Joint Tumor Segmentation in PET-CT Images Using Co-Clustering and Fusion Based on Belief Functions

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

Joint Tumor Segmentation in PET-CT Images using Co-Clustering and Fusion based on Belief Functions

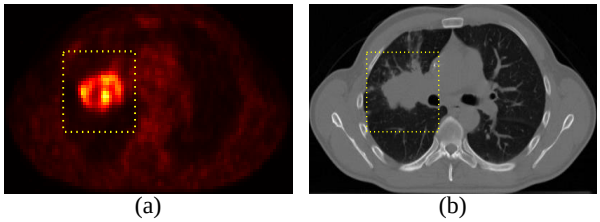
Chunfeng Lian, Su Ruan*, Thierry Denœux, Hua Li, and Pierre Vera

***Abstract*—Precise delineation of target tumor is a key factor to ensure the effectiveness of radiation therapy. While hybrid positron emission tomography-computed tomography (PET-CT) has become a standard imaging tool in the practice of radiation oncology, many existing automatic/semi-automatic methods still perform tumor segmentation on mono-modal images. In this paper, a co-clustering algorithm is proposed to concurrently segment 3-D tumors in PET-CT images, considering that the two complementary imaging modalities can combine functional and anatomical information to improve segmentation performance. The theory of belief functions is adopted in the proposed method to model, fuse, and reason with uncertain and imprecise knowledge from noisy and blurry PET-CT images. To ensure reliable segmentation for each modality, the distance metric for the quantification of clustering distortions and spatial smoothness is iteratively adapted during the clustering procedure. On the other hand, to encourage consistent segmentation between different modalities, a specific context term is proposed in the clustering objective function. Moreover, during the iterative optimization process, clustering results for the two distinct modalities are further adjusted via a belief-functions-based information fusion strategy. The proposed method has been evaluated on a dataset consisting of twenty-one paired PET-CT images for non-small cell lung cancer (NSCLC) patients. The quantitative and qualitative evaluations show that our proposed method performs well compared with the state-of-the-art methods.**

***Index Terms*—Tumor Co-Segmentation, Co-Clustering, Context Information, Information Fusion, Adaptive Distance Metric, Spatial Regularization, Belief Functions, PET-CT.**

I. Introduction

PRECISE segmentation of target tumor is indispensable in the practical process of radiation therapy. It assists in maximizing dose delivery in tumor tissues while minimizing side effects in organs at risk (OARs). Also, it is the basis for developing robust radiomic approaches for the prediction of therapy outcomes [1]–[3]. Positron emission tomography (PET), with the radio-tracer fluoro-2-deoxy-D-glucose (FDG), is a functional imaging modality that can non-invasively monitor metabolic activities of tumor cells. Target tumor usually demonstrates high standardized uptake values (SUVs) in FDG-PET images, which facilitates its differentiation from adjacent OARs (e.g. Fig. 1(a)). However, due to imprecise acquisition system and limited spatial resolution, positive tissues in PET images are blurry and noisy, which make it a challenging work for clinicians to precisely delineate tumors using solely PET images. Compared with PET, computed tomography (CT) is poor in contrast for the discrimination between tumor tissues and OARs, but it have relatively high spatial resolution, and can provide detailed anatomical information as well (e.g. Fig. 1(b)). In addition, the calculation of radiation dose depends on knowledge from CT images. The development of hybrid PET-CT scanner effectively combines the complementary information from the two distinct imaging modalities. While PET-CT has become a standard imaging technique in radiation oncology, many existing automatic/semi-automatic methods still delineate tumor volumes in mono-modal images.



(a) (b)

Fig. 1. A tumor shown in integrated PET-CT images: (a) FDG-PET image in the axial plane and (b) the corresponding CT. The yellow dotted box indicates the volume of interest (VOI).

Diverse mono-modal methods for automatic/semi-automatic segmentation of target tumor in PET images have been proposed [4], [5], which include thresholding methods [6], [7], region growing methods [8], [9], methods based on level set/active contours [10], [11], graph-based methods [12]–[14], gradient-based methods [15], [16], statistical methods [17]–[19], supervised learning methods [20], [21], and clustering methods [22]–[26]. Deep convolutional neural networks (CNN) [27] have also been applied to segmenting PET tumors [28]. It is worth noting that clustering methods have shown competitive performance in tumor delineation, especially considering that tumor tissues of different patients usually present heterogenous uptakes and irregular contours [4]. For example, an improved FCM (i.e., FCM-SW) method was proposed in [23] for the segmentation of heterogeneous PET tumors. A generalized FCM (i.e., SPEQTACLE) method was proposed in [25], where the fuzzy distributions of PET images were taken into account during tumor delineation. A spatial evidential clustering (i.e., SECM) method was proposed in [24] for the segmentation of lung tumors in multi-tracer PET images. More

* Corresponding author.

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

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

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

H. Li is with the Department of Radiation Oncology, Washington University in St. Louis, Saint Louis, MO 63110, USA (e-mail: li.hua@wustl.edu).

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

recently, a novel evidential clustering method (i.e., ECM-MS) was proposed in our previous study [26] to delineate target tumor using solely PET images, where the most informative imaging features were automatically selected to reliably quantify clustering distortions and spatial smoothness.

Some other methods use information from one modality to guide the delineation of target tumor in another modality [29]–[31]. For example, Mu et al. [31] proposed an improved level set method to segment tumor in FDG-PET images, where knowledge from corresponding CT images was adopted to guide the initialization of zero level set. In [30], Wojak et al. proposed a joint variational segmentation method, in which PET information is regarded as a local constraint to adjust the segmentation results in CT images. These methods attempted to delineate a unique contour in PET and CT images, without considering that the two distinct imaging modalities actually demonstrate complementary while distinct characteristics of the same target.

Up to now, most existing methods for tumor co-segmentation in PET-CT images are graph-based [32]–[37]. In [34], by constructing a hyper graph, a random walk method was proposed for automatic co-segmentation of multi-modal medical images (e.g. PET-CT and PET-MRI). In [33], an extension of [32], the co-segmentation was formulated as a binary labeling problem of Markov random field (MRF) on a graph consisting of two sub-graphs. The two sub-graphs correspond to PET and CT images, respectively; the interaction between them was modeled by an adaptive context energy. A maximum flow graph cut algorithm was then adopted to solve the formulated MRF optimization problem for consistent co-segmentation. In [36], the random walk and graph cut methods were effectively combined, where the random walk was performed on PET images as a stable initialization to improve the co-segmentation performance of the subsequent maximum flow graph cut with a specific energy function. In [37], a topology graph was combined with an intensity graph to improve graph cut with topology modeling for semi-automatic segmentation of target tumor in PET-CT images. Although these graph-based methods are efficient and intuitive for the co-segmentation task, their performance may be influenced by the quality of predefined tumor and background seeds. In addition, they make decision only according to the information provided by intensity values of image voxels, while other imaging features, e.g., texture features [38]–[40], describing the spatial context of each voxel are ignored.

Apart from these graph-based approaches, some other methodologies, e.g., active contours [41], supervised learning [42], [43], and stochastic modeling [44], [45], have also been applied to co-segmenting multi-modal medical images. For example, in [42], [43], textural features were extracted from PET-CT images to train classification models for the identification of tumor voxels. A statistical method based on hidden Markov tree (HMT) was proposed in [44] for tumor co-segmentation in PET-CT images pre-processed by wavelet/contourlet enhancement. A hidden Markov fields (HMF) based method was proposed in [45] for joint tumor segmentation in multi-modal medical images, where a copula-based fusion strategy was adopted to model dependency between different modalities. One critical challenge for the application of such statistical methods on PET-CT co-segmentation is how to conjointly define stochastic distributions of PET and CT voxels that could be very different. CNN-based approaches have also shown promising performance for multi-source medical image segmentation. For example, a multi-scale CNN method was proposed in [46] to segment tumor in multi-modal MRI. More recently, a multi-channel multi-scale encoder-decoder network (i.e., M^2EDN) method was proposed in [47] to learn task-oriented fusion of multi-source information for brain structure segmentation in MRI. These deep learning methods are also applicable in PET-CT tumor co-segmentation, while limited number of training samples (i.e., PET-CT images) may potentially hamper the segmentation performance. Also, similar to [29]–[31], the above active contour-based, statistical, and learning-based co-segmentation methods [41]–[47] tend to yield a unique contour in both PET and CT images, ignoring the specificity of each modality.

Effectively modeling uncertainty and imprecision inherent in PET and PET-CT images is of great concern for accurate segmentation of blurry and inhomogeneous targets [19]. In our study, this critical issue is addressed via the theory of belief functions (BFT) [48]. As a powerful tool for representing and reasoning with partial (imprecise, uncertain) knowledge, BFT can provide reliable ways to fuse information from different sources, turning out to be effective for joint tumor delineation in PET-CT images. Based on BFT, we propose a co-clustering method to concurrently segment volumetric PET and CT images in a unified framework. Diverse imaging features are extracted for the comprehensive characterization of PET and CT voxels, while without prior knowledge concerning the most informative ones. As an extension of our previous mono-modal method [26], a joint feature selection and distance metric adaptation procedure is included in the proposed co-clustering method to ensure reliable quantification of clustering distortions for both PET and CT images. In addition, a specific spatial regularization is also included to protect local homogeneity in each mono-modality. To encourage consistency between PET and CT, a special context term is proposed to softly penalize conflicting segmentation in the two distinct but also complementary modalities. During the iterative process of clustering, results in PET and CT are further adjusted by fusing them via the Dempster's combination rule [48], considering that knowledge from one modality could be regarded as additional evidence to assist the segmentation in the other one.

