentation of PET images. Thirdly, region-based segmentation methods, which include region growing[18] and graph based[19] methods, also attract much attention. Most of these methods are based on the assumption that the lesions should be homogeneous[20], and may not obtain a satisfied results on heterogeneous tumors usually. Therefore many researches are focused on improving the traditional region-based method to adapt to the heterogeneous tumors in recent years, and some of them obtained stratified results[21]. The fourth kind of methods, represented by level set and active contours[22-25], are based on the boundary of the object. Though how to locate the boundaries in the noisy and low-resolution PET images is challenging, these methods could overcome the restricted condition of homogeneity and obtain more accurate results.

Compared to other tumors, cervical tumor segmentation presents more challenges because the tumor is isoattenuating to

the normal cervical stroma on CT images[26], and the urine in the adjacent bladder has a similar signal intensity to the tumor on PET images[27]. It is difficult to localize the tumor from CT images and to isolate the tumor region from the effect of the bladder on PET images with most of the present algorithms alone. Until now, there is a little literature concerning the cervical tumor segmentation. Miller *et al* used the traditional threshold-based segmentation techniques, but manual ROI placement to eliminate the bladder was needed in many cases[28]. Roman-Jimenez *et al* proposed a fusion- and Gaussian-mixture-based classification (FGMC) method to separate the bladder and tumor semi-automatically on the basis of visual interpretation[29]. Arbones *et al* applied a level set method based on the Chan-Vese (CV) model[30] to delineate heterogeneous PET-positive areas, but post-processing was also needed to exclude the bladder[31].

In this study, we propose a cervical tumor segmentation algorithm that can separate the bladder automatically and delineate the cervical tumor accurately. This method consists of two major steps: first, recognizing the rough tumor region (RTR) on a newly constructed hyper-image according to the tissue specificity to provide the initial contour of the tumor; second, delineating the accurate boundaries of the cervical tumor with an improved level set method based on the gradient fields of the PET image and the level set function (LSF). The proposed method is validated on both phantom simulation and clinical studies to demonstrate its effectiveness and robustness.

II. MATERIALS AND METHODS

A. The Chan-Vese (CV) model

The traditional level set method is suitable for the delineation of the heterogeneous and multicentric PET-positive volumes, as it can represent contours of complex topology and handle topological changes during evolution, such as merging of multiple regions. As a typical region-based active contour model, the CV model[30] is widely used in segmentation. This model extracts the objects by minimizing the Mumford-Shah [32] energy function using region properties in a level set framework, and the level set function evolution equation is:

$$\frac{\partial \phi}{\partial t} = \delta(\phi) \left\{ \alpha \cdot \text{div}(\nabla \phi / |\nabla \phi|) - \nu - \lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2 \right\}, \quad (1)$$

where ϕ is the level set function (LSF), I is the image data (i.e. the PET images in this work), α , ν , λ_1 and λ_2 are positive parameters. c_1 and c_2 are the average intensities of the regions inside and outside of the zero level set respectively:

$$\begin{cases} c_1(\phi) = \text{average}(I(\phi > 0)) \\ c_2(\phi) = \text{average}(I(\phi < 0)) \end{cases}, \quad (2)$$

which can be expressed as:

$$\begin{cases} c_1(\phi) = \frac{\int_{\Omega} I(x) \cdot H(\phi(x)) dx}{\int_{\Omega} H(\phi(x)) dx} \\ c_2(\phi) = \frac{\int_{\Omega} I(x) \cdot (1 - H(\phi(x))) dx}{\int_{\Omega} (1 - H(\phi(x))) dx} \end{cases}, \quad (3)$$

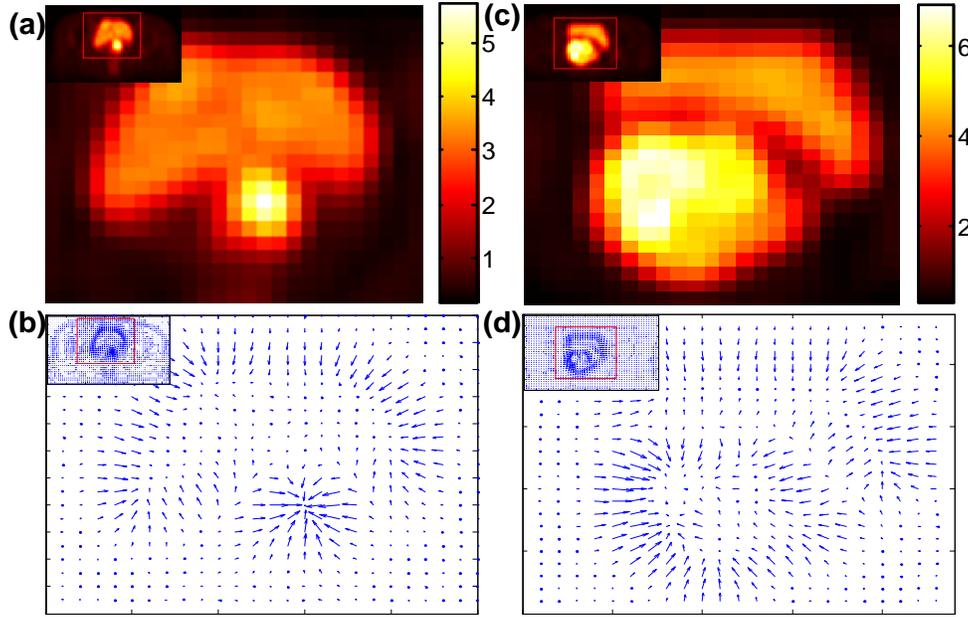


Fig.1. The top left corner of (a) and (c) are the slices of the Gaussian filtered PET images, and the top left corner of (b) and (d) are their gradient fields. The main parts of the four figures are the corresponding partial enlargements of the areas enclosed by the red squares.

where Ω is the domain of the image, and H is

$$H_\varepsilon(z) = \begin{cases} 1 & \text{if } z > \varepsilon \\ 0 & \text{if } z < \varepsilon \\ 1/2[1 + z/\varepsilon + 1/\pi \sin(\pi z/\varepsilon)] & \text{if } |z| < \varepsilon \end{cases} \quad (4)$$

The Dirac delta function δ is approximated by the following smooth function δ_ε [33]:

$$\delta_\varepsilon(x) = \begin{cases} \frac{1}{2\varepsilon} \left[1 + \cos\left(\frac{\pi x}{\varepsilon}\right) \right], & |x| \leq \varepsilon \\ 0, & |x| > \varepsilon \end{cases} \quad (5)$$

However, due to the accumulation of the urine, the cervical tumor has a similar signal intensity to the adjacent bladder on PET images. Therefore the classical CV model cannot separate the bladder and tumor. That is to say, when the evolution reaches the convergence, the segmentation results may include the bladder. Therefore, an improvement is greatly desired.

B. Modified level set method

The level set method delineates the contour by the zero level set in higher dimension, and formulates the motion of the contour as the evolution of the LSF[34]. For a given initial contour (i.e., the zero level set, ZLS), the evolution should push the zero level set to the boundary of the object. Given the fact that the centers of the bladder lumen and the tumor have a higher intensity than the periphery on the Gaussian filtered PET images (Fig.1(a) and (c)), the gradient fields of the boundary of the bladder and tumor should be opposite (Fig.1(b) and (d)). Therefore, an evolution equation adapted to cervical tumor segmentation could be constructed as:

$$\frac{\partial \phi(x)}{\partial t} = \left\{ -\lambda (I_\sigma - c_1)^2 + (I_\sigma - c_2)^2 \right\} \cdot \tanh(\cot \langle \nabla \phi(x), \nabla I_\sigma \rangle) \cdot \text{sgn}(\delta_{\varepsilon_i}(\phi)) \quad (6)$$

where I_σ represents the Gaussian filtered PET images using a

Gaussian kernel with a standard deviation σ , $\lambda > 0$, $\langle * \rangle$ stands for the angle between the two vectors (within the range of 0 to π), and $|*|$ stands for the magnitude of the vector. The definition of c_1 and c_2 are similar to CV model, and can be represented as the following equations in the 3D PET image segmentations:

$$\begin{cases} c_1(\phi) = \frac{\int_{\Omega} I_\sigma(x, y, z) \cdot H_{\varepsilon_2}(\phi(x, y, z)) dx dy dz}{\int_{\Omega} H_{\varepsilon_2}(\phi(x, y, z)) dx dy dz} \\ c_2(\phi) = \frac{\int_{\Omega} I_\sigma(x, y, z) \cdot (1 - H_{\varepsilon_2}(\phi(x, y, z))) dx dy dz}{\int_{\Omega} (1 - H_{\varepsilon_2}(\phi(x, y, z))) dx dy dz} \end{cases} \quad (7)$$

The term $\text{sgn}(\delta[\phi])$ is to ensure the expansion is performed in the narrow band around the ZLS. In order to regularize the LSF, Gaussian filter is applied after each iteration according to[35]:

$$\phi = G_\sigma * \phi, \quad (8)$$

where G_σ is a Gaussian kernel also with a standard deviation σ as the aforementioned, and $*$ stands for the convolution operation. The initial function ϕ_0 is a binary step function, and defined by:

$$\phi_0(x) = \begin{cases} c_0, & \text{if } x \in R_0 \\ -c_0, & \text{otherwise} \end{cases} \quad (9)$$

where c_0 is a positive constant (we set c_0 to 2 in this work), and R_0 is the area inside the initial contour of tumor.

The curve of the evolution function

$$L = \tanh(\cot(\theta)), \quad (10)$$

as plotted in Fig.2, has the following properties:

- 1) For θ within of 0 to $\pi/2$, which means the gradient field of I_σ and the LSF is coincident, the sign of this function is positive. The smaller the angle is, the larger the increment of L is.
- 2) For θ within the range of $\pi/2$ to π , which means the gradient fields of I_σ and the LSF is opposite, the expansion of the ZLS

will be stopped. The larger the angle is, the larger the decrement of L is.

