Region-Based Nasopharyngeal Carcinoma Lesion Segmentation from MRI Using Clustering- and Classification-Based Methods with Learning

Wei Huang · Kap Luk Chan · Jiayin Zhou

Published online: 2 August 2012
© Society for Imaging Informatics in Medicine 2012

Abstract In clinical diagnosis of nasopharyngeal carcinoma (NPC) lesion, clinicians are often required to delineate boundaries of NPC on a number of tumor-bearing magnetic resonance images, which is a tedious and time-consuming procedure highly depending on expertise and experience of clinicians. Computer-aided tumor segmentation methods (either contour-based or region-based) are necessary to alleviate clinicians' workload. For contour-based methods, a minimal user interaction to draw an initial contour inside or outside the tumor lesion for further curve evolution to match the tumor boundary is preferred, but parameters within most of these methods require manual adjustment, which is technically burdensome for clinicians without specific knowledge. Therefore, segmentation methods with a minimal user interaction as well as automatic parameters adjustment are often favored in clinical practice. In this paper, two region-based methods with parameters learning are introduced for NPC segmentation. Two hundred fifty-three MRI slices containing NPC lesion are utilized for evaluating the performance of the two methods, as well

as being compared with other similar region-based tumor segmentation methods. Experimental results demonstrate the superiority of adopting learning in the two introduced methods. Also, they achieve comparable segmentation performance from a statistical point of view.

Keywords Nasopharyngeal carcinoma · Magnetic resonance image · Tumor segmentation · Spectral clustering · Support vector machine · Metric learning

Introduction

Nasopharyngeal carcinoma (NPC), originating in the nasopharynx, is a malignant tumor lesion which often happens among a number of certain populations, including natives of South China, Southeast Asia, the Arctic, the Middle East, and North Africa [1]. According to a statistical survey conducted in [2], more than 80,000 new NPC cases were diagnosed worldwide in 2002, and 50,000 deaths were reported. Early diagnosis for NPC is critical for its later successful treatment [1, 2].

In clinical practice, magnetic resonance image (MRI) has been utilized as a preferred imaging modality for the evaluation of local, regional, and intracranial infiltration of NPC, due to its high spatial resolution for examining soft tissues [3]. NPC is tumor-staged by anatomical extension, and there are many studies reporting that the tumor volume is a dominant prognostic indicator for its disease progression [4–6]. Therefore, in order to measure the tumor volume precisely, accurate delineation of NPC in tumor-bearing MRI slices is necessary. In addition, accurate delineation of NPC is also a crucial issue in treatment planning, which is the mainstay of NPC treatment. This is especially relevant in the planning of intensity-modulated radiation therapy, where

W. Huang (✉)
Information Engineering School, Nanchang University, China,
No. 999, New Xuefu Road, Honggutan,
Nanchang, Jiangxi Province 330031, China
e-mail: n060101@e.ntu.edu.sg

K. L. Chan
School of Electrical and Electronic Engineering,
Nanyang Technological University,
Singapore, Singapore
e-mail: eklchan@ntu.edu.sg

J. Zhou
Institute for Infocomm Research,
Agency of Science Technology and Research,
Singapore, Singapore
e-mail: jzhou@i2r.a-star.edu.sg

tight margin around the tumor can be prescribed with highly conformed patterns of radiation dose distribution.

In the current commercially available diagnostic imaging or radiotherapy planning system, radiologists or radiotherapists often need to manually draw tumor margins within a stack of MRI slices. It is a tedious and time-consuming procedure, which highly depends on clinicians' expertise and prior experience. Studies reported on computer-aided segmentation of NPC from MRI slices are scarce. Among several existing studies, Hsu et al. [7] developed an automatic medical diagnosis approach for segmenting NPC with dynamic MRI and pharmacogenetic analysis. This method is only capable of identifying NPC, which makes it a type of "detection" rather than "segmentation" method. Also, only one section of the head can be examined by this method. Zhou et al. [8] proposed a knowledge-based fuzzy clustering method, which takes both T1-weight (T1W) image and contrast-enhanced T1-weight (CET1W) image into consideration. A symmetric property and 2D connectivity of segmented patches were used to exclude pixels which were unlikely to be one part of NPC. The major limitation of this method is that it defines a specific parameterized distribution so that a prior assumption on the data distribution is implicitly imposed. Since MRI data is quite diverse from person to person, this method may not work well for diverse cases. Lee et al. [9] used image masking, Bayesian probability calculation, and seed growing techniques on T2-weight (T2W) images as well as contrast enhancement ratio of T1W images for NPC extraction, but some processing assumptions in their scheme may not be always valid for various NPC cases. Generally speaking, conducting NPC segmentation on MRI is usually different from performing general image segmentation. In NPC-bearing MRI slices, the tumor region and the nasal cavity area containing capillary vessels often share similar visual attributes, after patients are injected with contrast agents for tumor lesion enhancement. General image segmentation techniques based on visual characteristics of images may not be able to differentiate areas of similar intensity in MRI slices. Hence, many factors, such as clinical and anatomic prior knowledge, characteristics of specific modalities, need to be considered in NPC tumor segmentation.