The rest of this paper is organized as follows. BFT is briefly recalled in Section II. The proposed co-segmentation method is then introduced in Section III. In Section IV, the proposed method is evaluated by a cohort of real-patient PET-CT images. Some discussions and the conclusion are presented in Section V and Section VI, respectively.

II. Theory of Belief Functions

The theory of belief functions (BFT) is also known as the Dempster-Shafer theory or evidence theory [48]. It has been successfully applied in diverse fields, including image analysis [49]–[52], and data classification/clustering [53]–[59], etc.

As a formal framework for modeling, fusing, and reasoning with partial (uncertain, imprecise) information, BFT extends both the set-membership approaches and probability theory, and has strong connections with other theories of uncertainty, e.g., possibility theory and imprecise probability. BFT has two main components, i.e., the quantification of pieces of evidence, and the combination of quantified evidence.

A. Evidence Quantification

In the framework of belief functions, we reason under uncertainty based on the modeling of evidence [48]. Let ω be a variable taking values in a finite domain $\Omega = \{\omega_1, \cdots, \omega_c\}$, called the *frame of discernment*. An item of uncertain evidence regarding the actual value of ω can be represented by a *mass function* m on Ω, defined from the powerset 2^Ω to the interval $[0, 1]$, such that

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

Each mass $m(A)$ denotes a *degree of support* attached to the proposition "the true value of ω is in the subset A", and to no more specific proposition. Function m is said to be normalized if $m(\emptyset) = 0$, where the empty set denotes hypothesis beyond the finite domain Ω. Any subset A with $m(A) > 0$ is called a *focal element*. In particular, if all focal elements of a mass function are singletons, it is said to be *Bayesian*, and is then equivalent to a probability distribution. A mass function with only one focal element is said to be *categorical*, and is then equivalent to a set.

We can associate *belief* and *plausibility* functions to a normalized mass function m, from 2^Ω to $[0, 1]$, which are defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \tag{2}$$

for all $A \subseteq \Omega$. Quantity $Bel(A)$ can be interpreted as the degree to which the evidence supports A (also known as *credibility*), while $Pl(A)$ can be interpreted as the degree to which the evidence is not contradictory to A. Functions Bel and Pl are in one-to-one correspondence with mass function m. For all $A \subseteq \Omega$, we always have $Bel(A) \leq Pl(A)$, which indicates that the belief (or probability) supporting the proposition "$\omega \in A$" is bounded by the corresponding credibility and plausibility.

B. Evidence Combination

A key operation in BFT is that beliefs are elaborated by fusing evidence from multiple information sources. The basic mechanism for evidence combination is *Dempster's rule* [48]. Let m_1 and m_2 be two mass functions derived from independent items of evidence. They can be fused using Dempster's rule to form a new mass function $m_{1\oplus2}$ defined as

$$m_{1\oplus2}(A) = \frac{1}{1-Q} \sum_{B \cap C = A} m_1(B) m_2(C), \tag{3}$$

for all $A \subseteq \Omega \setminus \emptyset$, where $Q = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$. It is worth noting that Dempster's rule generalizes set intersection and probabilistic conditioning, by which beliefs can be refined after information fusion.

For example, let m_1 and m_2 be two independent mass functions to help reasoning the actual value of an object. They are defined on a frame of discernment $\Omega = \{\omega_1, \omega_2\}$, which means that all possible values of the studied object are included in Ω. Thus, large value of $m_1(\Omega)$ or $m_2(\Omega)$ reflects that the corresponding piece of evidence is imprecise, since $m_1(\Omega)$ and $m_2(\Omega)$ quantify the degree of support attached to the hypothesis "the actual value of the studied object is in Ω", which does not provide any more specific indications. Assuming m_1 and m_2, and their combination via (3), i.e., $m_{1\oplus2}$, have the form of

A	$m_1(A)$	$m_2(A)$	$m_{1\oplus2}(A)$
$\{\omega_1\}$	0.5	0.2	0.56
$\{\omega_2\}$	0	0.2	0.11
Ω	0.5	0.6	0.33

According to $m_1(\Omega) = 0.5$ and $m_2(\Omega) = 0.6$, we can observe that both m_1 and m_2 are severly imprecise, while the fusion of them effectively reduces the imprecision to $m_{1\oplus2}(\Omega) = 0.33$. In addition, we can also find that the mass function $m_{1\oplus2}$ after information fusion also reduces the uncertainty in both m_1 and m_2, as the belief associated with the proposition "the actual value of the object is ω_1" has increased to 56%, i.e., $m_{1\oplus2}(\{\omega_1\}) = 0.56$.

However, to ensure the effectiveness of Dempster's combination rule, mass functions from distinct information sources should not contain high conflicts (which means large Q in (3)). To quantify the conflicts or inconsistence between different bodies of evidence, abounding dissimilarity measurements have been proposed, in which the metric d_J designed by Jousselme et al. [60] is a commonly used one. For any two mass functions m_1 and m_2, their dissimilarity $d_J(m_1, m_2)$ is quantified as

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)\mathbf{J}(m_1 - m_2)^T}, \tag{4}$$

where $\mathbf{J}$ is a positive definite matrix whose elements are Jaccard indexes, i.e., $\mathbf{J}(A, B) = |A \cap B|/|A \cup B|$, $\forall A, B \in 2^\Omega \setminus \emptyset$. The quantification (4) satisfies the requirements for a valid distance metric. In addition, it effectively accounts for the interaction between the focal elements of Ω [60], thus can truthfully reflect the conflicts between different pieces of evidence.

For instance, assume that m_1, m_2, and m_3 are three mass functions on $\Omega = \{\omega_1, \omega_2\}$, with the form of

A	$m_1(A)$	$m_2(A)$	$m_3(A)$
$\{\omega_1\}$	0.7	0.2	0.2
$\{\omega_2\}$	0	0	0.5
Ω	0.3	0.8	0.3

We can observe that m_1 is in consistence with m_2, as $m_1(\{\omega_2\}) = m_2(\{\omega_2\}) = 0$; while, on the other hand, both m_1 and m_2 contradicts with m_3, since the latter piece of evidence concentrated on ω_2 with $m_3(\{\omega_2\}) = 0.5$. If we use directly the Euclidean metric to quantify the dissimilarity between these three mass functions, $||m_1 - m_2||_2 =$

$||m_1 - m_3||_2 = 0.71$ is inappropriate, because it does not reflect the actual situation. However, if (4) is applied with

$$\mathbf{J} = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix} \tag{5}$$

to calculate the dissimilarity, the results are $d_J(m_1, m_2) = 0.35$ and $d_J(m_1, m_3) = 0.50$, which effectively reflect that m_1 is more consistent with m_2 than with m_3.

III. Method

First of all, image features are extracted from a 3-D region of interest (ROI) in each mono-modality. The ROI is a cuboid defined by users, which fully includes the target tumor, such as the yellow dotted box shown in Fig. 1. Let $\{X_i^{pt}\}_{i=1}^n$ be feature vectors in $\mathbb{R}^p$ for n voxels in PET, while $\{X_i^{ct}\}_{i=1}^n$ for the corresponding n voxels in CT. Using these extracted features, the proposed method co-segments tumor in the two distinct modalities via concurrently looking for matrices $\mathbf{M^{pt}} = \{m_i^{pt}\}_{i=1}^n$ and $\mathbf{M^{ct}} = \{m_i^{ct}\}_{i=1}^n$, where m_i^{pt} and m_i^{ct} are, respectively, the mass functions for voxel i in PET and CT. For PET (resp. CT), m_i^{pt} (resp. m_i^{ct}) quantifies the mass of belief attached to all possible propositions with respect to its cluster. We assume that all the voxels belong either to the background (i.e. hypothesis ω_1) or to the positive tissue (i.e. hypothesis ω_2), without existence of outliers. Thus, the whole frame of clusters is set as $\Omega = \{\omega_1, \omega_2\}$. The mass function for each voxel, both in PET and CT, obeys $m_i(\{\omega_1\}) + m_i(\{\omega_2\}) + m_i(\Omega) \equiv 1$. As $m_i(\Omega)$ measures the ambiguity regarding the clusters ω_1 and ω_2, blurring boundary and severe heterogeneous regions will have large mass on Ω. Finally, the crisp segmentation is obtained by making decisions based on $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$.

As a co-clustering algorithm, the proposed method searches for a pair of $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$ via minimizing a global cost function, which is constructed by independent costs in each modality, and context cost deduced by inconsistent segmentation between the two distinct but complementary modalities.

