



**HAL**  
open science

# Spatial Evidential Clustering with Adaptive Distance Metric for Tumor Segmentation in FDG-PET Images

Chunfeng Lian, Su Ruan, Thierry Denoeux, Hua Li, Pierre Vera

► **To cite this version:**

Chunfeng Lian, Su Ruan, Thierry Denoeux, Hua Li, Pierre Vera. Spatial Evidential Clustering with Adaptive Distance Metric for Tumor Segmentation in FDG-PET Images. *IEEE Transactions on Biomedical Engineering*, 2018, 65 (1), pp.21-30. 10.1109/TBME.2017.2688453 . hal-01637193

**HAL Id: hal-01637193**

**<https://hal.science/hal-01637193v1>**

Submitted on 5 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatial Evidential Clustering with Adaptive Distance Metric for Tumor Segmentation in FDG-PET Images

Chunfeng Lian, Su Ruan, Thierry Dencoux, Hua Li, and Pierre Vera

**Abstract**—While the accurate delineation of tumor volumes in FDG-PET is a vital task for diverse objectives in clinical oncology, noise and blur due to the imaging system make it a challenging work. In this paper, we propose to address the imprecision and noise inherent in PET using Dempster-Shafer theory, a powerful tool for modeling and reasoning with uncertain and/or imprecise information. Based on Dempster-Shafer theory, a novel evidential clustering algorithm is proposed and tailored for the tumor segmentation task in 3D. For accurate clustering of PET voxels, each voxel is described not only by the single intensity value but also complementarily by textural features extracted from a patch surrounding the voxel. Considering that there are a large amount of textures without consensus regarding the most informative ones, and some of the extracted features are even unreliable due to the low-quality PET images, a specific procedure is included in the proposed clustering algorithm to adapt distance metric for properly representing the clustering distortions and the similarities between neighboring voxels. This integrated metric adaptation procedure will realize a low-dimensional transformation from the original space, and will limit the influence of unreliable inputs via feature selection. A Dempster-Shafer-theory-based spatial regularization is also proposed and included in the clustering algorithm, so as to effectively quantify the local homogeneity. The proposed method has been compared with other methods on the real-patient FDG-PET images, showing good performance.

**Index Terms**—Lung Tumor Segmentation, Evidential  $c$ -Means, Adaptive Distance Metric, Feature Selection, Spatial Regularization, Dempster-Shafer Theory, PET Images.

## I. INTRODUCTION

**P**OSITRON emission tomography (PET), with the radio-tracer fluoro-2-deoxy-D-glucose (FDG), is an advanced imaging tool generally used in clinical oncology for diagnosis, staging, and restaging of tumors. In recent years, FDG-PET has also played an important role in adaptive radiation therapy treatment planning process. The goal of adaptive radiation therapy is to improve radiation treatment by incorporating

the specificities of individual patients, as well as those of target tumors, to re-optimize the treatment plan early on during the course of treatment [1]. The utilization of FDG-PET in adaptive radiation therapy has great benefits [2], including 1) as a complement to computed tomography (CT), FDG-PET can help to modify the gross tumor volume (GTV) definition; 2) FDG-PET images can be used to define subvolumes, namely biological target volumes (BTVs), within the tumor target, so as to include tumor biological characteristics in adaptive radiation therapy; 3) some studies, e.g., [3]–[5], have shown that the functional information provided by PET images can predict early the treatment outcome before the end of therapy, offering significant evidence for the adaptation of a more effective treatment plan.

While the accurate delineation of tumor volumes in FDG-PET is a pivotal step for all the purposes discussed above, noisy and blurring images due to the acquisition system make it a challenging work. To this end, diverse automatic or semi-automatic PET segmentation algorithms have been proposed, which include thresholding methods [6], [7], region growing and level set [8], [9], statistical methods [10], graph-based methods [11], [12], and clustering methods [13]–[15], etc.

As the most commonly used approach owning simple and intuitive nature, thresholding methods usually define a constant [6] or an adaptive [7] threshold to differentiate lesions from background. The disadvantage is that these methods are sensitive to noise, and have limited performance facing small or heterogeneous positive tissues. Region growing methods also need to select a threshold value as the stopping criterion. To improve the robustness of thresholding segmentation against noise, region growing methods take into account the spatial context in PET images; however, the performance of these methods usually depends heavily on the initialization of the segmentation. Statistical methods assume that positive tissues and surrounding volumes obey different statistical distributions of intensities, e.g., a mixture of Gaussian densities [10]. This kind of methods is robust to noise and partial volume effect caused by the low-resolution imaging system; however, they are sensitive to heterogeneous uptake of positive tissues. Graph-based algorithms, e.g., random walks (RW) [16], can effectively combine global cue with local smoothness by defining foreground and background seeds as hard constraints. Based on previous work with good performance [11], [16], an improved version of the classical RW, namely 3D-LARW method [12], was proposed recently for the segmentation of inhomogeneous or small tumor volumes. The potential

Chunfeng Lian is with the Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France, and also with the Université de Rouen, QuantIF, LITIS, Rouen, France (e-mail: chunfeng.lian@utc.fr).

Su Ruan is with the Université de Rouen, QuantIF, LITIS, Rouen, France (e-mail: su.ruan@univ-rouen.fr).

Thierry Dencoux is with the Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France (e-mail: thierry.dencoux@hds.utc.fr).

Hua Li is with the Department of Radiation Oncology, Washington University School of Medicine, Saint Louis, Missouri, United States (e-mail: huli@radonc.wustl.edu).

Pierre Vera is with the Department of Nuclear Medicine, Centre Henri-Becquerel, Rouen, France, and also with the Université de Rouen, QuantIF, LITIS, Rouen, France (e-mail: pierre.vera@chb.unicancer.fr).

disadvantage of these RW methods is that their performance can be influenced by the quality of seeds.

Clustering methods are suitable for PET image segmentation, especially considering that positive tissues are usually inhomogeneous with varying shapes among different patients, which are hard to be dealt with supervised learning strategies. In view of the wide application of fuzzy  $c$ -means (FCM) in multimodality medical image segmentation tasks [9], [17], Belhassen et al. proposed a robust approach, called FCM-SW [14], working specifically for segmenting heterogeneous tumors in PET images. In the objective function of FCM-SW, the spatial context of image voxels is included for modeling the uncertainty and inaccuracy inherent in PET, thus leading to more stable segmentation than the classical FCM. As an extension of FCM and possibilistic clustering [18], an evidential  $c$ -means algorithm (ECM) [19] was proposed in the framework of Dempster-Shafer theory [20]. A spatial version of ECM, namely SECM [15], was then proposed recently for lung tumor delineation in multi-tracer PET images. In the objective function of SECM, the local homogeneity is quantified by the weighted sum of the intensity distances from the neighborhood of each voxel to the cluster prototypes. Finding an alternative way to directly model the spatial information in the framework of Dempster-Shafer theory seems to be more appropriate, and may also further enhance the performance of ECM in low-quality PET images.

It is also worth noting that in the clustering methods mentioned above only intensity values were used to assign voxels into different clusters. Textural features [21]–[24], which describe the spatial environment surrounding each voxel, are very likely to provide complementary information for more accurate segmentation. However, the challenge to include textures in tumor segmentation is that a large amount of textures can be extracted, but there is no consensus regarding the most informative ones; in addition, some of the extracted features may be unreliable or inaccuracy due to the noisy and blurring nature of PET images. Abounding researches [25]–[27] have shown that a distance metric adapted to the data at hand could effectively improve the performance of clustering algorithms, and of other unsupervised or semi-supervised learning methods. However, since many of the existing methods were not designed specifically for tackling high-dimensional data that contain unreliable input features, their performance may decline with this kind of imperfect information.

