Kinetic Modeling based Probabilistic Segmentation for Molecular Images

Ahmed Saad[1,2], Ghassan Hamarneh[1], Torsten Möller[2], and Ben Smith[1,2]

[1] Medical Image Analysis Lab, [2] Graphics, Usability, and Visualization Lab,
School of Computing Science, Simon Fraser University, Canada,
{aasaad, hamarneh, torsten, brsmith}@cs.sfu.ca

Abstract. We propose a semi-supervised, kinetic modeling based segmentation technique for molecular imaging applications. It is an iterative, self-learning algorithm based on uncertainty principles, designed to alleviate low signal-to-noise ratio (SNR) and partial volume effect (PVE) problems. Synthetic fluorodeoxyglucose (FDG) and simulated Raclopride dynamic positron emission tomography (dPET) brain images with excessive noise levels are used to validate our algorithm. We show, qualitatively and quantitatively, that our algorithm outperforms state-of-the-art techniques in identifying different functional regions and recovering the kinetic parameters.

1 Introduction

Molecular imaging is a new scientific area merging concepts of molecular biology with non-invasive imaging technologies. In dPET [1], for example, radioactive tracers can act as biomarkers to track spatio-temporal molecular and cellular processes. Kinetic modeling (KM) is an essential step for molecular image quantification [2]. KM parameters of tissue perfusion, tracer transport or receptor binding are typically calculated by fitting the average time activity curves (TACs) within a region of interest (ROI) to underlying mathematical models. Manual ROI delineation of functional regions is a tedious, time-consuming, and error-prone task, that relies on subjective user assessment and is influenced by image quality. This leads to inaccuracies in KM parameter estimation.

In this paper, we focus on segmentation and KM parameter estimation from dPET as it is considered the leading technology for molecular imaging due to its high specificity and sensitivity. Nevertheless, our proposed method is applicable to other modalities whose analysis utilize KM such as dynamic contrast enhanced magnetic resonance imaging and contrast enhanced ultrasound.

Several methods had been proposed to segment dPET images into different functional regions. Barber applied factor analysis to planar dynamic Gamma camera images to identify functional regions [3]. Gou et al. applied hierarchical TAC clustering [4]. This algorithm was purely based on the TAC dynamics ignoring the spatial-domain information. Kamasak et al. simultaneously clustered and estimated each cluster's TAC directly from the sinogram data, without the need for tomographic reconstruction [5]. However, using such a projection domain

based approach makes it difficult to incorporate much needed user knowledge given the low SNR. Maroy et al. extracted the dPET TACs in the organ cores, where they are least affected by motion and spillover effects [6]. This algorithm might be problematic for regions with small cores and low SNR, since a large number of TACs is needed to achieve acceptable estimates of organ kinetics. Saad et al. incorporated a KM based image prior constraint during the segmentation instead of considering only the observed TACs [7]. However, their approach did not guarantee globally optimal segmentations nor did it address PVE.

In this paper, we develop a multi-class, seed-initialized, iterative segmentation algorithm for molecular images. Random walker (RW) [8] is adopted at the core of our algorithm (Sec. 2.1) guaranteeing a global optimal segmentation in each iteration with automatically updated seeds. A KM image prior term is incorporated into RW to capture the underlaying physiological information (Sec. 2.2). An iterative, self-learning, uncertainty-based approach is developed to overcome the low SNR for image prior estimation, especially with a low number of user provided seeds (Sec. 2.3). The KM parameter fitting utilizes a confidence-weighted averaging of TACs in each region, which addresses the problem of PVE (Sec. 2.4).

2 Method

The details of our KM based RW with self-learning (RWSL-KM) method are given in the following sections.