A. Cost for Segmentation in Mono-Modality

The cost for each mono-modality is an extension of a classical evidential clustering algorithm [53], with a form similar to that in our previous mono-modal method [26]. The novelty is that a specific spatial regularization and distance metric adaptation procedure are integrated for clustering image voxels. Without loss of generality, the following discussion taking the cost for PET, i.e., $\mathcal{J}_{mono}(\mathbf{M^{pt}})$, as an example:

Let cluster ω_1 (resp. ω_2) be represented by a center V_1^{pt} (resp. V_2^{pt}) in $\mathbb{R}^p$. For each nonempty subset $A_j \subseteq \Omega \setminus \emptyset$, we assume that its prototype $\bar{V}_j^{pt}$ is defined as the barycenter of the centers associated to the singletons composing A_j, i.e.,

$$\bar{V}_j^{pt} = \frac{1}{c_j} \sum_{k=1}^{2} s_{kj} V_k^{pt}, \tag{6}$$

where s_{kj} is binary, and it equals 1 if and only if $\omega_k \in A_j$; while $c_j = |A_j|$ denotes the cardinality of A_j. To learn $\mathbf{M^{pt}}$ in an unsupervised way, $\mathcal{J}_{mono}(\mathbf{M^{pt}})$ is then defined as

$$\begin{aligned} \mathcal{J}_{mono}(\mathbf{M^{pt}}) = & \sum_{i=1}^{n} \sum_{A_j} c_j^2 [m_i^{pt}(A_j)]^2 \left[d^2(X_i^{pt}, \bar{V}_j^{pt})\right] \\ & + \eta \sum_{i=1}^{n} \sum_{t \in \Phi(i)} \left[d_m^2(m_i^{pt}, m_{i,t}^{pt})\right] \left[d^2(X_i^{pt}, X_{i,t}^{pt})\right] \\ & + \lambda \mathcal{F} - \log\left(d^2(\bar{X}_{\omega_1}^{pt}, \bar{X}_{\omega_2}^{pt})\right), \end{aligned} \tag{7}$$

which consists of four terms:

1) Term 1: Similar to the original ECM, the first term of (7) denotes the restriction between clustering distortions and mass functions.

2) Term 2: The second term of (7) is a spatial regularization defined in the framework of belief functions for enhancing local homogeneity of neighboring voxels. According to the spatial prior of a PET volume, matrix $\mathbf{M^{pt}} = \{m_i^{pt}\}_{i=1}^n$ that we want to learn can be regarded as a specific random field, where each mass function m_i^{pt} is a random vector in $\mathbb{R}^3$, and its distribution depends on the mass functions of adjacent voxels. Let $\mathbf{\Phi} = \{\Phi(i)\}_{i=1}^n$ be a 3-D neighborhood system, where $\Phi(i) = \{1, \ldots, T\}$ is the set of the T neighbors for voxel i. The corresponding mass functions of voxels in $\Phi(i)$ are $\{m_{i,1}^{pt}, \ldots, m_{i,T}^{pt}\}$, while the feature vectors of these voxels are $\{X_{i,1}^{pt}, \ldots, X_{i,T}^{pt}\}$. Then, based on the above assumption about $\mathbf{M^{pt}}$, $\sum_{t \in \Phi(i)} \left[d_m^2(m_i^{pt}, m_{i,t}^{pt})\right] \left[d^2(X_i^{pt}, X_{i,t}^{pt})\right]$ of (7) quantifies the smoothness around voxel i, where $d^2(X_i^{pt}, X_{i,t}^{pt})$ denotes the distance between voxel i and its neighbor t in the feature space; while, $d_m^2(m_i^{pt}, m_{i,t}^{pt})$ measures the independence between m_i^{pt} and $m_{i,t}^{pt}$, i.e., *the inconsistency between the mass functions of voxel i and its neighbor t*. Based on (4), we set $d_m^2(m_i^{pt}, m_{i,t}^{pt}) = (m_i^{pt} - m_{i,t}^{pt})\mathbf{J}(m_i^{pt} - m_{i,t}^{pt})^T$, where matrix $\mathbf{J}$ is defined by (5).

3) Term 3: To describe image voxels, we include textural features as complementary information for voxel intensities. The challenge to this end is that a large amount of textural features can be extracted, but without prior knowledge concerning the most informative features. In addition, the quantified high-dimensional feature vectors may contain unreliable variables due to noise and limited resolution of PET images. To tackle this challenge, distance metric adaptation and/or feature selection (e.g., [61]) are desirable.

In our previous work [62], 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. Distinct from this previous work, here our goal is to adapt distance metric for given data in an unsupervised learning protocol, so as to improve the performance of clustering algorithms. Therefore, we look for a matrix $\mathbf{D^{pt}} \in \mathbb{R}^{p \times q}$ during clustering, under the constraint $q \ll p$, by which the dissimilarity between any two feature vectors, say X_1^{pt} and X_2^{pt}, can be quantified as

$$d^2(X_1^{pt}, X_2^{pt}) = (X_1^{pt} - X_2^{pt})\mathbf{D^{pt}}(\mathbf{D^{pt}})^T(X_1^{pt} - X_2^{pt})^T. \tag{8}$$

A desired matrix $\mathbf{D^{pt}}$ transforms the original feature space to a low-dimensional subspace, where discriminant input fea-

tures play a more significant role than uninformative ones in calculating the dissimilarity. To find such a $\mathbf{D^{pt}}$, the distortion between any X_i^{pt} and cluster centroid $\bar{V}_j^{pt}$, i.e., $d^2(X_i^{pt}, \bar{V}_j^{pt})$ in the first term of (7), is represented by (8). Moreover, the spatial regularization is also used to adapt the distance metric, where $d^2(X_i^{pt}, X_{i,t}^{pt})$ is also quantified by (8). During the iterative minimization of (7), a large dissimilarity $d_m^2(m_i^{pt}, m_{i,t}^{pt})$ between m_i^{pt} and $m_{i,t}^{pt}$ will indicate that $\mathbf{D^{pt}}$ obtained as current step is unfit. Then, it should be adjusted at the next step to reduce the dissimilarity between X_i^{pt} and $X_{i,t}^{pt}$, so as to bring adjacent voxels closer.

To ensure the effectiveness of metric updating, the third term of (7) is defined as the sparsity regularization $||\mathbf{D^{pt}}||_{2,1}$ of matrix $\mathbf{D^{pt}}$, i.e., $\mathcal{F} = \sum_{i=1}^{p} \sqrt{\sum_{j=1}^{q} (D_{i,j}^{pt})^2}$. It aims at selecting the most reliable input features to calculate distance $d^2(\cdot,\cdot)$ in the feature space (which exists in the other three terms). By forcing rows of $\mathbf{D^{pt}}$ to be zero, the proposed method only selects reliable features for clustering, while the influence of unreliable features is controlled. Scalar λ is a tuning parameter that controls the influence of this regularization.

4) Term 4: The last term of (7) is used to prevent the cost function being trivially solved with $\mathbf{D^{pt}} = 0$, which collapses all the features vectors into a single point. Vectors $\bar{X}_{\omega_1}^{pt}$ and $\bar{X}_{\omega_2}^{pt}$ are two seeds for the positive tissue and the background, respectively, which can be predetermined automatically. We empirically set a constant parameter (i.e., $= 1$) for this term, mainly considering it serves to prevent $\mathbf{D^{pt}} = 0$, in which case the penalty will be activated and the corresponding loss will get close to infinity. In addition, we experimentally found that this term is insensitive to changing parameters.

B. Cost for Inconsistency between PET and CT

The proposed method encourages consistent segmentations for PET and CT images, while not force them to be identical, as the information provided by the two distinct modalities may be partially contradicts with each other. These partial conflicts are possibly caused by multiple reasons, e.g., the low-contrast property of CT, the blurring nature of PET, and the impropriate correspondence between PET and CT voxels that are brought by the uncertainties in registration of these two distinct modalities.

According to the above assumption, the disagreement between PET and CT segmentations can be softly modeled by the dissimilarity between $\mathbf{M^{pt}} = \{m_i^{pt}\}_{i=1}^n$ and $\mathbf{M^{ct}} = \{m_i^{ct}\}_{i=1}^n$. Thus, based on the metric introduced in (4), the context penalty, i.e., $\mathcal{J}_{joint}(\mathbf{M^{pt}}, \mathbf{M^{ct}})$, is defined as

$$\mathcal{J}_{joint}(\mathbf{M^{pt}}, \mathbf{M^{ct}}) = \sum_{i=1}^{n} (m_i^{pt} - m_i^{ct}) \mathbf{J} (m_i^{pt} - m_i^{ct})^T, \quad (9)$$

where m_i^{pt} and m_i^{ct} are the mass functions of two corresponding voxels in PET and CT images, respectively; while, $\mathbf{J}$ is the matrix introduced in (5).

To sum up, according to the discussions in Section III-A and Section III-B, the proposed co-clustering method looks for a pair of $\mathbf{M^{pt}} = \{m_i^{pt}\}_{i=1}^n$ and $\mathbf{M^{ct}} = \{m_i^{ct}\}_{i=1}^n$ via minimizing a global cost function:

$$\begin{aligned} \mathcal{J}(\mathbf{M^{pt}}, \mathbf{M^{ct}}) =& \mathcal{J}_{mono}(\mathbf{M^{pt}}) + \mathcal{J}_{mono}(\mathbf{M^{ct}}) \\ &+ \gamma \mathcal{J}_{joint}(\mathbf{M^{pt}}, \mathbf{M^{ct}}), \end{aligned} \quad (10)$$

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

$$\sum_{A_j} m_{ij}^{pt} = 1, \text{ and } \sum_{A_j} m_{ij}^{ct} = 1, \quad (11)$$

for all $i = 1, \ldots, n$, and $A_j \subseteq \Omega \setminus \emptyset$. In (10), $\mathcal{J}_{mono}(\mathbf{M^{pt}})$ and $\mathcal{J}_{mono}(\mathbf{M^{ct}})$ denote, respectively, the cost in PET and CT that quantified by (7); while, $\mathcal{J}_{joint}(\mathbf{M^{pt}}, \mathbf{M^{ct}})$ is the context term (9) to quantify the inconsistence between segmentation in PET and CT; parameter γ controls the influence of this inconsistence.