At the beginning of the evolution, the voxels of the narrow band around the ZLS has similar intensities with the voxels inside of the initial contour, making the magnitude adjustment term $\{-\lambda(I - c_1)^2 + (I - c_2)^2\}$ positive, which is similar to the CV model. At the same time, the gradient fields of the LSF and the I_σ around the ZLS are coincident, and L is also positive. Both of the two positive terms increase the LSF, and the ZLS could expand. The smaller the angle is, the larger the increment of the LSF is, and the larger the probability of the expansion of the ZLS is. When the ZLS reaches the boundary of the tumor and the bladder, the magnitude adjustment term is positive, but L is negative for the gradient fields of LSF and I_σ around the ZLS are opposite. Therefore the expansion of the ZLS will be stopped. When the ZLS reaches the boundary of the tumor and the background, L is near to zero, and the magnitude adjustment term is negative. Both of them can ensure the termination of the expansion more effectively at the boundary of the tumors.

Finally, the largest connected region constituted by the voxels whose LSFs are positive is the ROI. By applying median filtering, we can delineate the cervical tumor.

The above evolution equation can be implemented with a finite difference scheme. Considering the 3-D case, the time dependent LSF $\phi(x, y, z, t)$ can be given in discretized form $\phi_{i,j,k}^n$ with spatial index (i, j, k) and temporal index n . Then (6) can be discretized as the following finite difference equation:

$$\frac{\phi_{i,j,k}^{n+1} - \phi_{i,j,k}^n}{\Delta t} = L(\phi_{i,j,k}^n), \quad (11)$$

which can be expressed as:

$$\phi_{i,j,k}^{n+1} = \phi_{i,j,k}^n + \Delta t \cdot L(\phi_{i,j,k}^n), \quad n = 0, 1, 2, \dots \quad (12)$$

The final LSF can be obtained by the above iterative process, and the voxels within the region where $\phi > 0$ are extracted. The step length Δt was set to equal to σ in our work.

C. Automatic definition of the initial tumor contour

In order to define the initial tumor contour (i.e. ZLS), the rough tumor region (RTR) is needed to be recognized firstly. Taking advantage of the fuzzy boundaries of uptake regions in PET images, the FCM algorithm could be used. The FCM method was proposed by Dunn[36] and later improved by Bezdek[37]. In general, the FCM method divides the image into c clusters by minimizing the following objective function:

$$J_{FCM} = \sum_{i=1}^N \sum_{j=1}^C (u_{ij})^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty, \quad (13)$$

where x_i is the i th measured data of dimension d , c_j is the d -dimension center of the j th cluster, u_{ij} is the degree to which data element x_i belongs to cluster c_j , m is the factor to adjust the membership degree weighting effect ($m=2$ in this work), and $\|\cdot\|$ stands for the Euclidian norm. The fuzzy clustering result is calculated by iteratively optimizing this objective function with the update of u_{ij} and c_j by:

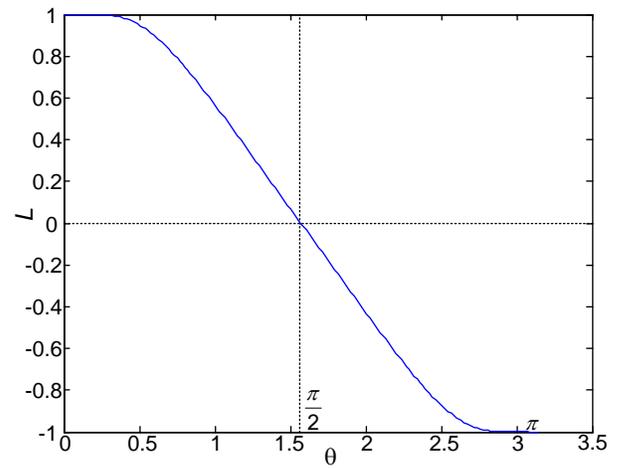


Fig.2. The curve of the function L

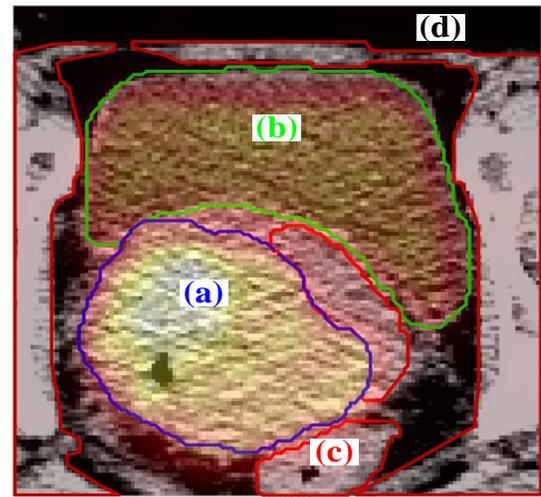


Fig.3. The diagram of the four different kinds of regions in the constructed hyper-image

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad c_j = \frac{\sum_{i=1}^N u_{ij}^m x_j}{\sum_{i=1}^N u_{ij}^m}, \quad 1 \leq i \leq N, 1 \leq j \leq C. \quad (14)$$

The iteration will not stop until $\max_{ij} \{|u_{ij}^{k+1} - u_{ij}^k|\} < \varepsilon$, where ε is the termination criterion within the range of 0 to 1 (10^{-5} was used in our work), and k is the iteration steps.

According to the tissue specificity, which refers to the fact that bladder content and the cervical tumor have similar degrees of FDG uptake compared with other surrounding tissues, but different structural information due to different levels of attenuation on CT, we could construct a hyper-image utilizing both the metabolic and anatomic information. For each voxel of this hyper-image, three features are considered: the SUV normalized to the SUVmax on PET images, the Hounsfield unit (HU) density values normalized to the maximum HU values on CT images, and the product of the above two features. Then the constructed hyper-image can be divided into four regions (shown as Fig.3) based on the tissue

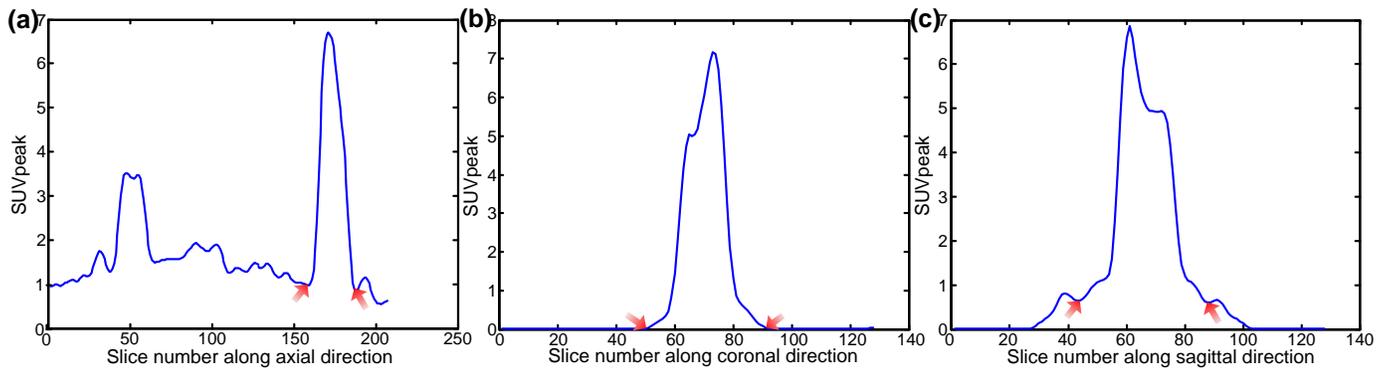


Fig.4. Per-slice SUVpeak of a PET scan, where the horizontal axis indicates the slice number along the axial, coronal, and sagittal directions, respectively, the vertical axis represents the SUVpeak of the corresponding slice. The slices between the two red arrows include the cervical tumor.

specificity: (a) a region with high SUV and moderate HU that corresponds to the tumor (regions delineated in blue); (b) a region with high SUV and low HU that corresponds to the bladder contents (regions delineated in green); (c) a region with low SUV and moderate HU that corresponds to the other soft tissues (regions delineated in red); and (d) a region with low SUV and low HU that corresponds to the background.

The detailed description of the proposed methodology to define the initial tumor contour is given as follows.

- 1) Localization of the cervical tumor: The whole PET images were processed by Gaussian filtering. Then SUVpeak (a local average SUV value in a group of voxels surrounding the voxel with $SUV_{max}[38]$) of each slice along the axial direction was calculated, and the corresponding curve was smoothed and plotted as Fig. 4a. The slices corresponding to the two adjacent valleys in front of and after the last significant peaks (pointed by the red arrows in Fig.4a) were the slices including cervical tumor. Similarly, the SUVpeak of each slice along the coronal and sagittal directions could also be obtained, and the slices between two adjacent valleys in front of and after the main peaks (pointed by the red arrows in Fig.4b and c) were used to define a smaller region (as Fig.3).
- 2) Constructing the hyper-image: The construction needed to up sample the PET images to CT resolution firstly so that each voxel in PET images corresponded to only one voxel in CT images. Then the hyper-image was constructed using the normalized SUV, HU and their product.
- 3) Separation of the RTR: The constructed hyper-image was partitioned into 4 clusters using the FCM method. For each of the obtained clusters, a simple morphological operation (erosion and dilation with a proper kernel) was performed, and the largest connected component was extracted.
- 4) Distinction of the RTR: In order to find the RTR from the four clusters, the mean attribute values of each cluster was calculated. The cluster with the largest sum of the SUV and HU was the RTR in our work. For some exceptional cases, we needed to consider the location of each cluster as the cervix is close to the back of the body.
- 5) Determination of the final RTR: Given that the obtained RTR may include the bladder wall, as it has similar intensity with the tumor both in CT images and PET

images due to their similar attenuation and the partial volume effect, an erosion operation was performed. Then we acquired the final RTR by down sampling to the PET resolution.