2.1 Random walker

As proposed in [8], the RW segmentation is formulated in a graph theoretic setting where image pixels are graph vertices v and edge weights w reflect similarity between neighboring pixels. The probability of assigning label $q \in \{1, 2, ..., Q\}$ to vertex v_i reflects the probability of a random walker starting at v_i reaching the seeds of region q out of Q possible regions. This results in a probability vector x_i^q. The set of user-labeled seeds is denoted V_L and the set of unlabeled vertices is denoted V_U. The probability of a labeled vertex $v_i \in V_L$ with label s is given by $x_i^q = 1$ if $q = s$ and 0 otherwise. Grady et al.[8] showed that an x^q that minimizes the energy functional

$$E_{total}^q(x^q) = \gamma_{reg} E_{reg}^q(x^q) + \gamma_{prior} E_{prior}^q(x^q), \tag{1}$$

can be obtained by solving the following linear system of equations

$$(\gamma_{reg} \triangle_{UU} + \gamma_{prior} \sum_{q=1}^{Q} \varLambda_U^q) x_U^q = -\gamma_{reg} \triangle_{UL} x_L^q + \gamma_{prior} \lambda_U^q \tag{2}$$

where E_{prior} and E_{reg} are the data fidelity and regularization energy terms, respectively. $\triangle_{UU}$ and $\triangle_{UL}$ are partitions from the combinatorial graph Laplacian matrix defined in [8]. λ_i^q represent the likelihood that vertex v_i belongs to class q. $\varLambda^q$ is a diagonal matrix with values λ^q along the diagonal. γ_{reg} and γ_{prior} are weighting parameters; setting their values is discussed in (Sec. 2.3). With x_i^q in

hand, vertex v_i is assigned a label using argmax x_i^q.

We extend the classical RW formulation to deal with time-varying dPET images in the following ways:

- Each vertex v_i now represents a TAC I_i, where $I_i = [I_i^1 I_i^2 ... I_i^T]$ and T is the total number of time steps. I_i is typically sampled non-uniformly with time, with smaller sampling intervals suffering from lower SNR, which in turn is due to the lower Gamma photon counts acquired in a shorter interval.
- The similarity between two TACs I_i and I_j is measured using the weighted L_2 distance $g_{ij} = \sqrt{\sum_{t=1}^{T}(I_i^t - I_j^t)^2 z_t}$, where z_t is a weight that accounts for the non-uniform time sampling encountered in dPET. Shorter sampling intervals with lower SNR are weighted less.
- A reciprocal weighting function is used to calculate the edge weights $w_{ij} = \frac{\beta}{g_{ij} + \epsilon}$, where β is a mapping parameter and ϵ is a small number for stability.

2.2 Kinetic modeling based image prior terms

In order to calculate the likelihood term λ_i^q in (2), we need to estimate the TAC distribution in different classes. However, higher dimensional density estimation is impractical beyond a few dimensions [9], which is the case for dPET data since a TAC at each pixel is a roughly 40 dimensional vector ($T \approx 40$ time steps). Further, there are no training images that we can use for the density estimation as different tracers lead to completely different intensity distributions. Here, we assume that the TAC classes are multivariate Gaussian distributions with the same scaled identity covariance matrix but with different means, which is the same assumption as in the widely used K-means clustering algorithm. We can now consider the distance to the class mean TAC as the inverse of the likelihood term with proper mapping, as follows

$$u_i^q = \sqrt{\sum_{t=1}^{T} \|I_i^t - \mu_q^t\|^2 z_t} \tag{3}$$

where u_i^q represents the weighted L_2 intensity difference between TAC I_i and mean TAC μ_q for class q.

As dPET tracer kinetics are often described using compartmental KM, we apply the KM process to the means μ^q for the Q classes prior to applying (3). We thus ensure that the prior terms capture the physiological phenomena under consideration and not only the observed TACs from the dPET data. Consequently, (3) becomes

$$u_{fit_i}^q = \sqrt{\sum_{t=1}^{T} \|I_i^t - \mu_{fit_q}^t\|^2 z_t} \tag{4}$$

where μ_{fit_q} is the activity TAC produced by solving a specific KM (e.g. two tissue compartmental KM) for each region mean activity μ_q [7]. The likelihood at each vertex is then estimated by

$$\lambda_{fit_i}^q = \Upsilon / (u_{fit_i}^q + \epsilon) \tag{5}$$

where Υ is a free mapping parameter. By incorporating KM TAC likelihood terms into RW, (2) becomes

$$(\gamma_{reg}\triangle_{UU} + \gamma_{prior}\sum_{q=1}^{Q}\Lambda_{U_{fit}}^{q})x_{U}^{q} = -\gamma_{reg}\triangle_{UL}x_{L}^{q} + \gamma_{prior}\lambda_{U_{fit}}^{q}. \qquad (6)$$