Noise and imprecision modeling is of great concern for reliable PET image segmentation [28], e.g., for the blurring and inhomogeneous positive tissues shown in Fig. 1. In our study this critical issue is addressed via Dempster-Shafer theory (DST), a powerful tool for modeling and reasoning with uncertain and/or imprecise information. As a solid foundation for the proposed approach, DST is an extension of both the set-membership approach and probability theory, and has been proven to have greater expressive power [20]. In the framework of DST, we propose a new clustering algorithm tailored for the delineation of tumor volumes in low-quality 3D PET images. The proposed method has three main objectives: 1) to add textural features as complementary information for intensities that used in the above methods, so as to obtain

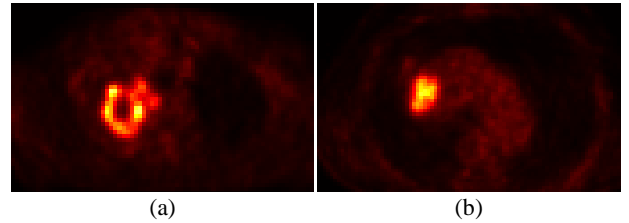


Fig. 1. Blurring FDG-PET images shown in the axis plane for two different patients, where large intra- and inter-tumor heterogeneity can be observed.

more accurate segmentation; 2) to properly adapt distance metric for well representing the clustering distortions and the similarities between neighboring voxels rather than using the simple Euclidean distance directly. Different from the existing methods (e.g. [25]), a sparsity constraint is included in the distance metric updating procedure to realize a feature selection via a low-dimensional feature transformation, thus limiting the influence of unreliable input features on the output segmentation; 3) to define a new energy function in the framework of DST using the concept of Markov random field (MRF). By reason that MRF offers a reliable way to consider spatial information [29], [30], the new MRF-based energy function is included in the objective function of ECM, and acts as a spatial regularization to effectively quantify the local homogeneity of PET image voxels.

The rest of this paper is organized as follows. The background on DST and the original ECM is recalled in Section II. The proposed method is then introduced in Section III. In Section IV, the proposed method is evaluated by a cohort of real-patient FDG-PET images, and the segmentation performance is compared with that of other methods. Finally, we conclude paper in Section V.

## II. BACKGROUND

As our method is based on Dempster-Shafer theory (DST) and the original ECM, they are briefly reviewed in Section II-A and II-B, respectively.

### A. Dempster-Shafer Theory

DST is also known as the theory of belief functions or evidence theory. As a generalization of both probability theory and the set-membership approaches, it has shown remarkable applications in divers fields, such as image and vision analysis [31]–[33], data classification [34]–[39], and data clustering [40]–[43], etc. In our previous work, two different methods [44], [45] based on DST were also proposed to predict cancer treatment outcome using PET images and clinical characteristics. DST has two main components, i.e., the quantification of a piece of evidence and the combination of different items of evidence.

1) *Evidence Quantification*: DST is a formal framework for reasoning under uncertainty based on the modeling of evidence [46]. Let  $\omega$  be a variable taking values in a finite domain  $\Omega = \{\omega_1, \dots, \omega_c\}$ , called the *frame of discernment*. An item of evidence regarding the actual value of  $\omega$  can be represented by a *mass function*  $m$  on  $\Omega$ , defined from the

powerset  $2^\Omega$  to the interval  $[0, 1]$ , such that  $\sum_{A \subseteq \Omega} m(A) = 1$ . Each number  $m(A)$  denotes a *degree of belief* attached to the hypothesis that " $\omega \in A$ ". Function  $m$  is said to be normalized if  $m(\emptyset) = 0$ , where the empty set denotes hypothesis beyond the finite domain  $\Omega$ . Any subset  $A$  with  $m(A) > 0$  is called a *focal element*. If all focal elements are singletons,  $m$  is said to be *Bayesian*; it is then equivalent to a probability distribution. A mass function  $m$  with only one focal element is said to be *categorical* and is equivalent to a set.

Corresponding to a normalized mass function  $m$ , we can associate *credibility* and *plausibility* function from  $2^\Omega$  to  $[0, 1]$  defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \quad (1)$$

Quantity  $Bel(A)$  can be interpreted as the degree to which the evidence supports  $A$ , while  $Pl(A)$  can be interpreted as the degree to which the evidence is not contradictory to  $A$ . A mass function  $m$  can also be transformed into a probability function for reasoning and decision-making. In Smet's transferable belief model [46], the *pignistic probability* corresponding to  $m$  is defined as:

$$BetP(\omega_q) = \sum_{A \subseteq \Omega; \omega_q \in A} \frac{m(A)}{|A|}, \quad \forall \omega_q \in \Omega, \quad (2)$$

where  $|A|$  is the cardinality of the subset  $A$ .

2) *Evidence Combination*: In DST, beliefs are elaborated by aggregating different items of evidence. The basic mechanism for evidence combination is *Dempster's rule* of combination [20]. Let  $m_1$  and  $m_2$  be two mass functions derived from independent items of evidence. They can be fused via Dempster's rule to induce a new mass function  $m_1 \oplus m_2$  defined as

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - Q} \sum_{B \cap C = A} m_1(B)m_2(C), \quad (3)$$

where  $Q = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  measures the *degree of conflict* between evidence  $m_1$  and  $m_2$ .

### B. Evidential $c$ -Means

Let  $\{X_1, \dots, X_n\}$  be a collection of feature vectors in  $\mathbb{R}^p$  describing  $n$  objects belonging to the set of clusters  $\Omega = \{\omega_1, \dots, \omega_c\}$ . ECM is grounded on a new concept of partition, namely the *credal partition* [40], which extends the concepts of hard, fuzzy, and possibilistic partition by allocating, for each object, a mass of belief, not only to single clusters, but also to any subset of the whole frame  $\Omega$ . Each single cluster  $\omega_k$ ,  $k \in \{1, \dots, c\}$ , is represented by a prototype  $V_k \in \mathbb{R}^p$ . Then, for each nonempty subset  $A_j \subseteq \Omega$ , a centroid  $\bar{V}_j$  is defined as the barycenter of the prototypes associated with the singletons in  $A_j$ , i.e.,  $\bar{V}_j = \frac{1}{c_j} \sum_{k=1}^c s_{kj} V_k$ , where  $s_{kj}$  is binary, and it equals 1 iff  $\omega_k \in A_j$ ; while  $c_j = |A_j|$  denotes the cardinality of  $A_j$ .

Let  $\mathbf{V}$  denotes a matrix of size  $(c \times p)$  composed of the coordinates of the cluster centers such that  $V_{kq}$  is the  $q$ th component of the prototype  $V_k$ . ECM looks for a credal

partition matrix  $\mathbf{M} = (m_{ij})$  of size  $(n \times 2^c)$  and for a matrix  $\mathbf{V}$  by minimizing the following objective function:

$$\mathcal{J}_{ecm}(\mathbf{M}, \mathbf{V}) = \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (4)$$

subject to the constraints  $m_{ij} \geq 0$ ,  $m_{i\emptyset} \geq 0$ , and

$$\sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} m_{ij} + m_{i\emptyset} = 1, \quad \forall i = 1, \dots, n, \quad (5)$$

where Euclidean metric  $d_{ij}^2 = \|X_i - \bar{V}_j\|^2$  is adopted in the original ECM to quantify the distortion between  $X_i$  and the center of the focal set  $A_j$ . Scalar  $m_{ij}$  denotes the mass of the object  $X_i$  allocated to the credal cluster  $A_j$ ; while  $m_{i\emptyset}$  denotes that allocated to the empty set, and  $\delta$  is a weighting parameter. The empty set is used for the detection of outliers. Coefficient  $\alpha \geq 0$  controls the degree of penalization of the subsets according to their cardinality ( $\alpha = 2$  by default), and coefficient  $\beta > 1$  controls the fuzziness of the credal partition ( $\beta = 2$  by default).