- 6) Definition of the initial tumor contour: The voxels with SUV larger than 40% of the SUV_{max} (threshold commonly used in clinical practice) [8] in the RTR were regarded as R_0 in the initial level set function ϕ_0 to ensure the inclusion of the all the center of the tumor.

D. Datasets and validation

1) Phantom simulation study

Since it is difficult to obtain the ground truth in PET imaging, phantom simulations were used to validate the efficiency of segmentation methods. We used a part of an anthropomorphic abdomen phantom containing a urinary bladder and a cervix, which was derived from a whole-body CT scan of a human patient. In order to verify the efficiency and robustness of the proposed method, we simulated both homogeneous and heterogeneous tumors, whose ground truth segmentations were available from CT images.

a. Homogeneous tumor phantoms

The SUV of the background compartment (the normal part of the cervix) was set to 0.5. Two spheres with 4-voxel radii (with volumes of 268 voxels) representing tumors were both set with SUVs equal to 4, making TBRs (tumor-to-background ratio) equal to 8, and BTRs (bladder lumen-to-tumor ratio) equal to 1 and 2. Another two tumors, represented by spheres with 6-voxel radii (with volumes of 905 voxels) were set with SUVs equal to 8, making TBRs equal to 16, and BTRs equal to 1 and 2. These four phantoms were referred to phantom a, b, c, and d, respectively.

b. Heterogeneous tumor phantoms

To isolate the effects of tumor shape and homogeneity, we created three irregularly shaped tumor phantoms of different volumes (349, 934, and 2105 voxels) with an averaged intensity histogram generated from 27 clinical cervical tumors, which were represented by phantom e, f, and g, respectively. Each voxel of the simulated tumors was chosen randomly to have an SUV drawn from this histogram[39]. The shape and volume of the tumors were based on clinical PET images. The simulation of the bladder was similar to that of the tumor. The

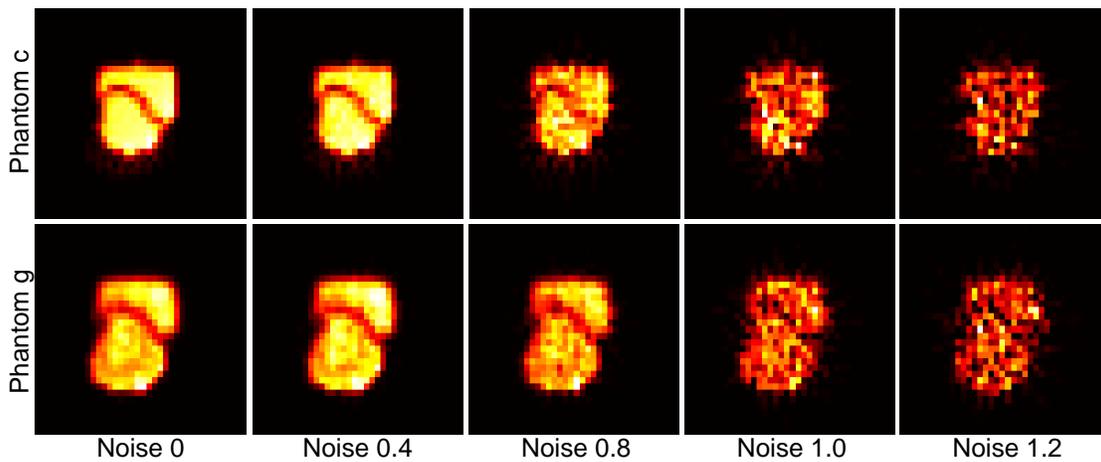


Fig.5. Simulation results of the homogenous and heterogeneous phantoms on different noise levels before Gaussian filtering.

SUV of the cervix was set to 10% of the SUVmin of the tumor.

The simulation, which was done with a PET camera (ECAT EXACT HR+®, CTI Siemens, Knoxville TN, USA), was reconstructed using the ordered subsets expectation maximization (OSEM) algorithm with five iterations, and the anisotropic resolution was $5.47 \times 5.47 \times 3.27 \text{ mm}^3$. Considering the noise in the raw PET data (i.e. the sinogram data), we used a common noise model to simulate the noise in the simulation study as :

$$S^* = S + \sigma \times N(0,1), \quad (15)$$

where S is the simulated noise-free raw PET data, S^* is the simulated noise raw PET data, and $N(0,1)$ represents the standard normal distribution. Assuming the measurement noise is subject to a Poisson distribution, σ can be modeled by:

$$\sigma = \alpha \sqrt{\frac{S}{\Delta T}}, \quad (16)$$

where α is a proportionality constant that determines the noise level and ΔT is the scan time. In our study, α was set to 0, 0.4, 0.8, 1.0, and 1.2, and twenty realizations for each noise level were simulated. Fig.5 illustrated the simulation results of the homogenous and heterogeneous phantoms on different noise levels before Gaussian filtering.

2) Clinical PET/CT Studies

This retrospective study was approved by the institutional review board of Cancer Institute and Hospital, Chinese Academy of Medical Sciences, and the absence of informed consent was also approved for the anonymous analysis of the data. Clinical PET/CT datasets were from 27 patients with squamous cell carcinoma of the uterine cervix at different stages. Patients had been injected with 200–400 MBq of ^{18}F -FDG, depending on body weight. The PET/CT images (GE Discovery ST16 PET-CT®, GE Healthcare, Milwaukee WI, USA) were acquired 50–90 minutes after the intravenous administration of the radioactive tracer. Each PET acquisition measured $128 \times 128 \times 207$ voxels with anisotropic resolution of $5.47 \times 5.47 \times 3.27 \text{ mm}^3$, while each CT acquisition measured $512 \times 512 \times 207$ voxels with anisotropic resolution of $0.98 \times 0.98 \times 3.27 \text{ mm}^3$.

The PET images were converted into SUV by normalizing the activity of the voxel to the dosage of injected ^{18}F -FDG and

the patient body weight. The ground truth segmentations were constructed by averaging the segmentations results delineated by two experienced physicians.

3) Evaluation Criteria

To evaluate the accuracy of the segmentation, two common metrics, dice similarity coefficient (DSC) [40], and Hausdorff distance (HD) [41] were used to perform direct comparison between the segmented volume and the ground truth.

Assuming the auto-segmented tumor and the ground truth are denoted by U_1 and U_2 , respectively. DSC is expressed as follows:

$$DSC(U_1, U_2) = 2 \frac{|U_1 \cap U_2|}{|U_1| + |U_2|} = \frac{2 \cdot TPFV}{(FPVF + TPVF) + (FNVF + TPVF)} \quad (17)$$

where TPVF, FPVF, and FNVF stands for the true positive, false positive, and false negative volume fractions, respectively. This metric measures the spatial overlap between the segmented object and the ground truth. The higher the DSC is, the more accurate the segmentation method is.

HD is a metric that measures the distance of the most mismatched boundary points between the segmented tumor and the ground truth, and is defined as:

$$HD(X, Y) = H(\partial U_1, \partial U_2) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\} \quad (18)$$

where ∂ denotes the boundary of segmented volumes such that $X = \partial U_1$ and $Y = \partial U_2$, d is the Euclidean distance between x and y , sup and inf denote supremum and infimum. As a shape dissimilarity metric, the lower HD means the segmented volumes has a more similar shape to the ground truth, which indicates that the segmentation method is more accurate.

To evaluate the accuracy of the quantification of the cervical tumors based on the proposed method, the clinically significant quantitative measures, the SUVmean (the mean SUV of the delineated volume) and the metabolic tumor volume (MTV) obtained from the auto-segmented tumors were compared with the reference values obtained from the ground truth



Fig.6. The mean DSCs and HDs and the relative standard deviations obtained over different phantoms using the LS-GF method (a–g) at different noise levels. (a) – (d) are homogeneous phantoms with different radii, TBRs and BTRs: (a) 4 voxels, 8:1,1:1; (b) 4 voxels, 8:1, 2:1; (c) 6 voxels, 16:1,1:1; (d) 6 voxels, 16:1, 2:1. (e) – (g) are heterogeneous phantoms with volumes of 349, 934, and 2105 voxels, respectively.

TABLE I
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON HOMOGENEOUS PHANTOMS UNDER DIFFERENT NOISE LEVEL

	DSC(%)				HD(mm)			
	LS-GF	T40%	Otsu	RW	LS-GF	T40%	Otsu	RW
Noise 0	93.23 ± 1.32	92.10 ± 0.21	89.45 ± 2.69	88.36 ± 1.41	6.02 ± 1.09	7.71 ± 0	7.71 ± 0	7.71 ± 0
<i>P</i> -value		<10 ⁻⁶ *	<10 ⁻⁶ *	<10 ⁻⁶ *		<10 ⁻⁶ *	3.37×10 ⁻⁶ *	9.08×10 ⁻⁶ *
Noise 0.4	93.39 ± 1.09	91.02 ± 0.61	90.56 ± 1.94	90.38 ± 1.41	6.02 ± 1.09	7.71 ± 0	7.71 ± 0	7.71 ± 0
<i>P</i> -value		<10 ⁻⁶ *	<10 ⁻⁶ *	<10 ⁻⁶ *		<10 ⁻⁶ *	3.92×10 ⁻⁵ *	7.73×10 ⁻⁵ *
Noise 0.8	91.71 ± 2.34	90.40 ± 1.57	88.71 ± 1.91	90.24 ± 1.75	7.27 ± 1.42	11.05 ± 8.81	14.93 ± 9.30	11.05 ± 8.81
<i>P</i> -value		1.50×10 ⁻² *	<10 ⁻⁶ *	1.74×10 ⁻⁵ *		1.50×10 ⁻² *	5.56×10 ⁻⁴ *	1.10×10 ⁻³ *
Noise 1.0	90.40 ± 2.49	88.13 ± 1.91	86.42 ± 2.82	88.13 ± 1.91	7.88 ± 0.60	16.19 ± 8.75	22.26 ± 11.21	16.14 ± 8.75
<i>P</i> -value		6.57×10 ⁻⁶ *	<10 ⁻⁶ *	6.65×10 ⁻⁴ *		<10 ⁻⁶ *	<10 ⁻⁶ *	7.54×10 ⁻⁶ *
Noise 1.2	88.67 ± 2.83	85.95 ± 4.00	86.48 ± 4.37	85.95 ± 4.00	9.08 ± 0.27	20.40 ± 10.94	23.47 ± 11.87	20.40 ± 10.94
<i>P</i> -value		<10 ⁻⁶ *	3.96×10 ⁻³ *	<10 ⁻⁶ *		<10 ⁻⁶ *	1.07×10 ⁻⁶ *	3.96×10 ⁻⁶ *

* The difference between the T40% (or Otsu, RW) and the LS-GF is significant at the 0.05 level.