2.3 Self-learning using uncertainty principles

The main assumption underlying RW is that labels are diffused on the graph according to the similarity between TACs. On the one hand, low SNR affects the similarity calculations, which in turn affects the energy term E_{reg}. On the other hand, we can not build a good estimate for E_{prior} with the low number of seeds expected from the user. Here, we develop a self-learning technique to boost the number of vertices used in the prior estimation E_{prior}. Self-learning is a commonly used concept in semi-supervised learning, in which a classifier is first trained with a small amount of labeled data and then used to classify the unlabeled data. Then, the most confident unlabeled vertices, together with their predicted labels, are added to the training set. The classifier is re-trained and the procedure repeated [10]. To identify the most confident vertices, we use uncertainty principles from information theory. The entropy at v_i is defined as

$$h(v_i) = \sum_{q=1}^{Q} x_i^q \log_2(x_i^q). \qquad (7)$$

Vertices with sufficiently high certainty or confidence will have a sufficiently low entropy, i.e. $h(v_i) < \aleph$, where $\aleph$ is a confidence threshold, and are, hence, dubbed *confident*. Nevertheless, confident vertices are not at par with user-specified seeds. Therefore, they are added to a new set $V_{confident}$, rather than to the seeds set V_L (i.e. $x^q \neq 1$ for $v_i \in V_{confident}$ whereas $x_L^q = 1$ for $v_i \in V_L$), to allow them to be relabeled as the algorithm iterates. Each region mean TAC μ_q is estimated by averaging all TACs $\forall v_i \in V_{confident} \cup V_L$ with label q.

Energy minimizing segmentation methods, RW included, are generally sensitive to the choice of weighting between spatial regularization and conformance to certain appearance or pixel intensity. In (6), such weighting balance is dictated by the weighting parameters γ_{req} and γ_{prior}. We devise a novel, automatic iteration-dependant schedule for setting these weighting parameters. Given an initial small number of seeds, we wish to weight the E_{reg} term more to ensure that the vertices added to $V_{confident}$ will be spatial neighbors and similar to the seeds. As the algorithm iterates, we wish to weight the E_{prior} more as we will have a better estimate of the TAC distribution in different classes given the larger number of vertices added to $V_{confident}$. So (6) is modified as follows

$$(\alpha\triangle_{UU} + (1-\alpha)\sum_{q=1}^{Q}\Lambda_{U_{fit}}^{q})x_{U}^{q} = -\alpha\triangle_{UL}x_{L}^{q} + (1-\alpha)\lambda_{U_{fit}}^{q}. \qquad (8)$$

The α scheduling can be written as $\alpha(k+1) = \zeta\alpha(k)$ where $0 < \zeta < 1$ and k is the iteration index.

2.4 Certainty weighted mean TAC

After the labeling process, the standard KM proceeds by *equally-weighted* averaging of all the TACs belonging to each functional region to produce μ_{final_q} [2]. Here, we propose a *certainty-weighted* averaging to calculate each region's mean TAC by utilizing the confidence of the labeling process. This compensates for the PVE by weighting pure tissue TACs differently, compared to those of mixture-tissues using the uncertainty principles. More specifically

$$\mu_{final_q} = \frac{\sum_{v_i \text{ labeled } q} \psi_i I_i}{\sum_{v_i \text{ labeled } q} \psi_i} \quad \text{where} \quad \psi_i = 1/(h(v_i) + \epsilon). \tag{9}$$

Our RWSL-KM method is summarized in Alg.1.