Algorithm 1: Iterative minimization of the cost

Input feature vectors $\{X_i^{pt}\}_{i=1}^n$ and $\{X_i^{ct}\}_{i=1}^n$; spatial neighborhood $\Phi(i)$ of each voxel i; tuning parameters η, λ, and γ; initial $\mathbf{M^{pt}}_{(0)}$, $\mathbf{M^{ct}}_{(0)}$, $\mathbf{V^{pt}}_{(0)}$, $\mathbf{V^{ct}}_{(0)}$, $\mathbf{D^{pt}}_{(0)}$, and $\mathbf{D^{ct}}_{(0)}$; tumor and background seeds ;

for $l = 1, 2, \ldots, l_{\max}$ **do**

Step 1. Optimization in PET:

- E-step: calculate $\mathbf{M^{pt}}_{(l)}$ using the efficient interior-point algorithm [63] with (12), $\mathbf{M^{ct}}_{(l-1)}$, $\mathbf{M^{pt}}_{(l-1)}$, $\mathbf{V^{pt}}_{(l-1)}$, and $\mathbf{D^{pt}}_{(l-1)}$;
- M-step I: calculate $\mathbf{V^{pt}}_{(l)}$ using (14) and $\mathbf{M^{pt}}_{(l)}$;
- M-step II: calculate $\mathbf{D^{pt}}_{(l)}$ via the Beck-Teboulle proximal gradient algorithm [64] with (15), $\mathbf{M^{pt}}_{(l)}$, $\mathbf{V^{pt}}_{(l)}$, and $\mathbf{D^{pt}}_{(l-1)}$;

Step 2. Optimization in CT:

- E-step: calculate $\mathbf{M^{ct}}_{(l)}$ using the efficient interior-point algorithm [63] with (12), $\mathbf{M^{pt}}_{(l)}$, $\mathbf{M^{ct}}_{(l-1)}$, $\mathbf{V^{ct}}_{(l-1)}$, and $\mathbf{D^{ct}}_{(l-1)}$;
- M-step I: calculate $\mathbf{V^{ct}}_{(l)}$ using (14) and $\mathbf{M^{ct}}_{(l)}$;
- M-step II: calculate $\mathbf{D^{ct}}_{(l)}$ via the Beck-Teboulle proximal gradient algorithm [64] with (15), $\mathbf{M^{ct}}_{(l)}$, $\mathbf{V^{ct}}_{(l)}$, and $\mathbf{D^{ct}}_{(l-1)}$;

Step 3. Adjust $\mathbf{M^{pt}}_{(l)}$ *by fusion with* $\mathbf{M^{ct}}_{(l)}$ via (3), with more information presented in Section III-C4;

if *no significant change of (10)* **then**

break;

end

end

Output a desired pair of $\mathbf{M^{pt}}_*$ and $\mathbf{M^{ct}}_*$;

C. Iterative Minimization of the Cost

To find the best pair of $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$, we propose to minimize the global cost function (10) in an iterative scheme. The optimization procedure is summarized in Algorithm 1, which can be detailed as follows.

1) Initialization: We initialize the mass functions (i.e. $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$) and the singleton cluster centers (i.e. $\mathbf{V^{pt}}$ and $\mathbf{V^{ct}}$) for both PET and CT via the original ECM algorithm. Based on the initial mass functions that obtained in PET, a very

limited number of tumor and background seeds are predefined. After that, the two feature vectors, i.e., $\bar{X}^{pt}_{\omega_1}$ (resp. $\bar{X}^{ct}_{\omega_1}$) and $\bar{X}^{pt}_{\omega_2}$ (resp. $\bar{X}^{ct}_{\omega_2}$), used in the last term of (7) are calculated as the barycenters of the tumor and background seeds, respectively. The output dimension, namely the number of columns in feature transformation matrix $\mathbf{D^{pt}}$ (resp. $\mathbf{D^{ct}}$), is then determined by applying principle component analysis on all the feature vectors $\{X^{pt}_i\}^n_{i=1}$ (resp. $\{X^{ct}_i\}^n_{i=1}$). The initial $\mathbf{D^{pt}}$ (resp. $\mathbf{D^{ct}}$) is constructed by the top 95% eigenvectors.

The optimization procedure alternates between three parts: the clustering in PET, the clustering in CT, and the fusion of $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$ at current step. For the clustering in each mono-modality, the optimization iterates between cluster assignment (i.e. $\mathbf{M^{pt}}$ or $\mathbf{M^{ct}}$ estimation) in the E-step, and both prototype determination (i.e. $\mathbf{V^{pt}}$ or $\mathbf{V^{ct}}$ estimation) and distance metric adaptation (i.e. $\mathbf{D^{pt}}$ or $\mathbf{D^{ct}}$ estimation) in the M-step.

2) Optimization in PET: This procedure only relates to the minimization of the first term and the last term of (10), which is performed in an EM-like protocol.

a) E-Step: Given $\mathbf{V^{pt}}$, $\mathbf{D^{pt}}$, and $\mathbf{M^{ct}}$, the minimization of (10) turns to be a quadratic problem with respect to $\mathbf{M^{pt}} = \{m^{pt}_i\}^n_{i=1}$. The derivative of (10) concerning the mass function m^{pt}_i ($\in \mathbb{R}^3$), $\forall i = 1, \ldots, n$, can be written as

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial m^{pt}_i} =& 2m^{pt}_i\mathbf{B} + 2\eta \sum_{t\in\Phi(i)} \left[d^2(X^{pt}_i, X^{pt}_{i,t})\right](m^{pt}_i - m^{pt}_{i,t})\mathbf{J} \\ &+ 2\gamma(m^{pt}_i - m^{ct}_i)\mathbf{J},\end{aligned} \tag{12}$$

where the matrix $\mathbf{J}$ is defined by (5), $d^2(X^{pt}_i, X^{pt}_{i,t})$ is measured by (8), m^{ct}_i is the mass function for the corresponding ith voxel in CT, and

$$\mathbf{B} = \begin{pmatrix} c^2_1 d^2(X^{pt}_i, \bar{V}^{pt}_1) & 0 & 0 \\ 0 & c^2_2 d^2(X^{pt}_i, \bar{V}^{pt}_2) & 0 \\ 0 & 0 & c^2_3 d^2(X^{pt}_i, \bar{V}^{pt}_3) \end{pmatrix}, \tag{13}$$

where $\bar{V}^{pt}_j$ is determined by (6), $\forall A_j \in \{\{\omega_1\}, \{\omega_2\}, \Omega\}$, and $d^2(X^{pt}_i, \bar{V}^{pt}_j)$ is calculated by (8). Based on the derivation (12), and using $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$ obtained by the last step as initializations, an efficient interior-point algorithm with a limited-memory BFGS approximation of the Hessian matrix [63] is adopted to solve the quadratic problem, so as to obtain a desired matrix $\mathbf{M^{pt}}$ at current step.

b) M-step I: Updating matrix $\mathbf{V^{pt}}$ of clustering centers is only influenced by the first term of (10). Let $f_j = \sum^n_{i=1} c^2_j (m^{pt}_{ij})^2$ and $g_j = \sum^n_{i=1} c^2_j (m^{pt}_{ij})^2 X^{pt}_i$, $\forall A_j \in \{\{\omega_1\}, \{\omega_2\}, \Omega\}$. According to (6), the centers of clusters ω_1 and ω_2 are calculated, respectively and directly, as

$$\begin{cases} V^{pt}_1 = \frac{2f_2(2g_1+g_3)+f_3(g_1-g_2)}{4f_1f_2+f_3(f_1+f_2)}; \\ V^{pt}_2 = \frac{2f_1(2g_2+g_3)+f_3(g_2-g_1)}{4f_1f_2+f_3(f_1+f_2)}. \end{cases} \tag{14}$$

c) M-step II: Similar to M-step I, the optimization of $\mathbf{D^{pt}}$ relates to the first term of (10), i.e., $\mathcal{J}^{ms}_{ecm}(\mathbf{M^{pt}})$ quantified by (7). The objective function (7) is only partly differentiable as a function of $\mathbf{D^{pt}}$, where the sparsity regularization $||\mathbf{D^{pt}}||_{2,1}$ has a singularity at $\mathbf{D^{pt}} = 0$. The derivative of the differentiable part of (7) concerning $\mathbf{D^{pt}}$ can be written as

$$\begin{aligned}\frac{\partial(\cdot)}{\partial \mathbf{D^{pt}}} =& 2\sum^n_{i=1}\sum_{A_j} c^2_j [m^{pt}_i(A_j)]^2 (X^{pt}_i - \bar{V}^{pt}_j)^T (X^{pt}_i - \bar{V}^{pt}_j)\mathbf{D^{pt}} \\ &+ 2\eta \sum^n_{i=1}\sum^{\Phi(i)}_t \left[d^2_m(m^{pt}_i, m^{pt}_{i,t})\right](X^{pt}_i - X^{pt}_{i,t})^T (X^{pt}_i - X^{pt}_{i,t})\mathbf{D^{pt}} \\ &- \frac{2(\bar{X}^{pt}_{\omega_1} - \bar{X}^{pt}_{\omega_2})^T (\bar{X}^{pt}_{\omega_1} - \bar{X}^{pt}_{\omega_2})\mathbf{D^{pt}}}{(\bar{X}^{pt}_{\omega_1} - \bar{X}^{pt}_{\omega_2})\mathbf{D^{pt}}(\mathbf{D^{pt}})^T (\bar{X}^{pt}_{\omega_1} - \bar{X}^{pt}_{\omega_2})^T},\end{aligned} \tag{15}$$

based on which the Beck-Teboulle proximal gradient algorithm [64] is adopted to search for a qualified $\mathbf{D^{pt}}$ at current step, using $\mathbf{D^{pt}}$ obtained at the last iteration as the initialization.