For most conventional tumor segmentation methods, they can be categorized into two groups: *contour-based methods* and *region-based methods* [10–14]. For contour-based methods [10, 11], [14], the method of *active contour* has been widely utilized [15]. The main idea of active contour is to evolve a deformable curve to match the boundary of tumor lesion. The evolved curve is driven by various kinds of internal or external forces, as well as other additional constraints depending on specific characteristics of MRI slices. Although contour-based methods require a minimal user interaction to draw an initial contour inside or outside the

tumor lesion for further curve evolution, a number of parameters within such methods need to be determined by users manually [10, 11, 14, 15]. For clinicians who normally do not possess specific knowledge of such methods, it is quite inconvenient for them to apply these contour-based methods for tumor region segmentation in clinical practice. Hence, segmentation methods with parameters that can be tuned or learned automatically are preferred for real clinical usage.

Region-based methods, on the other hand, consider the tumor lesion segmentation task as a region extraction problem. In order to discern the tumor region from other non-tumor tissue regions, image pixels are either classified into different tissue categories or clustered into different groups, according to certain criteria [12, 13]. Finding a proper criterion measuring the similarity between pixels to differentiate tumor from non-tumor is often of great importance in these methods. Since MRI data often have large varying statistical properties across different slices, either making an assumption about the adopted similarity beforehand, or learning it based on some training images and then applying the learned result on other testing slices (as the conventional way in pattern recognition studies) may not fit the nature of the problem well. Therefore, it is preferable to conduct learning case by case.

With the requirements of fulfilling both a minimal user interaction and automatic parameters learning in clinical practice, this paper introduces two region-based NPC tumor segmentation methods. In Section "[Generative Clustering-Based Segmentation Method with Metric-Based Similarity Learning](#)", a generative clustering-based segmentation method is presented. It is also a novel semi-supervised segmentation method composed of two steps. First, a new spatially weighted metric-based similarity is proposed. Parameters within the similarity are learned via a spectral clustering algorithm in a supervised fashion, maximizing the separation between tumor pixels and non-tumor pixels. Second, learned similarities are used to complete the tumor segmentation in an unsupervised fashion by an out-of-sample extension strategy. In Section "[Discriminative Classification-Based Segmentation Method with Kernel Learning](#)", a discriminative classification-based segmentation method is depicted, in which support vector machine (SVM) is incorporated as the classifier and parameters in it is learned by a radius-margin bound method. In Section "[Experiments, Statistical Analysis, and Discussion](#)", segmentation performance of the two introduced methods has been evaluated by 253 pairs of tumor-bearing MRI slices. A comprehensive statistical analysis is conducted for quantitative evaluation of all obtained segmentation results, as well as being compared with other similar region-based tumor segmentation methods. Extended discussions are also elaborated in this section. In Section "[Conclusion](#)", the conclusion of this study is drawn.

Generative Clustering-Based Segmentation Method with Metric-Based Similarity Learning

A semi-supervised segmentation method based on a generative clustering technique is presented to perform NPC lesion segmentation from MRI. This method is made up of two steps. For the supervised step, a spatially weighted metric-based similarity is proposed to reflect the notion of similarity between pairwise pixels in MRI. Parameters within the proposed similarity are learned via a supervised spectral clustering algorithm. For the unsupervised step, an unsupervised out-of-sample extension strategy is incorporated to finish the whole tumor segmentation.