Algorithm 1 Kinetic modeling based random walker with self-learning

---
Input: (i) dynamic molecular image data and (ii) user seeds for different classes.
Result: Segmented functional regions with their corresponding kinetic parameters.
Initialize V_L from user seeds and $V_{confident} = \{\}$.
Initialize $k = 0$ and $\alpha(k) = 1$.
repeat
    Compute the probability $x_i^q \,\forall q$ at each $v_i \in V_U$ by applying (8).
    Assign a class to v_i using $\underset{q}{\operatorname{argmax}}\ x_i^q$.
    Update $V_{confident}$ by adding v_i to $V_{confident}$ iff $h(v_i) < \aleph$.
    Compute each region mean TAC μ_q using only $v_i \in V_{confident} \cup V_L$.
    Find KM parameters producing μ_{fit_q} closest to μ_q in the least-square sense [7].
    Update $\alpha(k+1) = \zeta\alpha(k)$ and $k = k + 1$.
until Convergence: No significant change in either μ_{fit_q} or $V_{confident}$
Compute *final* certainty-weighted region mean TAC μ_{final_q} by applying (9).

---

3 Materials and implementation

We used synthetic and simulated dPET data with known ground truth to evaluate the proposed algorithm. For the synthetic dPET data, we generated an [18F]FDG-PET image dataset with TACs corresponding to 6 functional regions ($Q = 6$): background (BG), skull (SK), grey matter (GM), white matter (WM), cerebellum (CM) and putamen (PN). A two-tissue compartmental KM was used to simulate the dynamics for different regions with real kinetic parameters from the dPET clinical literature. The resulting TACs are assigned to a 2D downsampled labeled MRI image that was blurred with a Gaussian filter (full width at half maximum = 8mm). We ran 5 trials each with 5 Gaussian noise levels $[1, 3, 5, 8, 10]\sigma$. σ is used to scale the unit variance of the random noise generator to the scale of the synthetic TAC intensity at each time step. The dPET image has dimensions of 91×109 with 46 time steps: 12×10s, 10×30s, 10×120s, 10×300s and finally 4×600s with an isotropic voxel size of 2mm. Fig 1 shows the MRI slice used and the last dPET time step with different noise levels. We show the last time step as it has the highest SNR, as is typical in dPET (i.e. the preceding time frames are even noisier).

For the simulated dPET data, we used the publicly available simulated 3D+time dataset PET-SORTEO with known ground truth [11]. Ten simulated
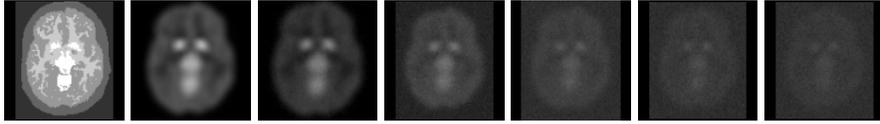
Fig. 1. From left to right: Labeled 2D Brain MRI slice. The last time step slice (highest SNR) of the synthetic dPET data blurred with a Gaussian kernel (i.e. introduces PVE) with increasing noise levels $[0, 1, 3, 5, 8, 10]\sigma$, where 0σ level has no additive noise.

[11C]Raclopride PET brain studies, accounting for inter-subject anatomical variability as well as different dPET image degradation factors, have been used. The PET-SORTEO image has dimensions of $128\times128\times63$ with 26 time steps: 6×30s, 7×60s, 5×120s and finally 8×300s with voxel size of $2.11 \times 2.11 \times 2.42$mm^3.

In order to validate our algorithm, we compare it with two of the state-of-the-art techniques in dPET segmentation: K-means with KM and a MRF-regularized version thereof [7], abbreviated KMN-KM and MRF-KM, respectively. We also compare to the original RW (with necessary extensions noted in (Sec. 2.1)) [8]. For the synthetic experiment, the comparison criterion is based on the recovery of the FDG glucose metabolic rate $K = K_1 k_3 / (k_2 + k_3)$ [4]. We define the KM recovery error as

$$RE^q = \|K_{fit}^q - K_{true}^q\|^2 \tag{10}$$

where K_{true}^q is the ground truth K used to generate the data in a functional region q. A perfect segmentation technique is the one that can recover K_{true}^q (yielding $RE^q = 0$). Each segmentation technique produces different K_{fit}^q.