3) Optimization in CT: The adaptation of $\mathbf{M^{ct}}$, $\mathbf{V^{ct}}$, and $\mathbf{D^{ct}}$, which relates to the last two terms of (10), is following the same way as that for $\mathbf{M^{pt}}$, $\mathbf{V^{pt}}$, and $\mathbf{D^{pt}}$ discussed above. To update $\mathbf{M^{ct}}$ at current step, $\mathbf{M^{pt}}$ obtained by Section III-C2 is utilized in (12).

4) Fusion of PET and CT: It is worth noting that $\mathbf{M^{pt}}$ and $\mathbf{M^{ct}}$ obtained in Section III-C2 and III-C3 are in fact two independent pieces of evidence regarding the same target tumor. They are complementary, as PET and CT can provide, respectively, functional and anatomical information. In addition, they are also consistent due to the soft context term defined by (9). Therefore, in this step, $\mathbf{M^{pt}}$ is further updated by fusing it with $\mathbf{M^{ct}}$ via the Dempster's rule (i.e. (3)). The updated $\mathbf{M^{pt}}$ is then used as the initialization for the optimization (i.e. Section III-C2) in the next iteration.

The whole optimization procedure will not terminate the alternation between the steps described in Section III-C2, III-C3, and III-C4, until the value of (10) has no significant change between two consecutive iterations.

D. Decision Making

Mass functions obtained by Algorithm 1 can be further refined. Without loss of generality, for an arbitrary voxel i, we assume that $\{m_{i,1}, \ldots, m_{i,T}\}$ are mass functions obtained by the clustering algorithm that correspond to voxels in its neighborhood $\Phi(i)$, i.e., adjacent voxels surrounding i. These mass functions can be regarded as T independent pieces of evidence to help reasoning the actual cluster of i. The reliability of evidence $m_{i,t}$ to voxel i, $\forall t \in \{1, \ldots, T\}$, is inversely proportional to the spatial distance between voxel i and t. Let this spatial distance be s^2_{it}, and based on Dempster's discounting procedure [48], $m_{i,t}$ can be weighted by a coefficient $\mu_t = \exp(-s^2_{it})$ to obtain a discounted mass function:

$$\begin{cases} wm_{i,t}(\{\omega_j\}) & = \mu_t m_{i,t}(\{\omega_j\}), \quad \forall j = 1, 2, \\ wm_{i,t}(\Omega) & = 1 - \sum^2_{j=1} wm_{i,t}(\{\omega_j\}). \end{cases} \tag{16}$$

Then, using the Dempster's rule of combination (3), all the T discounted mass functions are fused with the mass function m_i to output a renewed mass function m_i. This procedure is applied to both PET and CT. It can also be regarded as a BFT-based filtering operation in a small cubic window.

Finally, based on refined mass functions, crisp clustering of PET and CT voxels is performed via maximizing the plausibility values calculated by (2).

IV. Materials and Experiments

In this section, our proposed co-clustering algorithm was evaluated on 21 volumetric FDG-PET/CT images acquired for non-small cell lung cancer (NSCLC) patients. First, our proposed co-clustering method was compared with its mono-modal variant (i.e., ECM-MS [26]). Then, it was further compared with other sate-of-the-art segmentation methods, including four mono-modal methods, i.e., 3D-LARW [12], FCM-SW [23], SECM [24], and ECM [53], and two co-segmentation methods, i.e., HMT [44] and M^{2}EDN [47].

A. Material and Features

Totally 21 FDG-PET-CT images of NSCLC patients were studied, which were acquired on a Biograph LSO Sensation 16 (Siemens Medical Solutions, Hoffman Estates, Knoxville, TN, USA). The PET images were smoothed with a Gaussian filter (full width at half maximum = 5 mm) after reconstruction. They had the same anisotropic resolution of $4.06 \times 4.06 \times 2$ mm^3, and the size of the axial PET slices was 168×168. The spatial resolution of the corresponding CT images was $0.98 \times 0.98 \times 3$ mm^3, and the size of the axial CT slices was 512×512. The tumor lesions were manually delineated by experienced clinicians in PET by the guidance of the corresponding CT. These tumor lesions were mostly located in or near the mediastinum, and some of them were located in the apex, without significant respiratory displacements at visual inspection of the gated images. Generally, these tumor lesions demonstrated heterogenous FDG uptake, and had various sizes ranging from 1.9 mL to 135.8 mL.

Considering that the image resolution is anisotropic, in our experiments, a 3×3 window in 2-D was defined to extract features both in PET and CT images; moreover, for simplicity, the same kinds of features were extracted in both of them. Using the predefined box, the average intensity value, the maximum intensity, the minimum intensity, the range of intensity value (i.e., maximum$-$minimum), and the standard deviation of intenstiy were calculated as features for the centering voxel. The gray level size zone matrix (GLSZM) [40] was adopted to extract seven texture features, as its effectiveness in medical image characterization has already been evaluated [1]. Similarly, the gray-level co-occurrence matrix (GLCM) [38] was also utilized to extract fifteen features. To sum up, for each PET and CT voxel, a 28-dimensional feature vector was extracted, consisting of 6 intensity-based, 7 GLSZM-based, and 15 GLCM-based features.

In consideration of computational costs, after extracting features in PET and CT independently, data in CT were down-sampled to PET. Each voxel (and its feature vector) in CT corresponds to one voxel in PET.

B. Evaluation Criteria

The manual delineation by experienced clinicians was performed on PET images by the guidance of the corresponding CT images. Regarding the manual delineation as the reference, all the segmentation methods were comprehensively evaluated by five criteria, including Dice coefficient (DSC), sensitivity (SEN), positive predictive value (PPV), Hausdorff distance (HD), and mean absolute surface distance (MSD). Let TP, FP, TN, and FN denote the true positive, false positive, true negative, and false negative segmentation, respectively. We define $\mathrm{DSC} = 2\mathrm{TP}/\left(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}\right)$, $\mathrm{SEN} = \mathrm{TP}/\left(\mathrm{TP} + \mathrm{FN}\right)$, and $\mathrm{PPV} = \mathrm{TP}/\left(\mathrm{TP} + \mathrm{FP}\right)$. The HD metric quantifies the maximum distance between the boundary points of two different segmentations. Complementarily, the MSD metric measures the average difference between two segmentation surfaces.

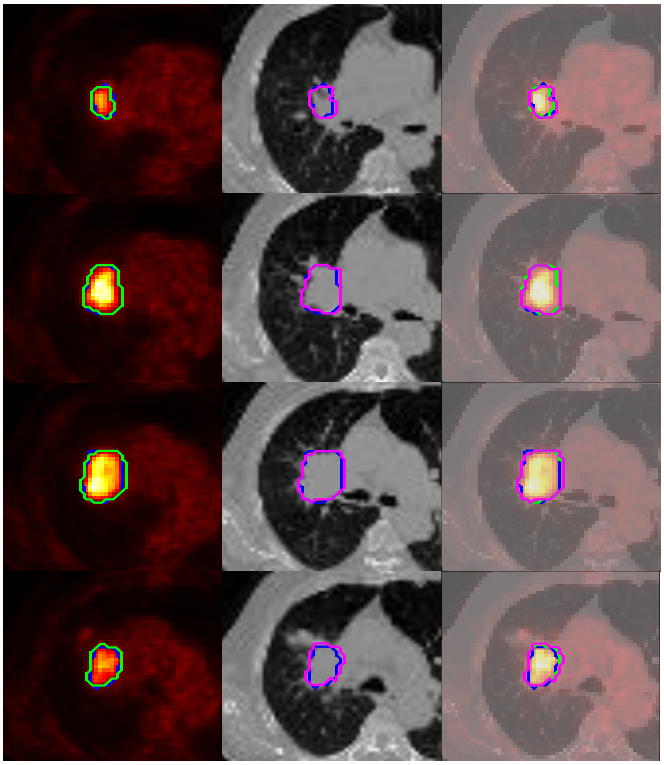


Fig. 2. A co-segmentation example shown in the axial plane, where contours delineated in PET (green) and CT (magenta) are compared to the ground truth (blue) in the first and the second column, respectively; in the last column, all the contours are overlaid in the fused images.