Evidential clustering has been successfully applied in various domains. For instance, Zhou et al. [43] proposed an evidential  $c$ -medoids method for the clustering of relational data with pairwise similarities. Makni et al. [41] extended the original ECM for the segmentation of prostate in multi-parametric magnetic resonance images. Liu et al. [42] proposed a variant of ECM, based on a modified definition of the clustering distortions. All the above applications or variants of evidential clustering are different from the method that will be proposed in Section III, as the latter one includes a specific spatial regularization and distance metric adaptation procedure to delineate tumor volumes in PET images.

## III. METHOD

A new approach, called Evidential  $c$ -Means integrating adaptive distance metric and spatial regularization (ECM-MS), is proposed in this section for tumor segmentation in PET.

Let  $\{X_1, \dots, X_n\}$  be a collection of feature vectors in  $\mathbb{R}^p$  describing  $n$  voxels in a volume of interest (VOI). The VOI is an user-defined box that includes the target tumor. We assume that all the voxels belong either to the background (i.e. hypothesis  $\omega_1$ ) or to the positive tissue (i.e. hypothesis  $\omega_2$ ), without existence of outliers. Thus, the whole frame of clusters is set as  $\Omega = \{\omega_1, \omega_2\}$ . Each mass function  $m$  satisfies  $m(\{\omega_1\}) + m(\{\omega_2\}) + m(\Omega) \equiv 1$ , without existence of  $m(\emptyset)$ . As  $m(\Omega)$  measures the ambiguity regarding the clusters  $\omega_1$  and  $\omega_2$ , blurring boundary and severe heterogeneous region will be assigned to  $m(\Omega)$ .

### A. Spatial Regularization

According to the spatial prior of a PET volume, the credal partition matrix  $\mathbf{M} = \{m_i\}_{i=1}^n$  that we want to learn can be viewed as a specific MRF, where each mass function  $m_i$  is a random vector in  $\mathbb{R}^3$ . Let  $\Phi = \{\Phi(i)\}_{i=1}^n$  be a 3D neighborhood system, where  $\Phi(i) = \{1, \dots, T\}$  is the set of the  $T$  neighbors of a voxel  $i$ , excluding  $i$ . The corresponding masses of voxels in  $\Phi(i)$  are  $\{m_1^i, \dots, m_T^i\}$ ,

while the feature vectors of these voxels are  $\{X_1^i, \dots, X_T^i\}$ . In the concept of MRF, the distribution of  $m_i$  is assumed to be depended on the predefined 3D neighborhood system, i.e.,  $p(m_i | \{m_j\}_{j \neq i}^n) = p(m_i | \{m_t^i\}_{t \in \Phi(i)})$ . Thus, the distribution of  $\mathbf{M}$  can be represented as  $p(\mathbf{M}) = Z^{-1} \exp\{-U(\mathbf{M})\}$ , where  $Z$  is a normalizing constant, and  $U(\mathbf{M})$  is an *energy function* of the form

$$U(\mathbf{M}) = \eta \sum_{i=1}^n \sum_{t \in \Phi(i)} C(i, t), \quad (6)$$

where scalar  $\eta > 0$  is a hyper-parameter, also called the inverse temperature in physics, which controls the degree of local homogeneity in a VOI. The potential function  $\sum_{t \in \Phi(i)} C(i, t)$  measures the smoothness around voxel  $i$ , in which  $C(i, t)$  denotes the inconsistency between voxel  $i$  and its neighbor  $t$ . In the framework of Dempster-Shafer theory (DST),  $C(i, t)$  can be defined as  $C(i, t) = \gamma_{it} dm_{it}^2$ , where  $dm_{it}^2$  denotes the *dissimilarity between  $m_i$  and  $m_t^i$* , while  $\gamma_{it}$  is a weighting factor that will be calculated in the feature space.

In this study, the metric defined by Jousselme et al. [47] is adopted to represent the dissimilarity between mass functions of any two adjacent voxels, as it has been commonly used to calculate the conflict between two different pieces of evidence that modeled by DST. As the result, the  $dm_{it}^2$  between  $m_i$  and  $m_t^i$ , where  $t \in \Phi(i)$ , is quantified as

$$dm_{it}^2 = (m_i - m_t^i) \mathbf{Jac}(m_i - m_t^i)', \quad (7)$$

where  $\mathbf{Jac}$  is a positive definite matrix whose elements are Jaccard indexes, i.e.,  $\mathbf{Jac}(A, B) = |A \cap B| / |A \cup B|$ ,  $\forall A, B \in 2^\Omega \setminus \emptyset$ . Since  $\Omega = \{\omega_1, \omega_2\}$  in this study, we have

$$\mathbf{Jac} = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}. \quad (8)$$

It is worth noting that (7) satisfies the requirements for a valid distance metric. In addition, it effectively accounts for the interaction between the focal elements of  $\Omega$  [47].

For instance, let  $m_i$ ,  $m_1^i$ , and  $m_2^i$  be three masses on the same frame  $\Omega = \{\omega_1, \omega_2\}$ , with the form of

$A$	$m_i(A)$	$m_1^i(A)$	$m_2^i(A)$
$\{\omega_1\}$	0.8	0.4	0.2
$\{\omega_2\}$	0	0	0.6
$\Omega$	0.2	0.6	0.2

According to this table,  $m_i(A)$  is more consistent with  $m_1^i(A)$  than with  $m_2^i(A)$ , as  $m_i(A)$  and  $m_1^i(A)$  both have mass of belief on  $\{\omega_1\}$  and no mass of belief on the opposite hypothesis  $\{\omega_2\}$ ; while  $m_2^i$  is strongly concentrated on  $\{\omega_2\}$ . As a comparison to (7), if we quantify the dissimilarities via the simple Euclidean metric,  $dm_{i1}^2$  and  $dm_{i2}^2$  will inappropriately be identical and equal 0.72. On the contrary, the dissimilarities deduced by (7) are  $\mathbf{dm}_{i1}^2 = 0.36$  and  $\mathbf{dm}_{i2}^2 = 0.72$ , respectively, which measure the distance more reasonably than the Euclidean metric. Therefore, this measure is used to define the specific MRF energy function (6). It acts as a spatial regularization to adaptively quantify the local homogeneity during the clustering.

The new objective function of ECM including this MRF-based spatial regularization is proposed as

$$\mathcal{J}_{ecm}^s(\mathbf{M}, \mathbf{V}) = \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^2 m_{ij}^2 d_{ij}^2 + \eta \sum_{i=1}^n \sum_{t \in \Phi(i)} \gamma_{it} dm_{it}^2, \quad (9)$$

subject to the constraints  $m_{ij} \geq 0$ , and

$$\sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} m_{ij} = 1, \quad \forall i = 1, \dots, n, \quad (10)$$

where  $\mathbf{V}$  and  $d_{ij}^2$  have the same form as that in (4). Matrix  $\mathbf{M} = (m_{ij})$  has  $n$  rows and 3 columns, in which  $m_{ij}$  is the mass of belief attached to the hypothesis that "the object  $X_i$  belongs to the credal cluster  $A_j$ ". The second term of (9) is the spatial regularization, in which  $dm_{it}^2$  (i.e. (7)) measures the dissimilarity between  $m_i$  and  $m_t^i$ , while  $\gamma_{it}$  is a weighting factor. The scalar  $\eta > 0$  controls the influence of this regularization. It should be predetermined by taking into account the size of the tumor for segmentation. More discussions with respect to  $\eta$  will be presented in Section IV-D3.