A Spatially Weighted Metric-Based Similarity

A spatially weighted metric-based similarity $d(x_i, x_j)$ reflecting the similarity between pixels x_i and x_j in a MRI slice is proposed as follows:

$$\begin{aligned} d(x_i, x_j) &= P(x_i, x_j) \cdot Q(x_i, x_j) \\ &= \exp\left(-\frac{\|p_i - p_j\|^2}{\sigma_p^2}\right) \\ &\quad \cdot \exp\left(-(s_i - s_j)^T A (s_i - s_j)\right) \end{aligned} \quad (1)$$

where, σ_p is a scalar and A is a full matrix; p_i (p_j) is a vector containing the normalized spatial coordinates of pixel x_i (x_j), and s_i (s_j) denotes the extracted low-level feature vector of pixel x_i (x_j). Obviously, $P(x_i, x_j)$ reflects the spatial similarity between x_i and x_j , which is constructed as an isotropically scaled Gaussian function. $Q(x_i, x_j)$ reveals the similarity of low-level features between x_i and x_j via a Mahalanobis form. The measured similarity in Eq. 1 emphasizes on spatial

localization, by which the similarity between pixels x_i and x_j decreases with the increase of their in-between distance. Therefore, two spatially nearby pixels will have more dominant influence on the similarity than two pixels that are far apart. The newly proposed similarity in Eq. 1 is different from ones used in other previous related works [16, 17]. In the *Normalized Cut* algorithm [16], Shi et al. did not include an explicit form of the term P of Eq. 1 in their similarity function and assumed that the matrix A in the term Q to be a diagonal matrix, which ignores the correlation among extracted features. In the *Ng–Jordan–Weiss* algorithm [17], only term Q of Eq. 1 is utilized in their similarity.

To ensure distances used in Eq. 1 as metrics, they should meet the four axioms of a metric (i.e., nonnegativity, identity of indiscernibles, symmetry, and triangle inequality). Therefore, the matrix A in the term Q should be at least positive semi-definite, i.e., $A \geq 0$. The purpose of learning a similarity is to find proper parameters of it, so that data from the same or different groups can be well grouped or differentiated, respectively. In this study, the purpose of learning such a similarity involves finding proper values for parameters A and σ_p^2 . Hence, parameters in our methods are determined algorithmically, not empirically.

Similarity Learning Via Supervised Spectral Clustering

Spectral clustering is a special clustering technique, which can be considered as an approximate solution to the *graph-cut* problem [18], achieving partitioning of an image by cutting weak links between graph nodes to separate an image into various segments. Similarity learning in this method is performed with a spectral clustering algorithm, whose procedure is based on [19] and presented in Table 1.

In [19], a theoretically proven cost function in Frobenius norm (i.e., the cost function in Eq. 2 of Table 1) is utilized,

Table 1 Algorithm of similarity learning via supervised spectral clustering

Input: a set of n points $S = \{s_1, s_2, \dots, s_n\}$ in $\mathbf{R}^d$

Algorithm:

1. Initialize parameters of the proposed similarity in Eq. 1: A and σ_p^2
2. Calculate the pairwise similarity $d(x_i, x_j)$ in Eq. 1 for each data pair. Use these computed $d(x_i, x_j)$ as elements to construct an affinity matrix $D \in \mathbf{R}^{n \times n}$ for all n points in S .
3. Form a new diagonal matrix C , whose (i, i) element is the sum of the i th row of D .
4. Construct a graph Laplacian matrix $L = C^{-1/2} \cdot D \cdot C^{-1/2}$ and find the k largest eigenvalues and their corresponding eigenvectors.
5. Form a new matrix $X \in \mathbf{R}^{n \times k}$ by stacking the k extracted eigenvectors in columns.
6. Solving A and σ_p^2 . Form an optimization function in Frobenius norm:

$$\min J(A, \sigma_p^2) = \frac{1}{2} \|XX^T - X_{\text{part}}X_{\text{part}}^T\|_F^2 \quad \text{s.t. } A \geq 0 \quad (2)$$

where $X_{\text{part}} = C^{-1/2}E(E^TCE)^{-1/2}$. B . E is an indicator matrix of data set partitions and B is an arbitrary orthogonal matrix. Apply $A \geq 0$ as a constraint to the optimization problem and find optimal solutions of A and σ_p^2 through a gradient descent algorithm.

Output: Learned parameters A and σ_p^2 of the metric.

in which its solution is defined up to orthogonal matrixes and subspaces spanned via columns of orthogonal matrixes are compared for learning a Mahalanobis distance-based similarity in spectral clustering. The similar learning strategy is incorporated in this study for learning our newly proposed metric-based similarity with its own constraint to be satisfied, viz. $A \geq 0$. Through a gradient descent method, parameters in Eq. 1 can be determined. Thus, a learned spatially weighted metric-based similarity is obtained.