The KM development is based on COMKAT [12]. The RW is partially based on the code accompanying [8]. We chose empirically the following values for the algorithm free parameters $\beta = 1$, $\Upsilon = 1$, $\alpha(0) = 1$, $\zeta = 0.9$ and $\aleph = 1.1863$ with $Q = 6$. The value of $\aleph$ corresponds to a minimum likelihood probability of 0.8 for one class and equal probabilities for the other classes. The same set of seeds was used for the initialization of different segmentation techniques: to calculate the initial means for KMN-KM and MRF-KM, and as initial seeds for RW and RWSL-KM. $|V_L|$ is 2.5% of the original image size in the 2D synthetic experiment and 0.02% in the simulated 3D experiment.

4 Results

Fig 2 qualitatively compares the four algorithms using the synthetic data with a noise level of 5σ. KMN-KM and MRF-KM result in isolated, scattered functional labels especially in the WM and GM regions. Further, the PN is underestimated as it has very similar kinetics to the CM region. Classical RW results in Voronoi regions with respect to the provided seeds due to the excessive noise (in agreement with [8]). RWSL-KM overcomes these two limitations and performs better than the other approaches. Most of the regions are sufficiently recovered especially in the PN region. Nevertheless, RWSL-KM misclassified a band around the CM as PN (green surrounding blue). Also, the area between the GM and the SK was misclassified as WM (the yellow perimeter). These misclassifications are the result of PVE. This behavior is not surprising and can further be explored by examining the uncertainty map (rightmost image in Fig 2). This image is a
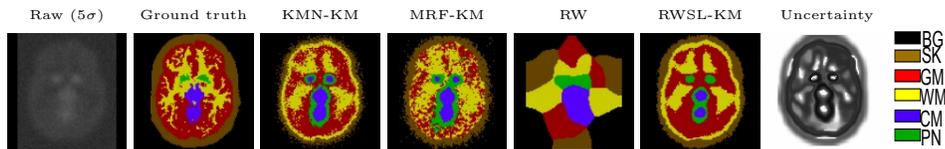
Fig. 2. Comparison between four segmentation algorithms. The rightmost figure is the negative entropy map generated by RWSL-KM (uncertain regions appear darker).
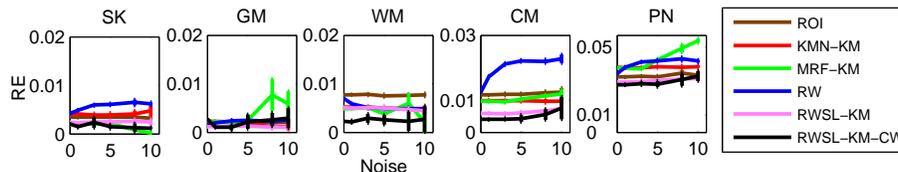

Fig. 3. Recovery error (RE) of the physiological parameters with different noise levels and multiple trials. ROI corresponds to the mean TAC extracted using the ground truth labeling. RWSL-KM-CW uses the certainty weighted mean TAC for each region. Note the different scale on the y-axis to demonstrate the differences between algorithms.

mapping of the negative entropy at each pixel to a grey level color: uncertain pixels appear darker. Misclassified regions with low certainty (darker) indicate that the classification of these regions is almost arbitrary. This supports our justification that the extracted mean TAC from each region should be weighted by the amount of certainty at each voxel (as in (9)) before estimating the kinetic parameters, which is the main objective of dPET analysis and not the segmentation per se.

Fig 3 shows a quantitative comparison between the different algorithms according to (10). It shows that K calculated from certainty-weighted mean TAC using RWSL-KM (RWSL-KM-CW) constantly outperforms other algorithms. Furthermore, Fig 3 shows that our method even outperforms ground truth labeling. This may seem unrealistic at first, however given that the ground truth labeling itself is a crisp (non fuzzy) labeling, TACs at the interface between functional regions will suffer from PVE and hence produce worse K estimates. This is particularly noticed in WM, CM and PN. Our certainty-weighted averaging, on the other hand, can handle the PVE well.