### B. Adaptive Distance Metric

Apart from intensity of voxels, in this study we also attempt to include textural features in ECM as complementary information for more accurate segmentation. The challenge to this is that a large amount of textures can be extracted, but without prior knowledge concerning the most informative features; additionally, these relatively high-dimensional feature vectors are very likely to contain unreliable variables due to the noisy and blurring nature of the PET imaging system. Hence, to obtain a desired segmentation, an adaptive distance metric and feature selection procedure is necessary.

In our previous work [37], a supervised method has been proposed to learn a low-rank dissimilarity metric for improving the performance of distance-based classifiers on high-dimensional datasets containing unreliable and imprecise features. Also, as has been discussed in Section I, a distance metric adapted to the data will be beneficial for ECM and other clustering methods. But adapting distance metric in clustering is distinct from metric learning in classification, as in the former case only weak knowledge can be gathered to guide the procedure. Therefore, in line with but different to [37], we look for a matrix  $\mathbf{D} \in \mathbb{R}^{p \times q}$  during clustering, under the constraint  $q \ll p$ , by which the dissimilarity between any two feature vectors, say  $X_1$  and  $X_2$ , can be represented as

$$d^2(X_1, X_2) = (X_1 - X_2) \mathbf{D} \mathbf{D}' (X_1 - X_2)'. \quad (11)$$

In other words, matrix  $\mathbf{D}$  transforms the original feature space to a low-dimensional subspace, where important input features will have a strong impact when calculating the dissimilarity. To find such a transformation matrix  $\mathbf{D}$ , the distances  $d_{ij}^2$  used in (9) is calculated via (11). The spatial regularization that defined by (6) is also used to adapt the distance metric. More specifically, for each voxel  $i$  and its neighbor  $t$ , we define the weighting factor that used in (9) as  $\gamma_{it} = (X_i - X_t^i) \mathbf{D} \mathbf{D}' (X_i - X_t^i)'$ . Then during the minimization of (9), a large dissimilarity  $dm_{it}^2$  between  $m_i$  and  $m_t^i$  will reveal that current distance

measure (11) is inadequate, and it should be adjusted at the next step to reduce the dissimilarity between  $X_i$  and  $X_t^i$ , so as to bring the two adjacent voxels closer together.

Based on the above analysis, the objective function (9) integrating adaptive distance metric can be updated as

$$\begin{aligned} \mathcal{J}_{ecm}^{ms}(\mathbf{M}, \mathbf{V}, \mathbf{D}) = & \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^2 m_{ij}^2 (X_i - \bar{V}_j) \mathbf{D} \mathbf{D}' (X_i - \bar{V}_j)' \\ & + \eta \sum_{i=1}^n \sum_{t \in \Phi(i)} dm_{it}^2 (X_i - X_t^i) \mathbf{D} \mathbf{D}' (X_i - X_t^i)' \\ & + \lambda \|\mathbf{D}\|_{2,1} - \log \left( (\bar{X}_{\omega_1} - \bar{X}_{\omega_2}) \mathbf{D} \mathbf{D}' (\bar{X}_{\omega_1} - \bar{X}_{\omega_2})' \right), \end{aligned} \quad (12)$$

subject to the constraints  $m_{ij} \geq 0$  and (10). In (12), matrix  $\mathbf{M} = (m_{ij})$  and  $\mathbf{V}$  have the same form as that in (9). The dissimilarities between neighboring mass functions, i.e.,  $dm_{it}^2$ , are still quantified by (7). The  $\ell_{2,1}$ -norm sparsity regularization (i.e. the third term)

$$\|\mathbf{D}\|_{2,1} = \sum_{i=1}^p \sqrt{\sum_{j=1}^q D_{i,j}^2} \quad (13)$$

is included to select input features during feature transformation. By forcing rows of  $\mathbf{D}$  to be zero, this sparsity term only selects the most reliable input features to calculate the linear transformation, thus controlling the influence of unreliable input features on the clustering result. Scalar  $\lambda$  is a hyper-parameter that controls the influence of this regularization. The last term of (12) is used to prevent the objective function being trivially solved with  $\mathbf{D} = 0$ , which collapses all the features vectors into a single point. Vectors  $\bar{X}_{\omega_1}$  and  $\bar{X}_{\omega_2}$  are two predetermined prototypes (or seeds) for the positive tissue and the background, respectively. A simple and easy initialization of them will be discussed in Section III-C.

Finally, a desired distance metric determined by (12) should satisfy 1) neighboring voxels are similar (realizing via the second term), and 2) the tumor seeds and the background seeds are widely separated (realizing via the last term).

### C. Optimization

Subject to  $m_{ij} \geq 0$  and (10),  $\forall i \in \{1, \dots, n\}$  and  $j \in \{1, 2, 3\}$ , the objective function (12) can be minimized in an EM-like iterative optimization scheme, so as to automatically determine  $\mathbf{M}$  for a given case. Based on the whole frame of clusters  $\Omega$  defined at the beginning of this section, the number of clusters  $c$  equals 2 in our applications.

1) *Initialization*: To guide the clustering procedure of ECM-MS, especially to control the integrated metric updating step, we firstly initialize the mass functions and the cluster centers via the original ECM algorithm. A very small number of voxels are then automatically selected as the seeds with predefined cluster labels. More specifically, based on the initial mass functions, image voxels are classified into three credal clusters, i.e., cluster  $\{\omega_1\}$ , cluster  $\{\omega_2\}$ , and cluster  $\Omega$ . To ensure the reliability and to control the number of the selected seeds, the tumor seeds are determined as the voxels whose

---

### Algorithm 1: ECM-MS

---

**Input** feature vectors  $\{X_1, \dots, X_n\} \in \mathbb{R}^p$ ; the spatial neighborhood  $\Phi(i)$  of each voxel  $i$ ; the hyper-parameters  $\eta$  and  $\lambda$ ; initial  $\mathbf{M}^{(0)}$ ,  $\mathbf{V}^{(0)}$ , and  $\mathbf{D}^{(0)}$ ; the tumor and background seeds ;

**for**  $l = 1, 2, \dots$  **do**

  E-step: calculate  $\mathbf{M}^{(l)}$  using the efficient interior-point algorithm [48] with (14),  $\mathbf{M}^{(l-1)}$ ,  $\mathbf{V}^{(l-1)}$ , and  $\mathbf{D}^{(l-1)}$  ;

  M-step I: calculate  $\mathbf{V}^{(l)}$  according to (16) and  $\mathbf{M}^{(l)}$  ;

  M-step II: calculate  $\mathbf{D}^{(l)}$  via the Beck-Teboulle proximal gradient algorithm [49] with (17),  $\mathbf{M}^{(l)}$ ,  $\mathbf{V}^{(l)}$ , and  $\mathbf{D}^{(l-1)}$  ;

**if** no significant change of  $\mathcal{J}_{ecm}^{ms}$  **then**  
   | break;

**end**

**end**

**Output** the final  $\mathbf{M}^*$ ,  $\mathbf{V}^*$ , and  $\mathbf{D}^*$ ;

---

intensity values are higher than that of the third quartile voxel in the cluster  $\{\omega_1\}$ ; while, the background seeds are determined as the boundary of the VOI. After that, the mass functions for the tumor and background seeds are fixed as  $m(\{\omega_1\}) \equiv 1$  and  $m(\{\omega_2\}) \equiv 1$ , respectively. In addition, the two prototypes, i.e.,  $\bar{X}_{\omega_1}$  and  $\bar{X}_{\omega_2}$ , used in (12) are calculated as the barycenters of the tumor and background seeds, respectively. The output dimension, namely the number of columns  $q$  in  $\mathbf{D}$ , is then determined by applying principle component analysis on all the feature vectors  $\{X_1, \dots, X_n\}$ . The initial  $\mathbf{D}$  is constructed by the top 95% eigenvectors.