TABLE II
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON HETEROGENEOUS PHANTOMS UNDER DIFFERENT NOISE LEVEL

	DSC(%)				HD(mm)			
	LS-GF	T40%	Otsu	RW	LS-GF	T40%	Otsu	RW
Noise 0	93.07 ± 2.00	90.91 ± 2.32	90.77 ± 2.10	90.60 ± 2.60	8.64 ± 2.84	12.74 ± 3.12	10.12 ± 1.91	9.35 ± 1.86
<i>P</i> -value		2.38×10 ⁻⁵ *	<10 ⁻⁶ *	2.53×10 ⁻⁶ *		<10 ⁻⁶ *	<10 ⁻⁶ *	4.70×10 ⁻³ *
Noise 0.4	93.59 ± 1.63	90.84 ± 2.53	91.55 ± 1.54	91.36 ± 1.74	8.92 ± 2.57	13.18 ± 2.84	10.61 ± 2.73	9.35 ± 1.86
<i>P</i> -value		<10 ⁻⁶ *	<10 ⁻⁶ *	<10 ⁻⁶ *		<10 ⁻⁶ *	1.26×10 ⁻⁶ *	0.039 *
Noise 0.8	92.02 ± 2.88	87.87 ± 2.36	89.86 ± 3.35	89.68 ± 2.29	9.08 ± 2.14	12.47 ± 4.92	10.77 ± 4.32	10.99 ± 4.70
<i>P</i> -value		<10 ⁻⁶ *	<10 ⁻⁶ *	<10 ⁻⁶ *		1.78×10 ⁻⁶ *	<10 ⁻⁶ *	10 ⁻⁶ *
Noise 1.0	90.93 ± 3.35	86.58 ± 2.74	88.61 ± 3.48	87.94 ± 3.00	10.75 ± 2.95	14.39 ± 2.08	14.44 ± 2.46	13.73 ± 2.79
<i>P</i> -value		<10 ⁻⁶ *	7.76×10 ⁻⁵ *	<10 ⁻⁶ *		4.99×10 ⁻⁶ *	3.63×10 ⁻⁶ *	2.60×10 ⁻³ *
Noise 1.2	89.27 ± 3.66	83.95 ± 4.35	87.01 ± 6.63	87.53 ± 3.54	10.94 ± 2.73	15.04 ± 3.83	14.60 ± 4.38	16.14 ± 1.26
<i>P</i> -value		<10 ⁻⁶ *	2.65×10 ⁻⁴ *	<10 ⁻⁶ *		<10 ⁻⁶ *	1.20×10 ⁻³ *	<10 ⁻⁶ *

* The difference between the T40% (or Otsu, RW) and the LS-GF is significant at the 0.05 level.

segmentation using Bland-Altman analysis, which calculates the mean and 95% confidence interval (CI) of the difference.

III. EXPERIMENTS AND RESULTS

The proposed method was implemented in Matlab R2012a on a desktop with an Intel(R) Core(TM) i5-650 @ 3.20GHz CPU using 4GB of RAM, and the reconstruction of the simulated PET images was implemented in Matlab R2012a on an Intel(R) Xeon(R) CPU E5-2420 v2 @ 2.20GHz CPU using 32GB of RAM.

We compared the proposed method with several traditional methods, such as the classical level set method based on Chan-Vese models, the traditional fixed thresholding method using 40% of the SUVmax (T40%), the Otsu method[9], and random walk (RW)[19]. Given the segmentation results obtained by the Chan-Vese method may include the bladder, which leads to the inaccurate quantitative analysis, we compared it with our method (hereafter referred to as the “LS-GF method”) in qualitative. Given the T40%, Otsu, and RW methods did not possess the ability to separate tumor from the bladder lumen, we performed them in the cervix region for the quantitative comparison.

A. Phantom simulation study

Since the simulated tumor was placed in the area of the eroded cervix, we regarded the eroded cervix as the RTR, and the phantom simulation study was mainly used to evaluate the proposed LS-GF method in terms of delineating the cervical tumor. Similarly, T40%, Otsu, and RW methods were also performed in the cervix of the simulated PET images.

Fig.6 displays the mean DSCs, HDs and the corresponding standard deviations of the 20 realizations for each phantom at different noise levels using the proposed method. As the figure illustrates, the proposed LS-GF method performs better with DSC rates higher than 90% for the larger simulated tumors

(phantoms c, d, f, and g), while the HDs are less than 11mm (about 2voxel) at all noise levels. For the smaller simulated tumors (phantoms a, b, and e), the proposed method can obtain nice results with DSC larger than 90% and HD smaller than 11mm at the low noise levels.

In order to present the improvement of the proposed LS-GF method compared to traditional level set method, we performed the two methods on the same case with the same ZLS under a series of different iteration numbers, which was shown in Fig.8 (a). From this figure, we can see that the evolution of the proposed method could be stopped at the boundary of the tumor adjacent to the bladder, while the CV model could not. The segmentation results obtained with the proposed method and other three different methods based on simulated phantom g are given in Fig.9(a), from which we can see the edge of the segmentation result obtained by the proposed method is more closely to the gold standard.

The quantitative comparison of different methods based on homogeneous phantoms and heterogeneous phantoms are listed in Tables I and II. It is apparent that the proposed method performs well for both homogeneous and heterogeneous phantoms with DSCs high than 92% and DSCs less than 11mm, which is much better than the other three methods under all of the different noise levels. At the same time, Students’ *t*-test was also performed to validate the significance of the proposed method. For both the homogeneous and heterogeneous phantoms, the differences of DSCs and HDs are statistically significant at all the different noise levels: between the LS-GF and T40% ($p < 0.05$), between the LS-GF and RW ($p < 0.05$), and between the LS-GF and Otsu ($p < 0.05$).

The above experiments were performed assuming that the RTR was appropriate, including the entire tumor but excluding the bladder, which is difficult to achieve using current algorithms in clinical use. The RTR is usually obtained with erosion to exclude the signal produced by the bladder, which

may cause the RTR to be too small to include the entire object. The dependence on RTR of these methods should be evaluated. We extracted RTRs of different sizes from the cervix in our work, and performed a group of experiments based on these RTRs. Fig.7 shows the quantitative comparison of the mean DSCs and HDs obtained by the different methods on different RTRs on the heterogeneous phantom with the largest tumor at clinical noise level (i.e. $\alpha = 0.4$).

From this figure, we find that even when the RTR is very small (40% of the real cervix mask), the proposed LS-GF method also gives a good result (the DSC is nearly 90%) compared to other methods (the DSC is less than 80%), which means our method is more robust for different RTRs.

B. Clinical PET/CT Studies

The automatic recognition method of the RTR based on the improved FCM method was satisfactory in 24 of 27 datasets. For the other three cases that part of the tumor was excluded, R_0 could be substituted by rough manual delineation, which did not affect the final segmentation results. Based on the extracted R_0 , the proposed LS-GF method produced satisfactory segmentation results in all of the clinical studies. The cervix region that the other three methods were performed in was segmented from the CT images by a semi-automated Skeleton cuts method[42], and the segmentation results obtained by the three methods were based on the best parameters for our datasets.

Fig. 8 (c) and (d) show the segmentation results of the clinical cases based on the proposed LS-GF method and the traditional level set method. Similar to the simulation studies, we could also find that the proposed method could separate the tumor from the adjacent bladder while the CV model could not.

Another series of qualitative comparisons between the proposed method and the T40%, Otsu, RW are provided in Fig.9(b-d), which shows the initial PET images of three typical patients and their corresponding segmentation results. Tumor b represents the tumors with obvious heterogeneity, tumor c has the similar intensity with the bladder, while the concentration of bladder is much higher than tumor d. The segmentation of both tumors c and d are affected by the adjacent bladders to a large extent. Due to the heterogeneity, the segmentation results of T40% method only include part of the tumors. Due to the partial volume effect, the segmentation results of the other three methods include the signal of the bladder. In brief, our method brings more competitive results than others .

In order to test the robustness of the LS-GF method, quantitative comparisons of the DSC and HD between it, T40%, Otsu, RW, and the standard reference of the 27 clinical datasets with the gold standard is shown in Table III. We can see that our LS-GF method produces an obviously larger DSC (more than 90%, which is generally accepted as excellent) and smaller HD (5.61–9.97mm, about 1–2voxel), while the DSCs of other methods are less than 85%, and the HDs are more than 11mm (2voxel). The Student's *t*-test of the DSCs and HDs between the LS-GF method and T40% ($p < 0.05$), Otsu ($p < 0.05$), RW ($p < 0.05$) both indicated significant differences.