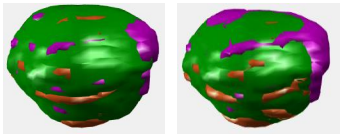


Fig. 3. Tumor volumes segmentated in PET (first column) and CT (second column), where, as compared to the ground truth, where, the green region consists of the true positive and true negative voxels, the magenta region consists of the false positive voxels, while the orange region consists of the false negative voxels.

C. Parameter Setting

The three tuning parameters utilized in the proposed method, i.e., η, λ, and γ, control, respectively, the influence of the spatial regularization, the influence of the sparsity regularization, and the consistence of the segmentation in PET and CT. In our experiments, λ and γ were empirically set to 8 and 0.001, respectively; while, to maximize the segmentation performance, the influence of the spatial penalty should be determined by taking into account the size of tumors. More specifically, η was empirically set to 0.003 and 0.2 for large and small tumors, respectively. To this end, we first roughly

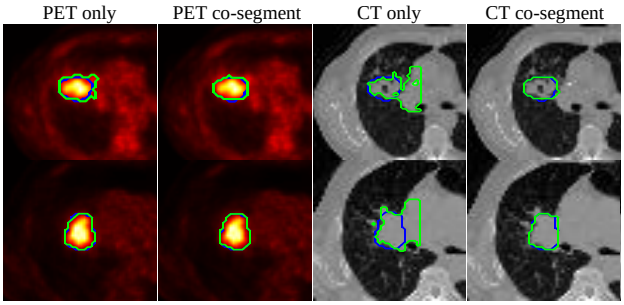


Fig. 4. Comparing mono-modal segmentation to co-segmentation in PET. The two rows correspond to two different examples. The green and blue contours denote, respectively, the automatic segmentation and the reference.

TABLE I
QUANTITATIVE SEGMENTATION RESULTS (MEAN±STANDARD DEVIATION) OBTAINED BY THE PROPOSED METHOD BASED ON CO-SEGMENTATION AND MONO-MODAL SEGMENTATION, RESPECTIVELY.

	PET only	Co-segment PET	CT only	Co-segment CT
DSC	0.86 ± 0.05	$\mathbf{0.87 \pm 0.04}$	0.29 ± 0.17	$\mathbf{0.86 \pm 0.04}$
SEN	0.84 ± 0.08	$\mathbf{0.86 \pm 0.06}$	0.73 ± 0.29	$\mathbf{0.86 \pm 0.06}$
PPV	$\mathbf{0.88 \pm 0.08}$	$\mathbf{0.88 \pm 0.06}$	0.20 ± 0.14	$\mathbf{0.86 \pm 0.06}$
HD	2.60 ± 1.34	$\mathbf{2.42 \pm 1.03}$	11.47 ± 2.20	$\mathbf{2.51 \pm 1.47}$
MSD	0.22 ± 0.10	$\mathbf{0.19 \pm 0.08}$	3.63 ± 1.14	$\mathbf{0.22 \pm 0.12}$

pre-estimated the size of the to-be-segmented tumor based on the initialization by the original ECM. Then, we empirically regarded tumors with the rough volume less than 40 mL as relatively small ones. The influence of these parameters will be further analyzed in the discussion part (i.e. Section V).

D. Results

1) Illustrative results of co-segmentation: An example to illustrate the co-segmentation performance of the proposed method is shown in Fig. 2, where each row represents a different slice in the axial plane of the same tumor. In the first and second column, contours delineated in PET (green) and CT (magenta), are compared respectively to the ground truth (blue); while, in the last column, all of them are overlaid in the fused images. As can be seen, segmentation in PET is in consistent with that in CT, while they are not identical. The 3-D tumor volumes in PET and CT are further shown in Fig. 3, where the green, magenta, and orange regions represent, respectively, true positive voxels, false positive voxels, and false negative voxels, as compared to the ground truth.

2) Co-segmentation vs. mono-modal segmentation: To demonstrate the effectiveness of the proposed co-segmentation method, its performance was compared with that of mono-modal segmentation in PET and CT. In our experiments, mono-modal segmentation was performed by ECM-MS [26], i.e., by setting γ in (10) to zero, and removing simultaneously the fusion procedure described in Section III-C4.

Two illustrative results are shown in Fig. 4, from which we can find that co-segmentation outperformed independent segmentation in PET and CT in both two cases. Correspondingly, the quantitative results on the 21 subjects are summarized in Table I, from which we can find that co-segmentation led to the best performance in terms of all the five metrics, compared to independent segmentation produced by our previous mono-modal method (i.e., ECM-MS [26]). More specifically, it is worth noting that our previous ECM-MS obtained good segmentation performance in segmenting PET images, which thanks to two main reasons: 1) PET images are high in contrast; 2) blur and heterogeneity inherent in them were effectively handled in the framework of belief functions, with the help of the proposed spatial regularization and distance metric adaptation procedure. The extension of ECM-MS, i.e., the proposed co-clustering algorithm, further improved the segmentation in PET images, as anatomical information from CT images was included as the complementary knowledge for more accurate tumor delineation. On the other hand, the segmentation in CT images was greatly improved by the co-segmentation strategy. It is mainly because the boundary between target and background is invisible in CT images, which was effectively tackled by co-segmentation via incorporating functional information from PET images.

3) Comparison with the state-of-the-art methods: In this group of experiments, our proposed co-clustering method was compared with four mono-modal methods: 1) 3D-LARW [12], 2) the original ECM [53], 3) SECM [24], and 4) FCM-SW [23]. In addition, it was also compared with two co-segmentation methods: 5) a statistical method using hidden Markov tree model (denoted as HMT) [44], and 6) M^{2}EDN [47], a CNN-based method implemented via using PET and CT as multi-channel inputs. For the supervised M^{2}EDN method, the segmentation results were quantified by 10-fold cross validation. To train a reliable network in each iteration, the training data was augmented on-the-fly via random rotations and small distortions.

The quantitative results of all competing methods are summarized in Table II, from which we can have at least two observations. *First*, the co-segmentation methods (i.e., HMT, M^{2}EDN, and our proposed method) yield overall better performance than the mono-modal methods (i.e., 3D-LARW, ECM, SECM, and FCM-SW), which demonstrates that the combination of anatomical CT information and functional PET information is beneficial for more accurate tumor segmentation. *Second*, our proposed co-clustering method outperforms the other two co-segmentation methods (i.e., HMT and M^{2}EDN). This is mainly due to: 1) Capitalizing on the specific context term (i.e., (9)) and fusion strategy defined in the framework of BFT, our proposed method provides an effective way to improve the delineation of target tumor. 2) The performance of HMT and M^{2}EDN may be hampered, respectively, by the difficulty in conjointly defining the distributions of PET and CT voxels, and the limited number of training samples for the construction of a deep neural network.

To be more comprehensive, the visual examples obtained by all competing methods are also presented in Fig. 5 for comparison. The first column of Fig. 5 presents the axial slices of four different tumors, where the first row is a slice corresponds to a large tumor, the second row represents a relatively small tumor, the third row represents a smaller tumor, and the last row represents a heterogenous tumor. The second column to the last column of Fig. 5 compare the contours delineated by different methods (green line) with those delineated by clinicians (blue line). We can observe that the contours delineated by the proposed method (the last

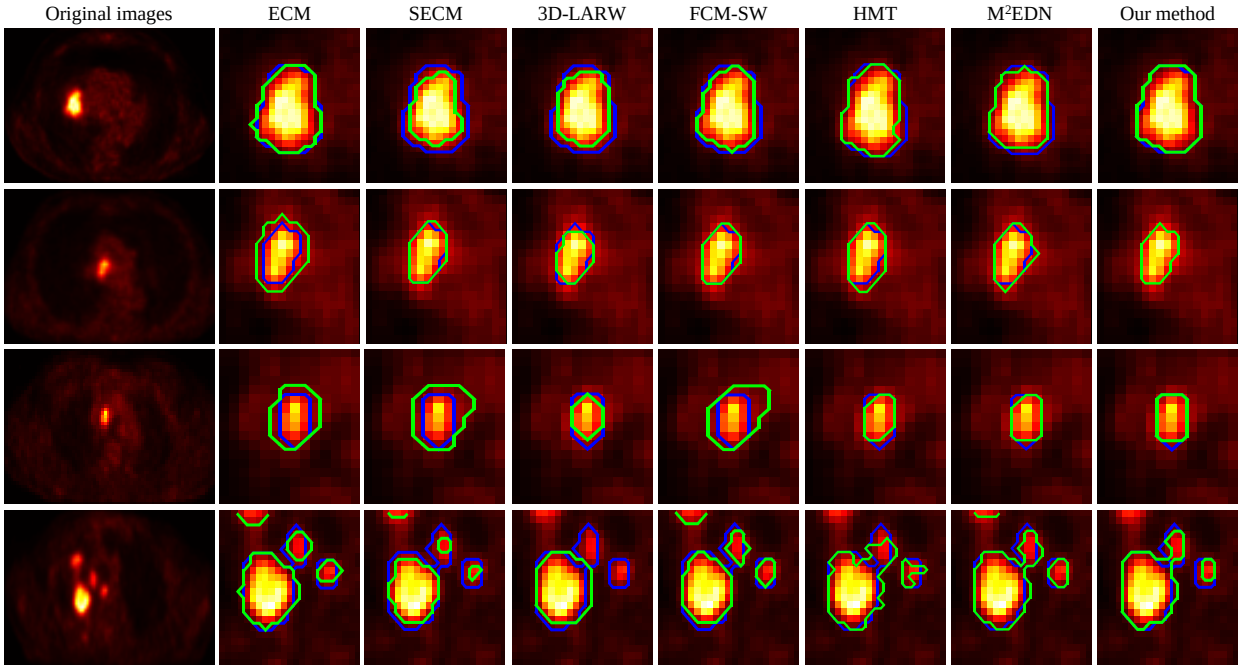


Fig. 5. Contours delineated by different methods (from the second column to the last column) for four different tumor volumes shown in the axial plane. The first column represents the input images. The contours delineated by the automatic/semi-automatic methods (green lines) are compared with those by clinicians (blue lines) in the following columns.