Then, the optimization procedure alternates between cluster assignment (i.e.  $\mathbf{M}$  estimation) in the E-step, and both prototype determination (i.e.  $\mathbf{V}$  estimation) and metric adaptation (i.e.  $\mathbf{D}$  estimation) in the M-step.

2) *E-step*: Given  $\mathbf{V}$  and  $\mathbf{D}$ , the minimization of (12) only relates to the first two terms, which turns to be a quadratic problem with respect to the mass functions  $\mathbf{M} = (m_{ij})$ . The derivative of (12) concerning the mass function  $m_i \in \mathbb{R}^3$ ,  $\forall i \in \{1, \dots, n\}$ , can be written as

$$\frac{\partial \mathcal{J}_{ecm}^{ms}}{\partial m_i} = 2m_i \mathbf{B} + 2\eta \sum_{j \in \Phi(i)} d_{ij}^2 (m_i - m_j) \mathbf{Jac}, \quad (14)$$

where the matrix  $\mathbf{Jac}$  is defined by (8),  $d_{ij}^2 = d^2(X_i, X_j)$  is measured by (11), and

$$\mathbf{B} = \begin{pmatrix} c_1^2 d^2(X_i, \bar{V}_1) & 0 & 0 \\ 0 & c_2^2 d^2(X_i, \bar{V}_2) & 0 \\ 0 & 0 & c_3^2 d^2(X_i, \bar{V}_3) \end{pmatrix}, \quad (15)$$

where  $d^2(X_i, \bar{V}_j)$  is also measured by (11). Based on the derivation (14), an efficient interior-point algorithm with a limited-memory BFGS approximation of the Hessian matrix [48] is adopted to solve the quadratic problem, so as to obtain the matrix  $\mathbf{M}$  at current step.

3) *M-step I*: The updating of the prototypes is only influenced by the first term of (12). Let  $f_j = \sum_{i=1}^n c_j^2 m_{ij}^2$  and

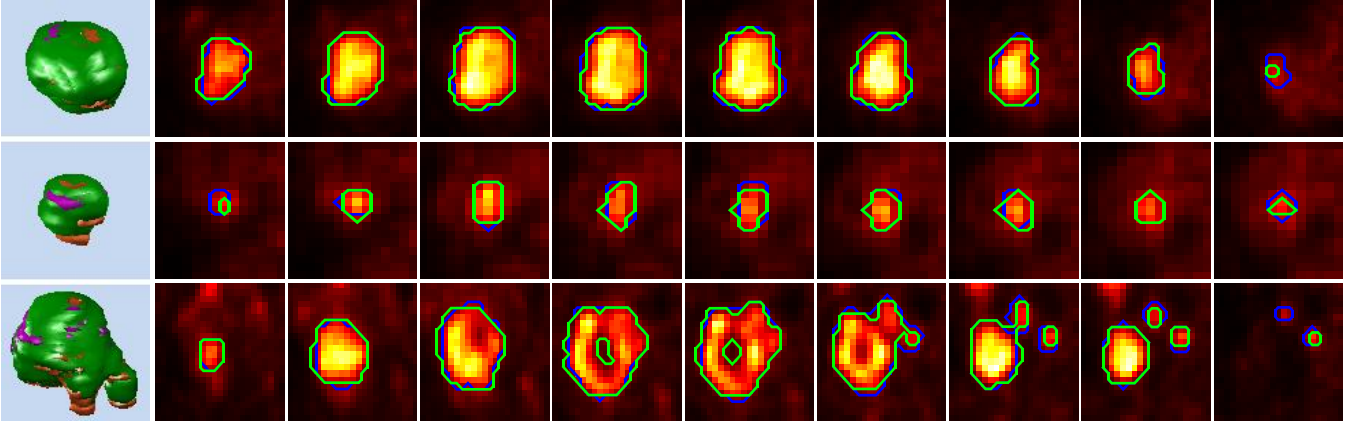


Fig. 2. Three different tumors delineated by ECM-MS. The first column demonstrates volumes in 3D, where, based on the manually segmentation by clinicians, the green region consists of the true positive and true negative voxels, the magenta region consists of the false positive voxels, while the orange region consists of the false negative voxels. For each tumor volume in the first column, more detailed results, slice by slice in the axial plane, are shown in the following columns correspondingly, where the contours delineated by ECM-MS (green line) are compared with that delineated by clinicians (blue line).

$g_j = \sum_{i=1}^n c_j^2 m_{ij}^2 X_i, \forall j \in \{1, 2, 3\}$ , the centers of the clusters  $\{\omega_1\}$  and  $\{\omega_2\}$  are calculated, respectively and directly, as

$$\begin{cases} V_1 = \frac{2f_2(2g_1+g_3)+f_3(g_1-g_2)}{4f_1f_2+f_3(f_1+f_2)}; \\ V_2 = \frac{2f_1(2g_2+g_3)+f_3(g_2-g_1)}{4f_1f_2+f_3(f_1+f_2)}. \end{cases} \quad (16)$$

4) *M-step II*: It is worth noting that the objective function (12), excluding the third term, is differentiable as a function of the transformation matrix  $\mathbf{D}$ ; while, the third term (i.e. the sparsity regularization) is only partly smooth with a singularity at  $\mathbf{D} = 0$ . For this reason, the proximal Forward-Backward splitting (FBS) algorithms [49] are efficient alternatives to solve the metric updating problem formulated in this step. More precisely, the derivative of the differentiable part of (12) concerning  $\mathbf{D}$  can be written as

$$\begin{aligned} \frac{\partial(\cdot)}{\partial \mathbf{D}} &= 2 \sum_{i=1}^n \sum_{A_j \neq \emptyset} c_j^2 m_{ij}^2 (X_i - \bar{V}_j)' (X_i - \bar{V}_j) \mathbf{D} \\ &+ 2\eta \sum_{i=1}^n \sum_{t \in \Phi(i)} dm_{it}^2 (X_i - X_t^i)' (X_i - X_t^i) \mathbf{D} \quad (17) \\ &- \frac{2(\bar{X}_{\omega_1} - \bar{X}_{\omega_2})' (\bar{X}_{\omega_1} - \bar{X}_{\omega_2}) \mathbf{D}}{(\bar{X}_{\omega_1} - \bar{X}_{\omega_2}) \mathbf{D} \mathbf{D}' (\bar{X}_{\omega_1} - \bar{X}_{\omega_2})'}, \end{aligned}$$

based on which the Beck-Teboulle proximal gradient algorithm [49], an improved version of the classical FBS methods with computational simplify and fast convergence rate, is executed to obtain a required distance metric at current step.

The optimization procedure of the proposed ECM-MS method is briefly summarized in Algorithm 1.