User Interaction

In this study, a minimal user interaction, similar towards what is often conducted in conventional contour-based methods, is incorporated. It allows clinicians to draw their own region of interest (ROI), in which the tumor lesion is assumed to be enclosed. The ROI can be in any arbitrary shape, and the enclosure does not have to be very close to the lesion's boundary. Such a ROI is helpful to reduce the computational cost of the whole method, as pixels outside ROI is assumed to be from non-tumor tissues and only pixels within ROI are candidates of tumor pixels.

Using such a ROI, input S in Table 1 can also be constructed conveniently. Points inside the enclosed region are sampled as positive training samples (i.e., from tumor tissue), while points outside are considered and sampled as negative training samples (i.e., from non-tumor tissues). In the implementation, positive samples often share similar intensities and are chosen from the center area of ROI. However, negative samples are usually highly diverse in intensity and are often more difficult to be sampled. A simple random sampling strategy is likely to under-represent some sub-populations (a.k.a. *stratum*, which represents the collection of points sharing similar visual attributes in this study) of negative points, making chosen negative samples less representative. It will bias the learning result, and degrade the tumor segmentation performance therein. Hence, in order to make negative samples more representative, a *stratified random sampling strategy* [20] is incorporated in sampling negative points. The used optimal stratum allocation scheme in this study is *Neyman allocation* [20], in which the number of sampled negative points n_l from a stratum l can be decided by:

$$n_l = n \cdot \frac{R_l \cdot \sigma_l}{\sum_{k=1}^L R_k \cdot \sigma_k}, \quad R_l = \frac{N_l}{N_1 + \dots + N_L} \quad (3)$$

where, n is the total number of negative points to be sampled; σ_l is the standard deviation of intensities of negative points located at stratum l ; R_l denotes the fraction of the number of negative points in the stratum l (N_l) with respect to negative points over all L stratum. It can be observed from Eq. 3 that, Neyman allocation intends to allow more negative pixels to be sampled from one particular stratum, which contains a larger fraction of negative points with largely varying intensities. In

this way, more representative negative training samples can be obtained in constructing the training set S for parameter learning in Table 1.

Tumor Segmentation Through Unsupervised Out-of-Sample Extension

The second step of the proposed semi-supervised segmentation method is to perform tumor segmentation from MRI by spectral clustering on pixels other than those sampled for similarity learning in previous sections. It is generally accepted that the main computational burden of spectral clustering usually comes from the eigen-decomposition step of the graph Laplacian matrix $\mathbf{L}$ of size $n \times n$, where n is the total number of pixels. In the previous learning step, this is not a problem as the number of sampled points in the training set is still small. If spectral clustering is used to group all other points in the ROI later, the size of matrix $\mathbf{L}$ will become extremely large and causes memory problems (e.g., for a 200×200 pixel-wise ROI, the constructed graph Laplacian $\mathbf{L}$ will be of the size $40,000 \times 40,000$, which is often difficult to load and handle for ordinary computers) as well as computational burden in eigen-decomposition (i.e., its computational cost is around $O(n^3)$). For this reason, an unsupervised *out-of-sample extension method* [21] is necessary in this segmentation method. The purpose of incorporating out-of-sample extension is to map points into the spectral domain directly using a mapping function, without performing the eigen-decomposition step. Therefore, the main computational burden of spectral clustering can be avoided.

For pixels in ROI, the similarity is calculated with respect to each sample in the training set S using the learned similarity function (Eq. 1). In this study, since the prior knowledge about the location of tumor lesion is incorporated by drawing ROIs on MRI slices, points outside ROI should not be considered as candidate points for NPC. Thus, a weighted adjacency matrix $\mathbf{W}'$ of the size: (number of remaining samples in ROI) $\times$ (number of samples in S) can be obtained. After normalization, the affinity matrix $\mathbf{W}'_{\text{norm}}$ is given by

$$W'_{\text{norm}}(i,j) = \frac{W'_{ij}}{\sqrt{\sum_{x \in S} W'_{ix} \sum_{x' \in S} W'_{jx'}}} \quad (4)$$

The spectral embedding of pixels in ROI can be computed using the below mapping function directly:

$$\mathbf{X}_{\text{mapped}} = \mathbf{W}'_{\text{norm}} \cdot \mathbf{X} \cdot \mathbf{V}^{-1} \quad (5)$$

where, $\mathbf{X}$ and $\mathbf{V}$ are the matrices of eigenvectors and eigenvalues obtained from the training data set S , respectively. After obtaining $\mathbf{X}_{\text{mapped}}$, conventional clustering algorithms can be utilized to finish segmenting the tumor lesion unsupervisedly therein. In this study, the classic *K-means* algorithm is

incorporated in this step: the pre-defined number of groups equals to 2 ($K=2$) for tumor and non-tumor groups.