TABLE II
QUANTITATIVE RESULTS OBTAINED BY DIFFERENT SEGMENTATION METHODS ON ALL THE 21 SETS OF PET-CT IMAGES. THE DSC, SEN, PPV, HD (IN VOXEL), AND MSD (IN VOXEL) ARE PRESENTED AS MEAN±STANDARD DEVIATION.

	3D-LARW	ECM	SECM	FCM-SW	HMT	M^{2}EDN	Our method
DSC	0.80 ± 0.08	0.72 ± 0.14	0.77 ± 0.13	0.83 ± 0.09	0.84 ± 0.04	0.84 ± 0.07	$\mathbf{0.87 \pm 0.04}$
SEN	0.76 ± 0.15	$\mathbf{0.92 \pm 0.07}$	0.84 ± 0.15	0.83 ± 0.11	0.83 ± 0.11	0.88 ± 0.11	0.86 ± 0.06
PPV	$\mathbf{0.88 \pm 0.09}$	0.63 ± 0.21	0.78 ± 0.23	0.88 ± 0.15	0.87 ± 0.11	0.83 ± 0.14	$\mathbf{0.88 \pm 0.06}$
HD	5.48 ± 5.48	5.96 ± 4.76	4.99 ± 3.78	3.98 ± 3.42	2.78 ± 1.00	3.58 ± 2.65	$\mathbf{2.42 \pm 1.03}$
MSD	0.55 ± 0.53	0.71 ± 0.69	0.54 ± 0.35	0.36 ± 0.29	0.26 ± 0.07	0.33 ± 0.19	$\mathbf{0.19 \pm 0.08}$

column) are more consistent with the reference contours.

Furthermore, it is worth mentioning that the segmentation results presented in this paper are also comparable to previous PET-CT co-segmentation studies. For example, our method yields DSC $= 0.87 \pm 0.04$, which is comparable to Song et al. [33] (DSC $= 0.81 \pm 0.08$), Ju et al. [36] (DSC $= 0.84 \pm 0.06$), Cui et al. [37] (DSC $= 0.87 \pm 0.04$), and Salah et al. [44] (DSC $= 0.89 \pm 0.13$). It potentially implies that our proposed co-clustering method is a competitive alternative compared with existing co-segmentation methods.

V. DISCUSSION

1) Effectiveness of textural features: Our proposed method adopts textural features in addition to intensity-based features for the characterization of image voxels, and then in an unsupervised way, informative features are automatically selected to further determine a discriminant feature transformation for the co-clustering of PET and CT voxels. To evaluate its effectiveness, we alternatively removed textural features extracted based on GLSZM and GLCM, and compared the corresponding segmentation results with those obtained by using all extracted features. The quantitative comparison in terms of DSC, SEN, and PPV is shown in Fig. 6. We can observe that including textural features could effectively improve the performance of our method, e.g., DSC and PPV

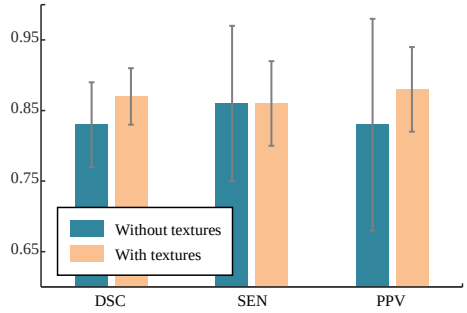


Fig. 6. The quantitative segmentation results (in terms of DSC, SEN, and PPV) obtained by our proposed method with and without using textural features, respectively.

were increased, respectively, from 0.83 ± 0.06 to 0.87 ± 0.04 and 0.83 ± 0.15 to 0.88 ± 0.06 in this group of experiments.

2) Effectiveness of information fusion based on Dempster's rule: To further refine the clustering results, the Dempster's combination rule (3) was adopted in our method for the fusion of complementary knowledge from PET and CT images (i.e. Section III-C4), as well as the fusion of information from neighboring voxels (i.e. Section III-D). To evaluate the effectiveness of such fusion strategy, we excluded it from the proposed method. The corresponding segmentation results are then summarized in Table III, from which we can find that the

TABLE III
SEGMENTATION PERFORMANCE OF THE PROPOSED METHOD, DSC AND HD (IN VOXEL), WITH/WITHOUT THE FUSION PROCEDURE BASED ON DEMPSTER'S RULE.

	without fusion	with fusion
DSC	0.86 ± 0.04	$\mathbf{0.87 \pm 0.04}$
HD	2.62 ± 1.30	$\mathbf{2.42 \pm 1.03}$

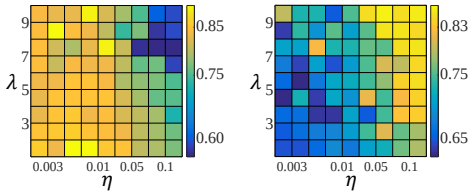


Fig. 7. The DSC, namely the intensity value, as a function of λ and η. The first and the second column correspond to two tumors with the size of 135.80 mL and 7.60 mL, respectively.

fusion procedure could help to refine the final segmentations.

3) Sensitivity to parameters: As an illustration, we set γ to 0.001, and orderly chose a η and a λ from $\{0.001, \ldots, 0.003, 0.07, 0.01, 0.03, \ldots, 0.07, 0.1, 0.2\}$ and $\{1, \ldots, 10\}$, respectively. Then, the proposed method was applied to segment a relatively large tumor (volume of 135.80 mL) and a relatively small tumor (volume of 7.60 mL). The obtained DSCs are summarized in Fig. 7. We can observe that the obtained DSCs as a function of λ is relatively stable in a wide range. On the other hand, for the relatively large tumor, our method had better performance with small η; while, on the contrary, large η is better for the relatively small tumor.

VI. CONCLUSION

In this study, a novel co-clustering algorithm based on BFT has been proposed to jointly segment tumors in PET-CT images. Image voxels have been described not only by intensities but also by complementary textural features. For reliable clustering of image voxels in both PET and CT images, a specific spatial regularization and distance metric adaptation procedure have been included in the proposed method. A specific context term has been proposed in the framework of belief functions to encourage consistent segmentation between the two distinct mono-modalities. To effectively combine complementary information in PET and CT images, during the minimization of the constructed cost function, clustering results in the two mono-modalities have been iteratively adjusted by fusing them via the Dempster's combination rule. The experimental results have shown that the proposed method performs well compared with existing mono-modal segmentation methods and co-segmentation methods.

REFERENCES

[1] M. Vallières, C. Freeman, S. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, p. 5471, 2015.

[2] C. Lian, S. Ruan, T. Denœux, F. Jardin, and P. Vera, "Selecting radiomic features from FDG-PET images for cancer treatment outcome prediction," *Medical Image Analysis*, vol. 32, pp. 257–268, 2016.

[3] M. Hatt, F. Tixier, D. Visvikis, and C. C. Le Rest, "Radiomics in PET/CT: more than meets the eye?" *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 365–366, 2017.

[4] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014.

[5] M. Hatt, J. A. Lee, C. R. Schmidtlein, I. E. Naqa, C. Caldwell, E. De Bernardi, W. Lu, S. Das, X. Geets, V. Gregoire *et al.*, "Classification and evaluation strategies of auto-segmentation approaches for PET: report of AAPM task group No. 211," *Medical Physics*, vol. 44, no. 6, 2017.

[6] S. Vauclin, K. Doyeux, S. Hapdey, A. Edet-Sanson, P. Vera, and I. Gardin, "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 & Biology*, vol. 54, no. 22, p. 6901, 2009.

[7] W. Jentzen, L. Freudenberg, E. G. Eising, M. Heinze, W. Brandau, and A. Bockisch, "Segmentation of PET volumes by iterative image thresholding," *Journal of Nuclear Medicine*, vol. 48, no. 1, pp. 108–114, 2007.

[8] H. Li, W. L. Thorstad, K. J. Biehl, R. Laforest, Y. Su, K. I. Shoghi, E. D. Donnelly, D. A. Low, and W. Lu, "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] S. Tan, L. Li, W. Choi, M. K. Kang, W. D DSouza, and W. Lu, "Adaptive region-growing with maximum curvature strategy for tumor segmentation in 18F-FDG PET," *Physics in Medicine & Biology*, vol. 62, no. 13, p. 5383, 2017.

[10] M. Abdoli, R. Dierckx, and H. Zaidi, "Contourlet-based active contour model for PET image segmentation," *Medical Physics*, vol. 40, no. 8, 2013.