#### D. Post-Processing & Decision-Making

To further improve the segmentation performance, mass functions obtained by Algorithm 1 can be post-processed. For each voxel  $i$ , the mass functions  $\{m_1^i, \dots, m_T^i\}$  in the 3D neighborhood  $\Phi(i)$ , i.e., voxels surrounding  $i$ , are viewed as  $T$  independent pieces of evidence regarding the cluster label of  $i$ . We assume that the reliability of each evidence  $m_t^i$  is inversely proportional to the spatial distance between  $i$  and  $t$ . Let this spatial distance be  $s_{it}^2$ . Then, based on Dempster's discounting

procedure [20], each piece of evidence  $m_t^i, \forall t \in \{1, \dots, T\}$ , can be weighted by a coefficient  $\mu_t = \exp(-s_{it}^2)$ , so as to obtain a discounted mass function

$$\begin{cases} wm_t^i(\{\omega_j\}) = \mu_t m_t^i(\{\omega_j\}), \quad \forall j = 1, 2, \\ wm_t^i(\Omega) = 1 - \sum_{j=1}^2 wm_t^i(\{\omega_j\}). \end{cases} \quad (18)$$

Using the Dempster's rule of combination (3), the discounted mass functions obtained by (18) are fused with the mass function  $m_i$  to output a renewed mass function  $m_i$ . On the other hand, the above procedure can also be regarded as a filtering operation in a small cubic window.

After that, a hard partition of image voxels can be obtained via maximizing the pignistic probability calculated by (2). In the final crisp segmentation, only the connected component with the maximum number of positive voxels will be recognised as the target tumor, thus effectively handling complicated segmentation tasks, e.g., the situation that target tumor is close to heart or mediastinum with high-uptake in the VOI. The same operation has also been applied to all the compared methods in Section IV, so as to improve their performance.

## IV. EXPERIMENTS AND RESULTS

In this section, the proposed ECM-MS was evaluated by the FDG-PET images acquired for non-small cell lung cancer (NSCLC) patients. The performance of ECM-MS was compared with that of a constant thresholding method using 40% of the maximum intensity in the lesion (T40%) [6], an adaptive thresholding method (TAD) [7], 3D-LARW [12], FCM-SW [14], SECM [15], and also the original ECM [19].

#### A. Material and Features

The FDG-PET images of 20 NSCLC patients were studied. These patients were injected by an average activity of FDG of  $261 \pm 48$  MBq. The obtained PET acquisitions have the same anisotropic resolution of  $4.06 \times 4.06 \times 2$  mm<sup>3</sup>, and were quantified using standardized uptake values (SUV). The tumor lesions were then manually delineated by experienced clinicians, with the volumes range from 1.9 mL to 135.8 mL.



TABLE I  
THE DICE COEFFICIENTS (DSC), THE HAUSDORFF DISTANCES (HD) IN Voxel, AND THE MEAN SURFACE DISTANCE (MSD) IN Voxel OBTAINED BY DIFFERENT SEGMENTATION METHODS ON THE FDG-PET IMAGES FOR THE NSCLC PATIENTS. ALL THE RESULTS ARE PRESENTED AS MEAN $\pm$ STD.

	T40%	TAD	3D-LARW	ECM	SECM	FCM-SW	ECM-MS
<b>DSC</b>	0.733 $\pm$ 0.114	0.706 $\pm$ 0.129	0.800 $\pm$ 0.078	0.722 $\pm$ 0.142	0.763 $\pm$ 0.13	0.839 $\pm$ 0.094	<b>0.862 <math>\pm</math> 0.048</b>
<b>HD</b>	5.119 $\pm$ 5.089	5.002 $\pm$ 5.528	5.642 $\pm$ 5.573	5.699 $\pm$ 4.732	5.125 $\pm$ 3.823	3.996 $\pm$ 3.514	<b>2.544 <math>\pm</math> 1.349</b>
<b>MSD</b>	0.647 $\pm$ 0.513	0.725 $\pm$ 0.627	0.560 $\pm$ 0.535	0.697 $\pm$ 0.701	0.559 $\pm$ 0.347	0.351 $\pm$ 0.295	<b>0.216 <math>\pm</math> 0.098</b>

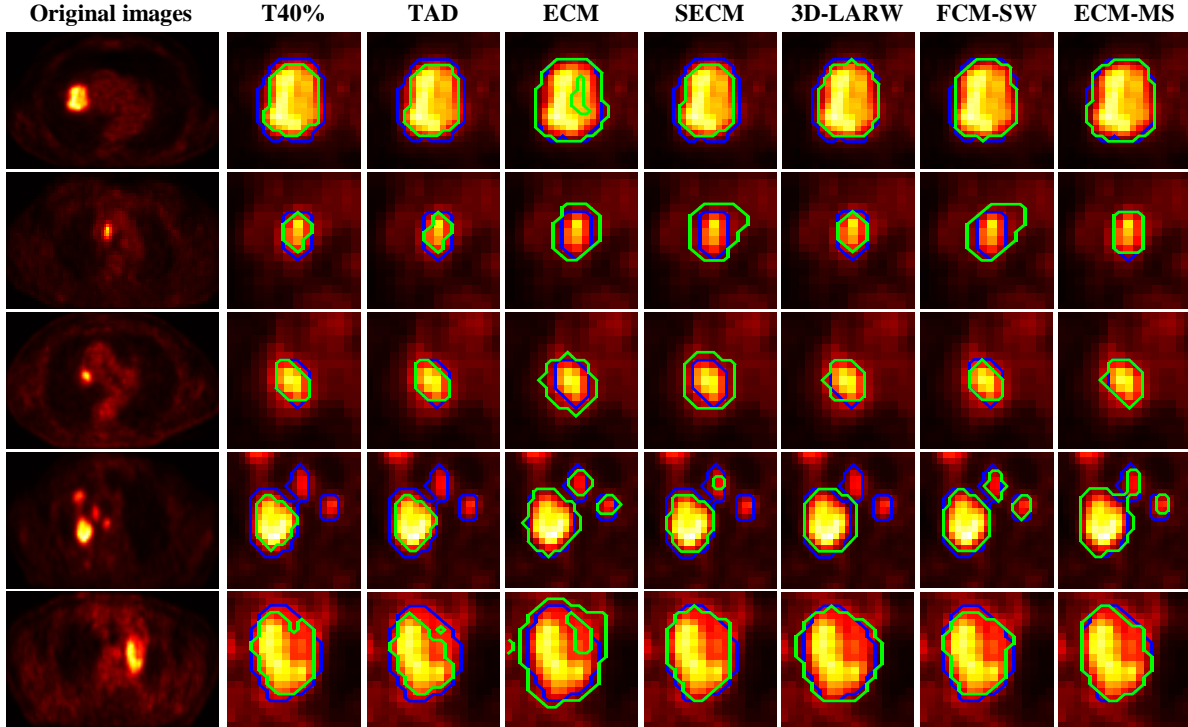


Fig. 3. Contours delineated by different methods (from the second column to the last column) for five different tumor volumes shown in the axis plane. The first column represents the input images with contours delineated by expert clinicians. The delineation by the seven algorithms (green line) is compared with that by clinicians (blue line) in the following columns.

Considering that the image resolution is anisotropic, a  $(3 \times 3)$  window in 2-D was defined to extract features to be used in the proposed ECM-MS. Using this window, the average SUV, the maximum SUV, the minimum SUV, the range of SUV (i.e. maximum–minimum), and the standard deviation of SUV were calculated as features for the centering voxel. The gray level size zone matrix (GLSZM) [23] was adopted to extract seven texture features, as its effectiveness in PET image characterization has already been evaluated [4]. Similarly, the gray-level co-occurrence matrix (GLCM) [21] was also utilized to extract fifteen features. To sum up, for each voxel, a 28-dimensional feature vector was extracted, consisting of 6 SUV-based, 7 GLSZM-based, and 15 GLCM-based features.