For the simulated dPET data, we don't have access to the generated kinetics, so we can not evaluate the algorithm according to (10). Hence, we chose the Dice coefficient to measure the overlap between a segmented region and ground truth. Note that the Dice metric doesn't take the segmentation confidence into account, so one of the features of our algorithm is unutilized. Fig 4 shows the Dice coefficient averaged over 10 patients. It shows the RWSL-KM outperforming the different algorithms except for the CM region and slightly in the GM region. This can be explained by knowing that the CM region should be devoid of the Raclopride tracer, which is the reason it is used as the reference tissue in graphical KM methods [2]. Further, RWSL-KM exhibits the lowest variability in different regions, especially in the GM region.
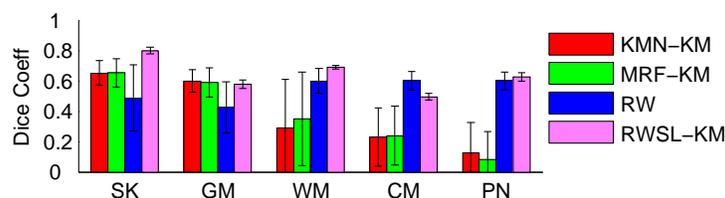
Fig. 4. Performance evaluation between the four algorithms using the Dice metric. The error bars show one standard deviation.

5   Discussion

In this paper, we developed a semi-supervised segmentation technique for molecular images incorporating spatial regularization, kinetic modeling and uncertainty principles. We showed qualitatively and quantitatively that RWSL-KM produces better results than other state-of-the-art techniques.

We intend to extend RWSL-KM by incorporating structural MRI or CT information for better localization of tracer activity. Further, we need to investigate the performance of our algorithm with real datasets from multiple modalities.

References

1. Rudin, M.: Molecular Imaging: Basic Principles and Applications in Biomedical Research. Imperial College Press (2005)
2. Morris, E.D., Endres, C.J., Schmidt, K.C., Christian, B.T.: Kinetic modeling in Positron Emission Tomography. In Wernick, M., Aarsvold, J.N., eds.: Emission Tomography: The Fundamentals of PET and SPECT. San Diego: Academic (2004)
3. Barber, D.C.: The use of principal components in the quantitative analysis of Gamma camera dynamic studies. Phys. Med. Biol. **25** (1980) 283–292
4. Guo, H., Renaut, R., Chen, K., Reiman, E.: Clustering huge data sets for parametric PET imaging. Journal of Biosystems **71** (2001) 81–92
5. Kamasak, M.E., Bayraktarb, B.: Unsupervised clustering of dynamic PET images on the projection domain. In: Proceedings of the SPIE Medical Imaging. (2006) 1539–1548
6. Maroy, R., Boisgard, R., Comtat, C., Frouin, V., Cathier, P., Duchesnay, E., Dolle, F., Nielsen, P.E., Trebossen, R., Tavitian, B.: Segmentation of Rodent whole-body dynamic PET images: An unsupervised method based on voxel dynamics. IEEE TMI **27** (2008) 342–354
7. Saad, A., Smith, B., Hamarneh, G., Möller, T.: Simultaneous segmentation, kinetic parameter estimation, and uncertainty visualization of dynamic PET images. In: Lecture Notes in Computer Science, (MICCAI). (2007) 726–733
8. Grady, L.: Multilabel random walker image segmentation using prior models. In: Proceedings of CVPR. (2005) 763–770
9. Scott, D.: Multivariate Density Estimation. New York: Wiley (1992)
10. Zhu, X.: Semi-supervised learning literature survey. Technical report, Department of Computer Sciences, University of Wisconsin, Madison (2005)
11. Reilhac, A., Batan, G., Michel, C., Grova, C., Tohka, J., Costes, N., Evans, A.C.: Validation of PET SORTEO: a platform for simulating realistic PET studies and development of a database of simulated PET volumes. IEEE Trans. Nucl. Sci **52** (2004) 1321–1328
12. Muzic, R., Cornelius, S.: COMKAT: compartment model kinetic analysis tool. JNM **42(4)** (2001) 636–645