TABLE III
QUANTITATIVE COMPARISON OF DIFFERENT METHODS
ON CLINICAL STUDIES

Method	DSC (%)	HD (mm)
LS-GF	91.80 ± 2.46	7.79 ± 2.18
T40%	67.00 ± 12.90	15.59 ± 10.78
<i>P</i> -value	<10 ⁻⁶ *	6.41 × 10 ⁻⁴ *
Otsu	80.48 ± 6.78	16.24 ± 9.17
<i>P</i> -value	1.13 × 10 ⁻⁶ *	2.32 × 10 ⁻⁴ *
RW	82.10 ± 5.50	13.87 ± 7.12
<i>P</i> -value	2.34 × 10 ⁻⁵ *	7.23 × 10 ⁻³ *

* The difference between the T40% (or Otsu, RW) and the LS-GF is significant at the 0.05 level.

Additionally, the LS-GF method has the smallest standard deviation, which means it was stable across all the 27 tested clinical studies.

Figs. 10 and 11 show the Bland-Altman analysis of the SUVmean and the MTV obtained from different methods performed on the 27 clinical datasets versus the reference values. As illustrated in Fig.10(a), Bland-Altman analysis of SUVmean obtained by the proposed method compared with the gold standard shows 96% (26 of 27) points within 1.96 SDs (i.e. the 95% limits of agreement), with a bias of 0.10 and 1.96 SD limits of agreement of -0.24 and 0.43. Fig.10 (b), (c), and (d) show the Bland-Altman analysis of SUVmean obtained by the T40% method, Otsu method and RW method compared with gold standard. Only the Otsu method have 96% points within CI, with the bias and the ±1.96 SD of 0.27, -0.77, and 1.32, respectively. Fig.11 gives the Bland-Altman analysis of MTV obtained by the proposed method and the other three methods compared with the gold standard. For the proposed method shown in Fig.11(a), only one difference outside the 95% limits of agreement, the bias is only -1.67cm³, and the 1.96 SD is -16.37 to 13.10 cm³. All of the rest three plots also have 96% points within CI, with biases of -39.70cm³, -15.90cm³, -16.03cm³, and the ±1.96 SD of -98.84 and 19.45cm³, -67.71 and 35.90cm³, -57.18 and 25.12cm³, respectively. From the above description, we can find that both the proposed method and the Otsu method has a good agreement with the gold standard, but the proposed method produces the smallest bias. Additionally, according to of Figs.10(a) and 11(a), the max absolute differences of SUVmean and MTV within the CI is -0.14 and -16.37cm³ (both represented by solid blue circles), which are negligible compared to the corresponding average values 3.54 and 104.9cm³, and are acceptable in clinical practice. Therefore, we can conclude that our method has the best agreement with the gold standard compared to other three method.

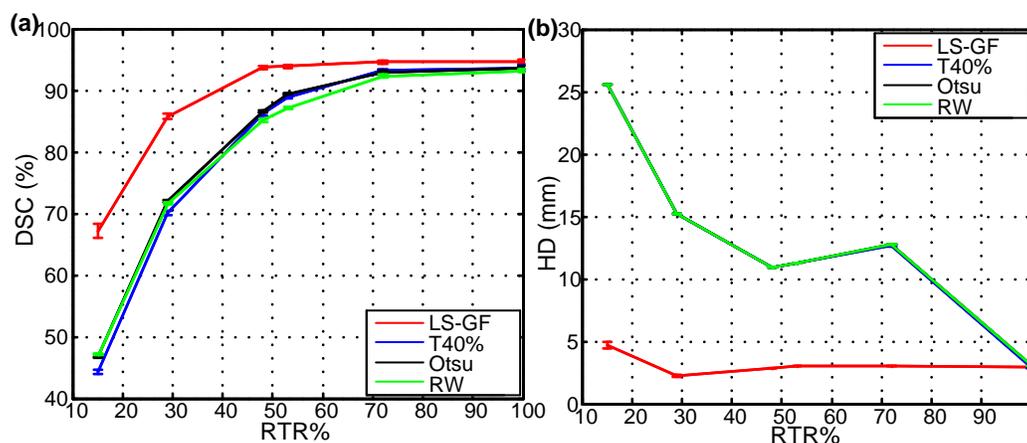


Fig.7. Quantitative comparison of the DSCs and HDs among different methods using different RTRs on heterogeneous phantom g. The horizontal axis, RTR% represents for the area of RTR/area of cervix $\times 100\%$, the vertical axis of (a) represents the mean and standard deviation of DSCs of the 20 realizations of phantom g at noise 0.4, and the vertical axis of (b) represents the mean and standard deviation of HDs.

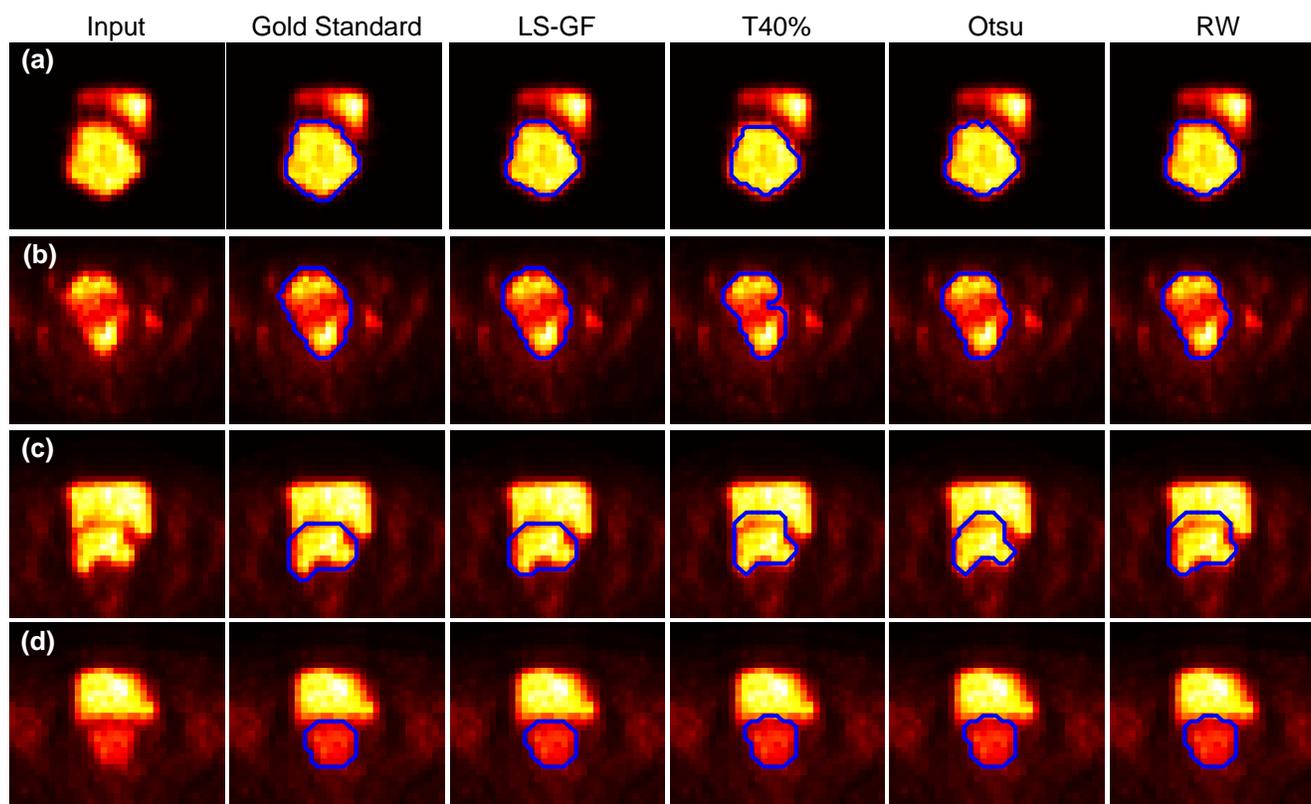


Fig.8. Qualitative comparisons of the evolution results under different iteration number between the proposed LS-GF method and the traditional level set method based on CV model: (a) and (b) are the segmentation results based on simulated phantom g at noise 0.4; (c) and (d) are the segmentation results based on a clinical case.