### B. Evaluation Criteria

Regarding the manual delineation by clinicians as the reference, all the segmentation methods were evaluated by three criteria, i.e., the Dice coefficient (DSC), the Hausdorff distance (HD), and the mean absolute surface distance (MSD). Let  $S_1$  and  $S_2$  be two segmentations with the corresponding boundaries  $B_1$  and  $B_2$ . Then,  $DSC = 2|S_1 \cap S_2| / (|S_1| + |S_2|)$ , which measures the overlap between the regions of two

different segmentations. The HD metric defines the maximum distance between the contours of two different segmentations; while, as the complementary, the MSD metric measures how much on average the two contours differ. More information with respect to HD and MSD can be found in [50].

### C. Results

To demonstrate the performance, as examples, three different PET volumes segmented by ECM-MS are shown in Fig. 2, where the three rows (from the top to the bottom) correspond to a large tumor, a small tumor, and a heterogenous tumor, respectively. The first column of Fig. 2 presents the tumor volumes in 3D. Using the manually segmentation by clinicians as the reference, the green region consists of the true positive and true negative voxels, the magenta region consists of the false positive voxels, while the orange region consists of the false negative voxels. For each tumor volume, the second column to the last column of Fig. 2 show the corresponding results slice by slice in the axis plane (from the top to the bottom), where the green and blue line represent the contours delineated by ECM-MS and clinicians, respectively. As can be seen, the delineation by ECM-MS is in consistent with that by clinicians for all the three examples. It is also worth noting that, for the severely heterogenous tumor shown in the third



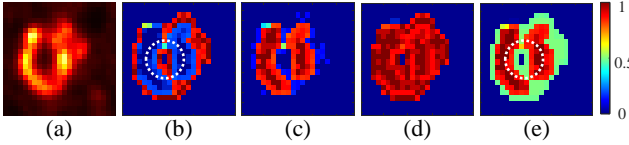


Fig. 4. (a) a heterogeneous FDG positive tissue; (b) represents the corresponding map of mass function  $m(\{\Omega\})$  obtained by ECM-MS; (c) and (d) are the credibility and plausibility map for the hypothesis of tumor (i.e.  $Bel(\{\omega_1\})$  and  $Pl(\{\omega_1\})$ ), respectively; while (e) is the corresponding pignistic probability map (i.e.  $BetP(\omega_1)$ ).

row, ECM-MS blocked some voxels out from the solid tumor delineated by clinicians. It indicates that the proposed method may could offer helpful information regarding the radiation necrosis during RT or ART. This property will be discussed in more detail in the next subsection.

The segmentation results obtained by all the methods on the 20 FDG-PET volumes are summarized and compared in Table I, from which we can find that the proposed method obtained better performance, both DSC and HD, than the other six algorithms. To be more comprehensive, the visual examples obtained by these methods are also presented in Fig. 3 for comparison. The first column of Fig. 3 presents the axis slices of five different tumors, where the first row is a slice corresponds to a large tumor, the second and the third rows represent two small tumors, while the last two rows represent two heterogenous tumors. The second column to the last column of Fig. 3 compare the contours delineated by the seven different methods (green line) with that delineated by clinicians (blue line). As can be seen, the contours delineated by the proposed method (the last column) are more in consistent with the reference contours in this experiment, especially for the small tumors and heterogenous tumors.

#### D. Discussion

1) *Uptake Analysis in FDG-PET*: In addition to the crisp segmentation results presented in Section IV-C, ECM-MS can also be adopted to gain deeper insight into the FDG uptake. As an example, a heterogeneous positive tissue is shown in Fig. 4 (a). Then, for all the voxels available in (a), subfigure (b) presents the masses of belief that attached to the whole frame of hypothesis (i.e.  $m(\Omega)$ ). It can be found that the possible radiation necrosis in the white circle, and the blurry boundary of the target are assigned high values of  $m(\Omega)$ , which means that these high uncertainty and imprecision regions were successfully recognised by ECM-MS. The similar knowledge can also be obtained in subfigure (e), which shows us the pignistic probability with respect to the hypothesis of tumor, i.e.,  $BetP(\omega_1)$  calculated by (2). Based on (1), the credibility and plausibility map for the hypothesis of tumor (i.e.  $Bel(\{\omega_1\})$  and  $Pl(\{\omega_1\})$ ) are presented, respectively, in (c) and (d) of Fig. 4. As a piece of additional information, they can be regarded as the lower and upper bounds for the probability of each voxel belonging to the cluster of tumor.

2) *Role of different modules in (12)*: To evaluate the influence of the spatial regularization, the sparsity regularization, and the post-processing step (i.e. fusion of mass functions updated by (18) in Section III-D) on the final segmentation,

TABLE II  
SEGMENTATION PERFORMANCE WITHOUT THE SPATIAL REGULARIZATION (NO SPATIAL), THE SPARSITY REGULARIZATION (NO  $\|\mathbf{D}\|_{2,1}$ ), AND THE UNCERTAINTY REDUCTION (NO POST-PROCESSING), RESPECTIVELY.

	No spatial	No $\ \mathbf{D}\ _{2,1}$	No post-processing	ECM-MS
<b>DSC</b>	$0.789 \pm 0.097$	$0.846 \pm 0.067$	$0.857 \pm 0.047$	<b><math>0.862 \pm 0.048</math></b>
<b>HD</b>	$3.535 \pm 2.385$	$2.685 \pm 1.449$	<b><math>2.544 \pm 1.349</math></b>	<b><math>2.544 \pm 1.349</math></b>
<b>MSD</b>	$0.389 \pm 0.205$	$0.248 \pm 0.117$	$0.217 \pm 0.098$	<b><math>0.216 \pm 0.098</math></b>

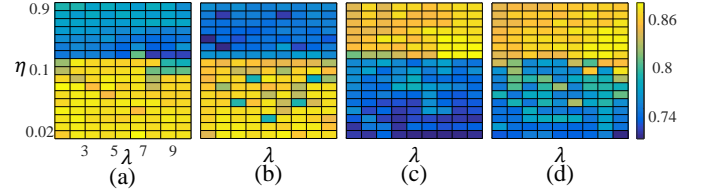


Fig. 5. The Dice coefficient (i.e. the intensity value) as a function of  $\lambda$  and  $\eta$ . (a) to (c) correspond to four different tumors delineated by clinicians with the volumes of 51.20 mL, 135.80 mL, 18.33 mL and 8.10 mL, respectively.

we orderly excluded them from ECM-MS. Then, the corresponding segmentation results are summarized in Table II. We can find that the spatial penalty and the sparsity penalty effectively improved the accuracy of the proposed method. Based on this foundation, the post-processing operation further enhanced the segmentation performance, as it can refine the contour of output target, and can also deal with false negative voxels inside the tumor contour.

3) *Parameter Setting*: The two hyper-parameters utilized in ECM-MS, i.e.  $\eta$  and  $\lambda$  of (12), control the influence of the spatial regularization and the influence of the sparsity regularization, respectively. To maximize the segmentation performance, they should be determined taking into account the size of the tumor. As an illustration, we orderly chose a  $\eta$  and a  $\lambda$  from  $\{0.01, \dots, 0.09, 0.1, \dots, 0.9\}$  and  $\{1, \dots, 10\}$ , respectively. Then, ECM-MS was applied on two relatively large tumors (volumes of 51.20 mL and 135.80 mL delineated by clinicians) and two relatively small tumors (volumes of 18.33 mL and 8.10 mL). The segmentation results were finally quantified using DSC, and is summarized in Fig. 5. It can be found that for the large tumors (i.e. (a) and (b)), ECM-MS had relatively better performance with  $\eta \in [0.01, 0.09]$ ; while, for the two small tumors (i.e. (c) and (d)),  $\eta \in [0.1, 0.9]$  is better. Thus, in our experiment,  $\lambda$  was set to 8, while  $\eta$  was set to 0.01 and 0.2, respectively, for large and small tumors.