[11] V. Jaouen, P. Gonzalez, S. Stute, D. Guilloteau, S. Chalon, I. Buvat, and C. Tauber, "Variational segmentation of vector-valued images with gradient vector flow," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4773–4785, 2014.

[12] D. Onoma, S. Ruan, S. Thureau, L. Nkhali, R. Modzelewski, G. Monnehan, P. Vera, and I. Gardin, "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] H. Mi, C. Petitjean, P. Vera, and S. Ruan, "Joint tumor growth prediction and tumor segmentation on therapeutic follow-up PET images," *Medical Image Analysis*, vol. 23, no. 1, pp. 84–91, 2015.

[14] R. R. Beichel, M. Van Tol, E. J. Ulrich, C. Bauer, T. Chang, K. A. Plichta, B. J. Smith, J. J. Sunderland, M. M. Graham, M. Sonka *et al.*, "Semiautomated segmentation of head and neck cancers in 18F-FDG PET scans: A just-enough-interaction approach," *Medical Physics*, vol. 43, no. 6, pp. 2948–2964, 2016.

[15] X. Geets, J. A. Lee, A. Bol, M. Lonneux, and V. Grégoire, "A gradient-based method for segmenting FDG-PET images: methodology and validation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, no. 9, pp. 1427–1438, 2007.

[16] M. Wanet, J. A. Lee, B. Weynand, M. De Bast, A. Poncelet, V. Lacroix, E. Coche, V. Grégoire, and X. Geets, "Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens," *Radiotherapy and Oncology*, vol. 98, no. 1, pp. 117–125, 2011.

[17] D. W. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," *Medical Physics*, vol. 34, no. 2, pp. 722–736, 2007.

[18] M. Hatt, C. C. Le Rest, A. Turzo, C. Roux, and D. Visvikis, "A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 881–893, 2009.

[19] A.-S. Dewalle-Vignion, N. Betrouni, R. Lopes, D. Huglo, S. Stute, and M. Vermandel, "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.

[20] M. S. Sharif, M. Abbod, A. Amira, and H. Zaidi, "Artificial neural network-based system for PET volume segmentation," *Journal of Biomedical Imaging*, vol. 2010, p. 4, 2010.

[21] B. Berthon, C. Marshall, M. Evans, and E. Spezi, "ATLAAS: An automatic decision tree-based learning algorithm for advanced image segmentation in positron emission tomography," *Physics in Medicine & Biology*, vol. 61, no. 13, p. 4855, 2016.

[22] B. Foster, U. Bagci, Z. Xu, B. Dey, B. Luna, W. Bishai, S. Jain, and D. J. Mollura, "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.

[23] 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.

[24] B. Lelandais, S. Ruan, T. Denœux, P. Vera, and I. Gardin, "Fusion of multi-tracer PET images for dose painting," *Medical Image Analysis*, vol. 18, no. 7, pp. 1247–1259, 2014.

[25] J. Lapuyade-Lahorgue, D. Visvikis, O. Pradier, C. Cheze Le Rest, and M. Hatt, "SPEQTACLE: An automated generalized fuzzy C-means algorithm for tumor delineation in PET," *Medical Physics*, vol. 42, no. 10, pp. 5720–5734, 2015.

[26] C. Lian, S. Ruan, T. Denoeux, H. Li, and P. Vera, "Spatial evidential clustering with adaptive distance metric for tumor segmentation in FDG-PET images," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 21–30, 2018.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[28] M. Hatt, B. Laurent, A. Ouahabi, H. Fayad, S. Tan, L. Li, W. Lu, V. Jaouen, C. Tauber, J. Czakon *et al.*, "The first MICCAI challenge on PET tumor segmentation," *Medical Image Analysis*, vol. 44, pp. 177–195, 2018.

[29] H. Gribben, P. Miller, G. G. Hanna, K. J. Carson, and A. R. Hounsell, "MAP-MRF segmentation of lung tumours in PET/CT image," in *2009 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2009, pp. 290–293.

[30] J. Wojak, E. D. Angelini, and I. Bloch, "Joint variational segmentation of CT-PET data for tumoral lesions," in *2010 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2010, pp. 217–220.

[31] W. Mu, Z. Chen, W. Shen, F. Yang, Y. Liang, R. Dai, N. Wu, and J. Tian, "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.

[32] D. Han, J. Bayouth, Q. Song, A. Taurani, M. Sonka, J. Buatti, and X. Wu, "Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method," in *International Conference on Information Processing in Medical Imaging (IPMI)*. Springer, 2011, pp. 245–256.

[33] Q. Song, J. Bai, D. Han, S. Bhatia, W. Sun, W. Rockey, J. E. Bayouth, J. M. Buatti, and X. Wu, "Optimal co-segmentation of tumor in PET-CT images with context information," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1685–1697, 2013.

[34] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. J. Mollura, "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.

[35] Z. Xu, U. Bagci, J. K. Udupa, and D. J. Mollura, "Fuzzy connectedness image co-segmentation for hybrid PET/MRI and PET/CT Scans," in *Computational Methods for Molecular Imaging*. Springer, 2015, pp. 15–24.

[36] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on PET-CT images," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5854–5867, 2015.

[37] H. Cui, X. Wang, J. Zhou, S. Eberl, Y. Yin, D. Feng, and M. Fulham, "Topology polymorphism graph for lung tumor segmentation in PET-CT images," *Physics in Medicine & Biology*, vol. 60, no. 12, pp. 4893–4914, 2015.

[38] 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.

[39] J. Zhang, J. Liang, and H. Zhao, "Local energy pattern for texture classification using self-adaptive quantization thresholds," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 31–42, 2013.

[40] G. Thibault, J. Angulo, and F. Meyer, "Advanced statistical matrices for texture characterization: application to cell classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 630–637, 2014.

[41] I. El Naqa, D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J. D. Bradley, P. Grigsby, and J. O. Deasy, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning," *Medical Physics*, vol. 34, no. 12, pp. 4738–4749, 2007.

[42] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning," *IEEE Transactions on Medical Imaging*, vol. 28, no. 3, pp. 374–383, 2009.

[43] D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, and A. Sun, "Automatic segmentation of lung carcinoma using 3D texture features in 18-FDG PET/CT," *International Journal of Molecular Imaging*, vol. 2013, 2013.

[44] H. Hanzouli-Ben Salah, J. Lapuyade-Lahorgue, J. Bert, D. Benoit, P. Lambin, A. Baardwijk, E. Monfrini, W. Pieczynski, D. Visvikis, and M. Hatt, "A framework based on hidden Markov trees for multimodal PET/CT image co-segmentation," *Medical Physics*, vol. 44, no. 11, pp. 5835–5848, 2017.

[45] J. Lapuyade-Lahorgue, J.-H. Xue, and S. Ruan, "Segmenting multi-source images using hidden Markov fields with copula-based multivariate statistical distributions." *IEEE Transactions on Image Processing*, vol. 26, no. 27, pp. 3187–3195, 2017.

[46] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.

[47] C. Lian, J. Zhang, M. Liu, X. Zong, S.-C. Hung, W. Lin, and D. Shen, "Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images," *Medical Image Analysis*, vol. 46, pp. 106–117, 2018.

[48] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976, vol. 1.

[49] X. Li, A. Dick, C. Shen, Z. Zhang, A. van den Hengel, and H. Wang, "Visual tracking with spatio-temporal Dempster–Shafer information fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3028–3040, 2013.

[50] Z. Liu, G. Li, G. Mercier, Y. He, and Q. Pan, "Change detection in heterogenous remote sensing images via homogeneous pixel transformation," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1822–1834, 2018.

[51] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence," *Image and Vision Computing*, vol. 24, no. 1, pp. 61–69, 2006.

[52] M. E. Y. Boudaren, L. An, and W. Pieczynski, "Unsupervised segmentation of SAR images using Gaussian mixture-hidden evidential Markov fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1865–1869, 2016.

[53] M.-H. Masson and T. Denœux, "ECM: An evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 41, no. 4, pp. 1384–1397, 2008.

[54] N. Makni, N. Betrouni, and O. Colot, "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.

[55] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, and Y. He, "Classifier fusion with contextual reliability evaluation," *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1605–1618, 2018.

[56] Y. Bi, J. Guan, and D. Bell, "The combination of multiple classifiers using an evidential reasoning approach," *Artificial Intelligence*, vol. 172, no. 15, pp. 1731–1751, 2008.

[57] L. Jiao, T. Denoeux, and Q. Pan, "A hybrid belief rule-based classification system based on uncertain training data and expert knowledge," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 12, pp. 1711–1723, 2016.

[58] F. Li, S. Li, and T. Denœux, "k-CEVCLUS: Constrained evidential clustering of large dissimilarity data," *Knowledge-Based Systems*, 2017.

[59] Z.-G. Liu, Q. Pan, J. Dezert, and A. Martin, "Combination of classifiers with optimal weight based on evidential reasoning," *IEEE Transactions on Fuzzy Systems*, 2017.

[60] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.

[61] M. Liu and D. Zhang, "Pairwise constraint-guided sparse learning for feature selection," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 298–310, 2016.

[62] C. Lian, S. Ruan, and T. Denœux, "Dissimilarity metric learning in the belief function framework," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 6, pp. 1555–1564, 2016.

[63] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical Programming*, vol. 107, no. 3, pp. 391–408, 2006.

[64] 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.