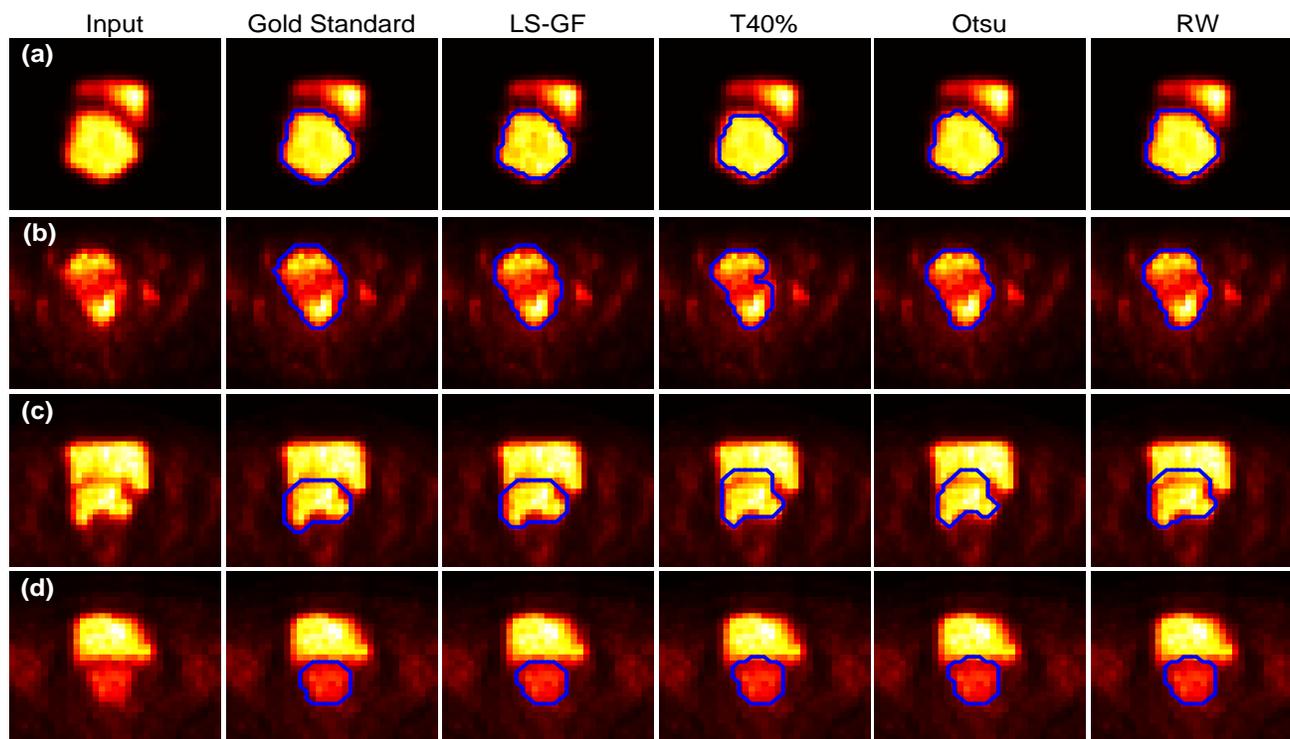


Fig.9. Cervical tumor segmentation results (delineated in blue) obtained with different methods in simulated phantom g at noise 0.4 (a) and three different clinical cases (b~d). For each case from left to right: a random slice of input PET images, the gold standard and the corresponding segmentation by our LS-GF method, T40%, Otsu, and RW.

IV. DISCUSSION

In this paper, we have proposed an automatic cervical tumor segmentation algorithm for the delineation of cervical tumors in PET/CT images. This method can be summarized as follows: 1) the construction of a novel evolution equation based on gradient fields to delineate the object automatically; 2) the construction of a hyper-image utilizing the anatomical and metabolic information to define the initial contour of the tumor. Owing to the above innovations, we were able to segment cervical tumors accurately and automatically, which is clinically relevant. Furthermore, we have also evaluated the feasibility of this algorithm in clinical practice by assessing the quantitative indices.

For the proposed LS-GF method, two hyper-parameters, i.e. λ and σ , need to be determined for a given c_0 ($c_0 = 2$ in our work). In fact, the larger the λ is, the larger the probability that the voxel belongs to the background is, which results to partitioning the fuzzy boundary of the tumor to the background with larger probability. In other words, the segmented tumor is smaller. On the contrary, the smaller the λ is, the larger the segmented tumor is, which is shown in Fig.12. Therefore, it is important to set λ properly. The choice of λ is influenced by the quality of the PET image and the size of the tumor. In real clinical applications, a certain amount of training datasets (five random patients were chosen in this work) are needed to find the proper λ . For different λ , the mean DSCs and HDs can be obtained as shown in Fig.13. From this figure, we could find when λ was set around 0.5, we could obtain nice results with larger DSC and smaller HD. Additionally, because of the

introduction of the Gaussian filter, the effect of the noise on LS-GF method is largely removed, especially for larger tumors, which can be seen from Fig.6. Even if the image is degenerated seriously at noise level of 1.2 as Fig.5, we could also obtain an acceptable segmentation results. In order to investigate the effect of the standard deviation σ , we performed additional experiments using a series of different σ values on five random clinical datasets. The obtained mean DSCs, HDs and the corresponding standard deviations of five random patients under different σ are shown in Fig.14. The figure illustrates that when σ was set within the range [0.5 1], the DSCs approached greater than 90% (accepted as excellent), and the HDs are also small (about 1–2 voxels) and stable.

Tables I, II, and III show that our method gives a much better results both in simulated studies and clinical studies compared to traditional T40%, Otsu, and RW methods, and the differences between our method and the traditional methods are statistically significant based on the Student's t -test both in the simulated studies at all the different noise levels and the clinical studies, which means the improvement of the new method is significant. However, when comparing the three tables, we can find that all the four methods perform well in the simulated studies. This is because the RTRs in the simulated studies include little signal of the bladder in our simulation, since the original intent of the simulation studies is to validate the accuracy of the method in the isolated tumors without the effect of the surrounding tissues. However, due to the short distance between the bladder and the cervix and the partial volume effect in the clinical studies, the border of the cervix may

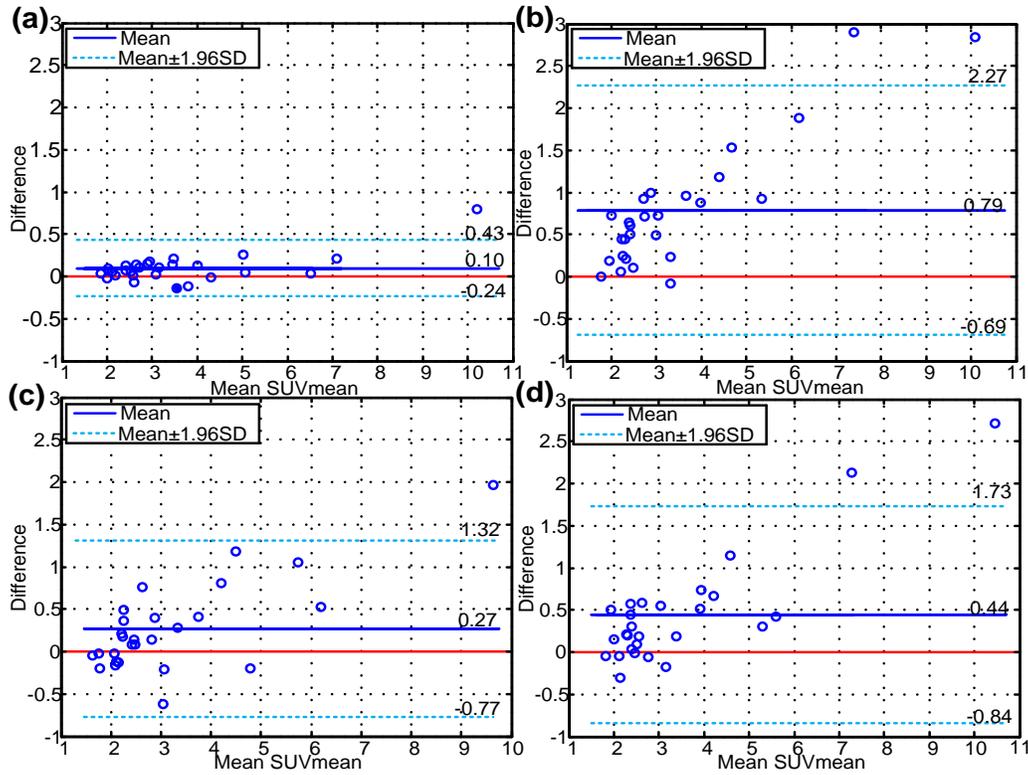


Fig.10. Bland-Altman plots of the difference of the SUVmean of tumors obtained by different automatic methods and the gold standard. The red solid line is the zero difference line. (a), (b), (c) and (d) show the SUVmean agreement between the LS-GF method, T40% method, Otsu method, RW method, and gold standard, respectively.

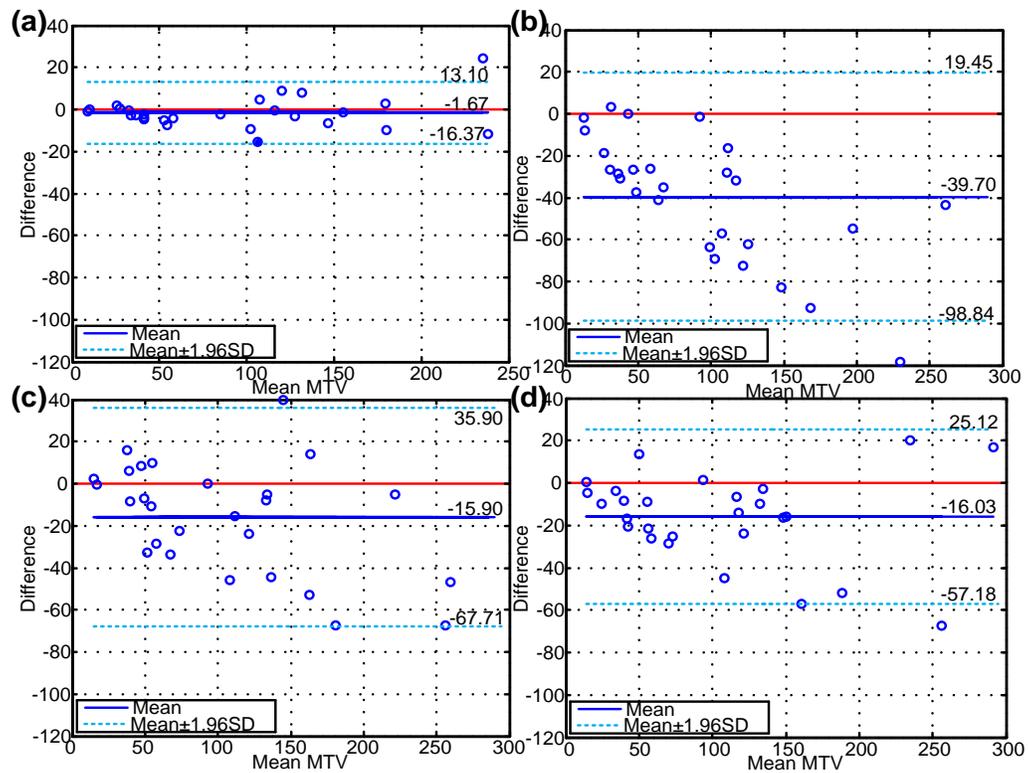


Fig.11. The Bland-Altman plots of the difference of the MTV of tumors obtained by different automatic methods and gold standard versus the mean of the two variables. The red solid line is the zero difference line. (a), (b), (c), and (d) show the MTV agreement between the LS-GF method, T40% method, Otsu method, RW method and the gold standard, respectively.