## V. CONCLUSION

In this study, we have investigated to address the imprecision and noise inherent in FDG-PET images via Dempster-Shafer theory (DST), a powerful tool for modeling and reasoning with uncertain and/or imprecise information. Based on DST, an evidential clustering algorithm integrating adaptive distance metric and MRF-based spatial regularization has been proposed for the delineation of tumor volumes in FDG-PET images. The experimental results obtained on twenty real-patient PET stacks have shown the effectiveness of the proposed method. Considering that DST is also widely used for the information fusion task, as the future work, we will study how to include the anatomical information provided by

CT into the proposed segmentation algorithm, so as to further improve the tumor delineation performance in FDG-PET.

## REFERENCES

- [1] D. Yan *et al.*, "Adaptive radiation therapy," *Physics in Medicine and Biology*, vol. 42, no. 1, p. 123, 1997.
- [2] D. Thorwarth *et al.*, "Physical radiotherapy treatment planning based on functional PET/CT data," *Radiotherapy and Oncology*, vol. 96, no. 3, pp. 317–324, 2010.
- [3] I. El Naqa *et al.*, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," *Pattern Recognition*, vol. 42, no. 6, pp. 1162–1171, 2009.
- [4] F. Tixier *et al.*, "Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *Journal of Nuclear Medicine*, vol. 52, no. 3, pp. 369–378, 2011.
- [5] H. Mi *et al.*, "Robust feature selection to predict tumor treatment outcome," *Artificial Intelligence in Medicine*, vol. 64, pp. 195–204, 2015.
- [6] Y. E. Erdi *et al.*, "Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding," *Cancer*, vol. 80, no. S12, pp. 2505–2509, 1997.
- [7] S. Vaucelin *et al.*, "Development of a generic thresholding algorithm for the delineation of 18FDG-PET-positive tissue: application to the comparison of three thresholding models," *Physics in Medicine and Biology*, vol. 54, no. 22, p. 6901, 2009.
- [8] H. Li *et al.*, "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours," *Medical Physics*, vol. 35, no. 8, pp. 3711–3721, 2008.
- [9] W. Mu *et al.*, "A segmentation algorithm for quantitative analysis of heterogeneous tumors of the cervix with 18 F-FDG PET/CT," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2465–2479, 2015.
- [10] M. Aristophanous *et al.*, "A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography," *Medical Physics*, vol. 34, no. 11, pp. 4223–4235, 2007.
- [11] U. Bagci *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," *Medical Image Analysis*, vol. 17, no. 8, pp. 929–945, 2013.
- [12] D. Onoma *et al.*, "Segmentation of heterogeneous or small FDG PET positive tissue based on a 3D-locally adaptive random walk algorithm," *Computerized Medical Imaging and Graphics*, vol. 38, no. 8, pp. 753–763, 2014.
- [13] B. Foster *et al.*, "Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 711–724, 2014.
- [14] S. Belhassen and H. Zaidi, "A novel fuzzy c-means algorithm for unsupervised heterogeneous tumor quantification in PET," *Medical Physics*, vol. 37, no. 3, pp. 1309–1324, 2010.
- [15] B. Lelandais *et al.*, "Fusion of multi-tracer PET images for dose painting," *Medical Image Analysis*, vol. 18, no. 7, pp. 1247–1259, 2014.
- [16] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.
- [17] M. N. Ahmed *et al.*, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 21, no. 3, pp. 193–199, 2002.
- [18] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
- [19] M.-H. Masson and T. Denceux, "ECM: An evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 41, no. 4, pp. 1384–1397, 2008.
- [20] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [21] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 2, pp. 780–795, 1999.
- [22] J. Zhang *et al.*, "Local energy pattern for texture classification using self-adaptive quantization thresholds," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 31–42, 2013.
- [23] G. Thibault *et al.*, "Advanced statistical matrices for texture characterization: application to cell classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 630–637, 2014.
- [24] J. Zhang *et al.*, "Automatic craniomaxillofacial landmark digitization via segmentation-guided partially-joint regression forest model and multiscale statistical features," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1820–1829, 2016.
- [25] E. P. Xing *et al.*, "Distance metric learning with application to clustering with side-information," in *Advances in Neural Information Processing Systems*, 2002, pp. 505–512.
- [26] F. Wang and J. Sun, "PSF: A unified patient similarity evaluation framework through metric learning with weak supervision," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1053–1060, 2015.
- [27] H. Jia *et al.*, "A new distance metric for unsupervised learning of categorical data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 5, pp. 1065–1079, 2016.
- [28] A.-S. Dewalle-Vignion *et al.*, "A new method for volume segmentation of PET images, based on possibility theory," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 409–423, 2011.
- [29] Y. Zhang *et al.*, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [30] A. Roche *et al.*, "On the convergence of EM-like algorithms for image segmentation using Markov random fields," *Medical Image Analysis*, vol. 15, no. 6, pp. 830–839, 2011.
- [31] S. Le Hegarat-Mascle *et al.*, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1018–1031, 1997.
- [32] A. Bendjebbour *et al.*, "Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 8, pp. 1789–1798, 2001.
- [33] H. Tabia *et al.*, "A new 3D-matching method of nonrigid and partially similar models using curve analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 852–858, 2011.
- [34] Z.-G. Liu *et al.*, "A new belief-based k-nearest neighbor classification method," *Pattern Recognition*, vol. 46, no. 3, pp. 834–844, 2013.
- [35] —, "Credal classification rule for uncertain data based on belief functions," *Pattern Recognition*, vol. 47, no. 7, pp. 2532–2541, 2014.
- [36] L. Jiao *et al.*, "Belief rule-based classification system: Extension of FRBCS in belief functions framework," *Information Sciences*, vol. 309, pp. 26–49, 2015.
- [37] C. Lian *et al.*, "Dissimilarity metric learning in the belief function framework," *IEEE Transactions on Fuzzy Systems*, vol. PP, no. 99, 2016.
- [38] Z.-G. Liu *et al.*, "Adaptive imputation of missing values for incomplete pattern classification," *Pattern Recognition*, vol. 52, pp. 85–95, 2016.
- [39] —, "Hybrid classification system for uncertain data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016.
- [40] T. Denceux and M.-H. Masson, "EVCLUS: evidential clustering of proximity data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 1, pp. 95–109, 2004.
- [41] N. Makni *et al.*, "Introducing spatial neighbourhood in evidential c-means for segmentation of multi-source images: application to prostate multi-parametric MRI," *Information Fusion*, vol. 19, pp. 61–72, 2014.
- [42] Z.-G. Liu *et al.*, "Credal c-means clustering method based on belief functions," *Knowledge-Based Systems*, vol. 74, pp. 119–132, 2015.
- [43] K. Zhou *et al.*, "ECMdd: Evidential c-medoids clustering with multiple prototypes," *Pattern Recognition*, vol. 60, pp. 239–257, 2016.
- [44] C. Lian *et al.*, "Dempster-Shafer theory based feature selection with sparse constraint for outcome prediction in cancer therapy," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Springer, 2015, pp. 695–702.
- [45] —, "Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction," *Medical Image Analysis*, vol. 32, pp. 257–268, 2016.
- [46] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [47] A.-L. Jousselme *et al.*, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [48] R. A. Waltz *et al.*, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical Programming*, vol. 107, no. 3, pp. 391–408, 2006.
- [49] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [50] G. Gerig *et al.*, "Valmet: A new validation tool for assessing and improving 3D object segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001*. Springer, 2001, pp. 516–523.