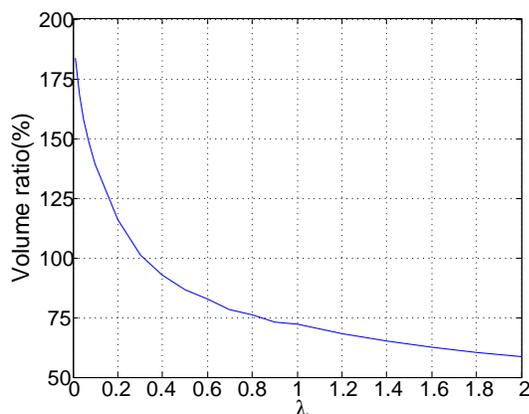


Fig. 12. The relationship between λ and the size of the segmented tumor on the clinical datasets. The horizontal axis represents the different value of λ , and the vertical axis represents the ratio of the segmented tumor volume to the gold standard.

include the signal belonging to bladder, which means, even if the RTR is exactly the accurate cervix, the three traditional methods T40%, Otsu, and RW might not delineate the tumor accurately, as illustrated by the segmentation results shown in Fig.9 (c and d). This effect can be decreased by the erosion of RTR, which may lead to the RTR too small to include the whole tumor. However, the three traditional methods could only be performed in the given region, which leads to insufficient segmentations. On the contrary, the LS-GF method achieves the final segmentation by expanding the ZLS centered on the RTR, and acting on the whole image rather than just the given regions, which could guarantee the sufficient segmentation under the premise of excluding the bladder. Fig.7 shows the quantitative comparison between these methods using different RTRs. Therefore, no matter how small the RTR is, the proposed method could obtain good results as long as RTR includes all the centers of the tumor.

Due to the introduction of the L term into the classical CV model, the proposed method possesses the advantage of separating the tumor and bladder automatically. Fig.8 shows the detailed evolution processes of the proposed method and the classical CV model, from which we can see the LS-GF method could exclude the bladder lumen successfully when the method achieves the convergence compared to traditional CV models. In other words, another important improvement of the proposed method is its automatic separating capacity besides its accuracy.

To validate the clinical application significance of the proposed method, we conducted indirect comparison of clinically significant quantitative measures SUVmean and MTV with Bland-Altman plots. From Figs.10 and 11, we can find that the proposed method produces the smallest bias and standard deviation, and more than 95% points are within the consistent interval. Additionally, the max absolute differences of SUVmean and MTV within the consistent interval is too small (only 3.95% and 15.6%) to be neglected compared to the corresponding average values. In brief, the proposed method has a high agreement with the clinical gold standard, and can be applied in clinical cervical tumor segmentation.

The most common used method in the CC segmentation is the traditional threshold-based segmentation techniques. However, the interactive morphologic operation and manual operation are needed to eliminate the bladder for every cases as Miller *et al* used[28], which is time consuming and subjective. Subsequently, Roman-Jimenez *et al* proposed a semi-automatic method by combining the region-growing adaptive threshold method (RGAT) and fusion- and Gaussian-mixture-based classification (FGMC) method to separate the bladder and tumor. On one hand, the RGAT suffers from poor reproducibility[43]; on the other hand, it was necessary to decide whether to perform FGMC visually for each case. In addition, the voxel seed, the location of which may affect the final segmentation results, was needed to be selected by interactive program. Compared to the above methods, our method could obtain the final segmentation results automatically and quickly without interactive manual interpretation, which is meaningful for the following quantitative analysis of large datasets. Recently, Arbones *et al* proposed an automatic tumor segmentation in cervical cancer based on the Chan-Vese (CV) models and post-processing[31]. This method obtains the final tumor through detecting and removing the bladder from the positive areas (i.e. the results of the CV method). The detection of the bladder is based on the assumption that it is the largest segmented object. When we tested this method on one of our clinical dataset whose tumor is larger than the bladder in the positive areas, we found this method may confuse the bladder and tumor, such as Fig.15 (a2). Additionally, this method only segmented part of the tumors in many of our cases, such as Fig.15 (b2). Moreover, this method performs poor on our datasets due to the low resolution. Though it could obtain the tumor through interpolation, it takes much longer time (about 40min) compared to our method (about 4s) for every cases under the same environment. Therefore, our method performs better both in accuracy and efficiency.

This method could also work well in the experiments of segmenting rectum cancer, which is also affected by the adjacent bladder in some cases. For other pathologies, such as lung tumors with heart or mediastinum physiological uptake and head and neck tumors with nearby uptake, the proposed LS-GF method could also work well based on the accurate initial contour. It should be noted that the automatic definition of the initial contour based on the FCM may not be applicable because of the different tissue specificities, which needs further study in the future.

V. CONCLUSION

An automatic segmentation algorithm for the quantitative analysis of cervical tumors was proposed and evaluated. Experiments on both phantom and clinical studies demonstrate the accuracy, effectiveness, and robustness of the LS-GF method. Further assessment of the quantitative indices indicates the feasibility of this method for accurate quantitative analysis of cervical tumors in clinical practice.

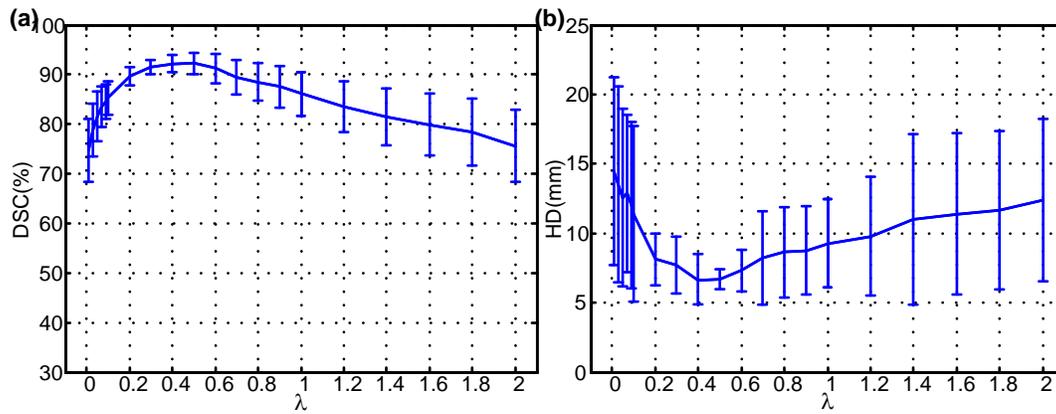


Fig.13. Profiles of mean DSCs and HDs obtained under different λ . The horizontal axis stands for λ , the vertical axis of (a) represents the mean and standard deviation of DSCs of the five cases, and the vertical axis of (b) represents the mean and standard deviation of HDs obtained from the five cases.

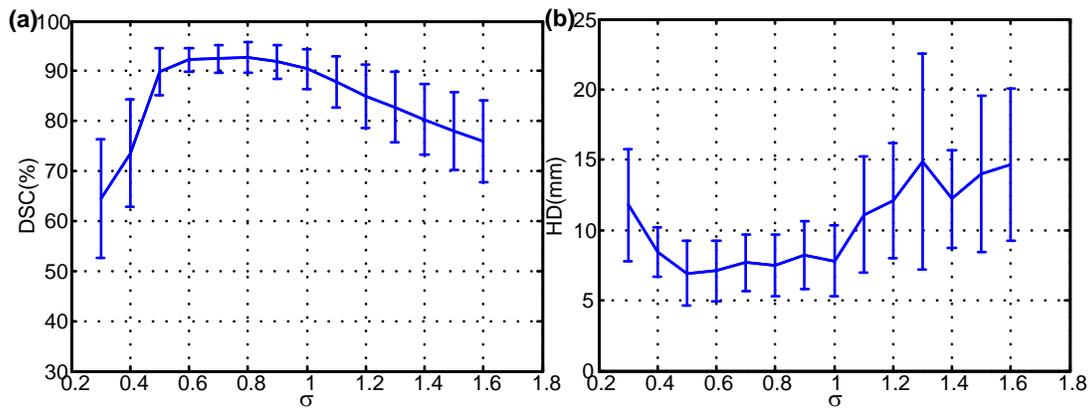


Fig.14. The mean DSCs and HDs of the random five cases obtained under different σ . The horizontal axis stands for sigma, the vertical axis of (a) represents the mean and standard deviation of DSCs, and the vertical axis of (b) represents the mean and standard deviation of HDs obtained from the five cases.

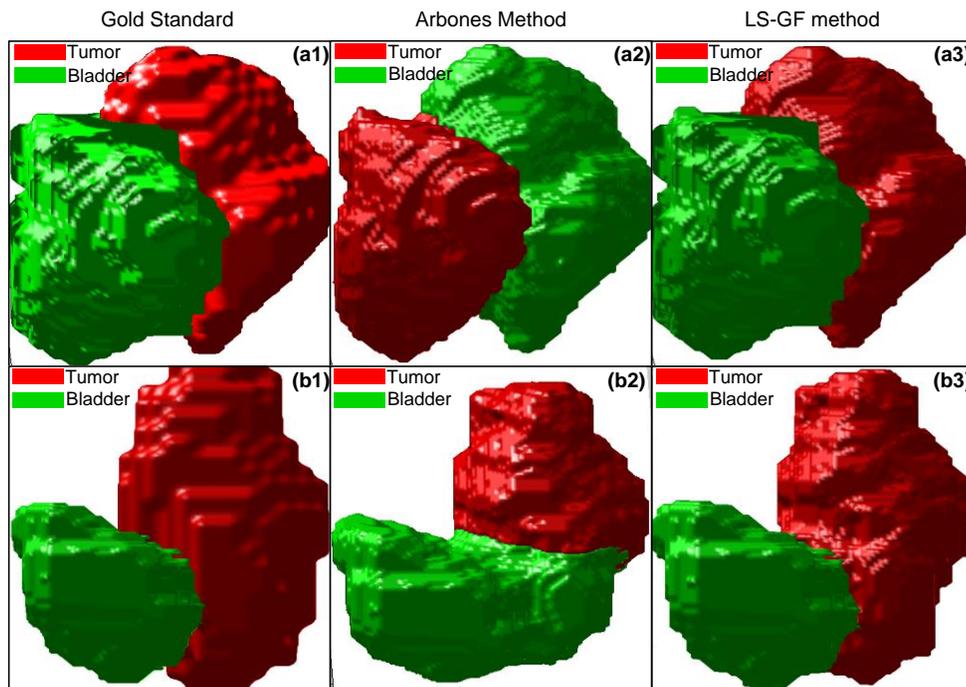


Fig.15. Cervical tumor segmentation results obtained with different methods of two typical clinical cases. The red part means the tumor while the green part means the bladder. For each case from left to right are: the gold standard, the corresponding segmentation by Arbones's Method and our LS-GF method.

REFERENCES

- [1] I. A. f. R. o. Cancer, *World Cancer Report 2014*. Lyon: International Agency for Research on Cancer Press, 2014.
- [2] B. S. Kim, I. J. Kim, S.-J. Kim, H.-Y. Nam, K. J. Pak, K. Kim, *et al.*, "The prognostic value of the metabolic tumor volume in FIGO stage IA to IIB cervical cancer for tumor recurrence: measured by F-18 FDG PET/CT," *Nucl Med Mol Imaging*, vol. 45, pp. 36-42, 2011.
- [3] S. Chicklore, V. Goh, M. Siddique, A. Roy, P. K. Marsden, and G. J. Cook, "Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis," *Eur J Nucl Med Mol Imaging*, vol. 40, pp. 133-140, 2013.
- [4] I. El Naqa, P. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, *et al.*, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," *attem Recogn*, vol. 42, pp. 1162-1171, 2009.
- [5] M. Hatt, M. Majdoub, M. Vallieres, F. Tixier, C. C. Le Rest, D. Groheux, *et al.*, "18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi-Cancer Site Patient Cohort," *J Nucl Med*, vol. 56, pp. 38-44, 2015.
- [6] M. Hatt, F. Tixier, C. Cheze Le Rest, O. Pradier, and D. Visvikis, "Robustness of intratumour 18F-FDG PET uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma," *Eur J Nucl Med Mol Imaging*, vol. 40, pp. 1662-1671, 2013/10/01 2013.
- [7] M. MacManus, U. Nestle, K. E. Rosenzweig, I. Carrio, C. Messa, O. Belohlavek, *et al.*, "Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007," *Radiother Oncol*, vol. 91, pp. 85-94, 2009.
- [8] U. Nestle, S. Kremp, and A.-L. Grosu, "Practical integration of [¹⁸F]-FDG-PET and PET-CT in the planning of radiotherapy for non-small cell lung cancer (NSCLC): The technical basis, ICRU-target volumes, problems, perspectives," *Radiother Oncol*, vol. 81, pp. 209-225, 2006.
- [9] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23-27, 1975.
- [10] L. Drever, W. Roa, A. McEwan, and D. Robinson, "Iterative threshold segmentation for PET target volume delineation," *Med Phys*, vol. 34, pp. 1253-1265, 2007.
- [11] J. A. van Dalen, A. L. Hoffmann, V. Dicken, W. V. Vogel, B. Wiering, T. J. Ruers, *et al.*, "A novel iterative method for lesion delineation and volumetric quantification with FDG PET," *Nucl Med Commun*, vol. 28, pp. 485-493, 2007.
- [12] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET volumes by iterative image thresholding," *J Nucl Med*, vol. 48, pp. 108-114, 2007.
- [13] S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET," *Med Phys*, vol. 37, pp. 1309-1324, 2010.
- [14] M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET," *IEEE T Med Imaging*, vol. 28, pp. 881-893, 2009.
- [15] M. Hatt, C. Cheze le Rest, P. Descourt, A. Dekker, D. De Ruyscher, M. Oellers, *et al.*, "Accurate Automatic Delineation of Heterogeneous Functional Volumes in Positron Emission Tomography for Oncology Applications," *International Journal of Radiation Oncology*Biophysics*, vol. 77, pp. 301-308, 5/1/ 2010.
- [16] M. Hatt, F. Lamare, N. Bousson, A. Turzo, C. Collet, F. Salzenstein, *et al.*, "Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET," *Phys Med Biol*, vol. 52, p. 3467, 2007.
- [17] Y. Guo, Y. Feng, J. Sun, N. Zhang, W. Lin, Y. Sa, *et al.*, "Automatic Lung Tumor Segmentation on PET/CT Images Using Fuzzy Markov Random Field Model," *Comput Math Method Med*, vol. 2014, 2014.
- [18] E. Day, J. Betler, D. Parda, B. Reitz, A. Kirichenko, S. Mohammadi, *et al.*, "A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients," *Med Phys*, vol. 36, pp. 4349-4358, 2009.
- [19] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, *et al.*, "Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images," *Med Image Anal*, vol. 17, pp. 929-945, 2013.
- [20] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Comput Biol Med*, vol. 50, pp. 76-96, 2014.
- [21] F. Hofheinz, J. Langner, J. Petr, B. Beuthien-Baumann, J. Steinbach, J. Kotzerke, *et al.*, "An automatic method for accurate volume delineation of heterogeneous tumors in PET," *Med Phys*, vol. 40, p. 082503, 2013.
- [22] M. Abdoli, R. Dierckx, and H. Zaidi, "Contourlet-based active contour model for PET image segmentation," *Med Phys*, vol. 40, p. 082507, 2013.
- [23] D. Markel, H. Zaidi, and I. El Naqa, "Novel multimodality segmentation using level sets and Jensen-Rényi divergence," *Med Phys*, vol. 40, pp. -, 2013.
- [24] C.-Y. Hsu, C.-Y. Liu, and C.-M. Chen, "Automatic segmentation of liver PET images," *Comput Med Imag Grap*, vol. 32, pp. 601-610, 2008.
- [25] X. Geets, J. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *Eur J Nucl Med Mol Imaging*, vol. 34, pp. 1427-1438, 2007/09/01 2007.
- [26] H. K. Pannu, F. M. Corl, and E. K. Fishman, "CT Evaluation of Cervical Cancer: Spectrum of Disease 1," *Radiographics*, vol. 21, pp. 1155-1168, 2001.
- [27] K.-P. Wong, X. Zhang, and S.-C. Huang, "Improved Derivation of Input Function in Dynamic Mouse [¹⁸F] FDG PET Using Bladder Radioactivity Kinetics," *Mol Imaging Biol*, vol. 15, pp. 486-496, 2013.
- [28] T. R. Miller and P. W. Grigsby, "Measurement of tumor volume by PET to evaluate prognosis in patients with advanced cervical cancer treated by radiation therapy," *Int J Radiat Oncol*, vol. 53, pp. 353-359, 2002.
- [29] G. Roman-Jimenez, J. Leseur, A. Devillers, and J. David, "Segmentation and characterization of tumors in 18F-FDG PET-CT for outcome prediction in cervical cancer radio-chemotherapy," presented at the Image-Guidance and Multimodal Dose Planning in Radiation Therapy: a MICCAI Workshop 2012.
- [30] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE T Image Process*, vol. 10, pp. 266-277, 2001.
- [31] D. R. Arbonès, H. G. Jensen, A. L. Jakobsen, P. M. af Rosenschöld, A. E. Hansen, C. Igel, *et al.*, "Automatic FDG-PET-based tumor and metastatic lymph node segmentation in cervical cancer," in *SPIE Medical Imaging*, 2014, pp. 903441-903441-8.
- [32] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun Pur Appl Math*, vol. 42, pp. 577-685, 1989.
- [33] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE T Image Process*, vol. 19, pp. 3243-3254, 2010.
- [34] J. A. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science* vol. 3: Cambridge university press, 1999.
- [35] K. Zhang, H. Song, and L. Zhang, "Active contours driven by local image fitting energy," *attem Recogn*, vol. 43, pp. 1199-1206, 2010.
- [36] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," 1973.
- [37] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*: Kluwer Academic Publishers, 1981.
- [38] M. C. Adams, T. G. Turkington, J. M. Wilson, and T. Z. Wong, "A systematic review of the factors affecting accuracy of SUV measurements," *Am J Roentgenol*, vol. 195, pp. 310-320, 2010.
- [39] F. J. Brooks and P. W. Grigsby, "The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake," *J Nucl Med*, vol. 55, pp. 37-42, 2014.
- [40] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp. 297-302, 1945.
- [41] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: Measuring error on simplified surfaces," in *Comput Graph Forum*, 1998, pp. 167-174.
- [42] D. Xiang, J. Tian, F. Yang, Q. Yang, X. Zhang, Q. Li, *et al.*, "Skeleton cuts—An efficient segmentation method for volume rendering," *IEEE T Vis Comput Gr*, vol. 17, pp. 1295-1306, 2011.
- [43] M. Hatt, N. Bousson, C. Cheze-Le Rest, O. Pradier, and D. Visvikis, "Metabolically active volumes automatic delineation methodologies in PET imaging: review and perspectives," *Cancer/Radiothérapie*, p. in press, 2011.