Discriminative Classification-Based Segmentation Method with Kernel Learning

In pattern recognition and machine learning, *classification* is the process to assign objects into a given number of groups, so that objects from the same group share certain similar characteristics. There are many popular methods proposed to fulfill the classification purpose, including naive Bayes Classifiers, decision trees, neural networks, SVM, etc. In this section, SVM, a popular and powerful binary-class classifier, is utilized in the discriminative classification-based method.

Support Vector Machine

Given a set of N pairs of data: $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, in which x represents the feature vector, y , whose value is either -1 or $+1$, is a label indicating the data's assigned group. A separating hyperplane (in Fig. 1) can be defined therein as $\mathbf{x}^T \boldsymbol{\beta} + \beta_0 = 0$, where $\boldsymbol{\beta}$ is a normal vector perpendicular to the hyperplane, and β_0 denotes its offset. Parameters $\boldsymbol{\beta}$ and β_0 need to be chosen properly, so that the margin measured by $1/|\boldsymbol{\beta}|$ between the separating hyperplane and either one of its parallel hyperplanes $\mathbf{x}^T \boldsymbol{\beta} + \beta_0 = \pm 1$ can be as far apart as possible to separate data from two classes (as indicated via hollow circles and solid circles in Fig. 1). For samples from one class with label $y=1$, ideally they are assumed to be above the upper parallel hyperplane $\mathbf{x}^T \boldsymbol{\beta} + \beta_0 = 1$ after classification, hence they satisfy the constraint $\mathbf{x}^T \boldsymbol{\beta} + \beta_0 \geq 1$. It is similar for samples from

the other class with label $y=-1$ having a constraint $\mathbf{x}^T \boldsymbol{\beta} + \beta_0 \leq -1$. Therefore, the two constraints can be summarized together as $y(\mathbf{x}^T \boldsymbol{\beta} + \beta_0) \geq 1$. The basic optimization problem of a SVM classifier can be written as

$$\min_{\boldsymbol{\beta}, \beta_0} \|\boldsymbol{\beta}\|, \quad \text{s.t. } y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq 1, \quad i = 1, 2, \dots, N \quad (6)$$

The above theory is for linear SVM classifiers. For non-linear SVM classifiers, the key point is to apply “*kernel trick*” to the maximum-margin hyperplane. It is the kernel function that makes the original SVM classifier fit the maximum-margin hyperplane in a high-dimensional transformed feature space, in which the resulted separating hyperplane may not be linear in its original feature space anymore and a non-linear SVM classifier is obtained therein. Generally speaking, polynomial kernel $K(x_i, x_j) = (1 + \langle x_i, x_j \rangle)^d$ (d denotes the degree of the polynomial function), Gaussian radius basis kernel $K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{c}\right)$ ($c > 0$ is the Gaussian width), and hyperbolic tangent kernel $K(x_i, x_j) = \tanh(\tau_1 \langle x_i, x_j \rangle + \tau_2)$ (τ_1 and τ_2 are coefficients) are widely utilized as the kernel function in non-linear SVM.

Kernel Learning in SVM

Conducting learning in a non-linear SVM classifier is mainly to tune parameters of the utilized kernel function. There are several methods in literature proposed to fulfill this task [22–28]. In general, there are two main learning strategies. One strategy is *cross-validation*, in which the training set is divided into n subsets of equal size. Sequentially, one subset is tested using a SVM classifier trained on the remaining $(n-1)$ subsets, and parameters are chosen with the highest cross-validation accuracy [22]. The other strategy is based on the maximum margin theory (i.e., the larger the margin, the better classification performance of its constructed hyperplane). Learning algorithms belonging to this category include span bound method [23], [24], Jaakkola–Haussler bound method [25], Oppen–Winther bound method [26], radius-margin bound method [27], support vector count method [28], etc.

Although SVM is popular and widely utilized in many medical image processing applications, it has never been incorporated in NPC segmentation from MRI with automated parameter tuning. In this study, binary-class SVM is chosen as the classifier in the discriminative classification-based method, and radius-margin bound method is implemented for parameters learning [27]. In radius-margin bound method, an upper bound specifying the number of classification errors in a leave-one-out procedure is defined using the margin of the SVM classifier as well as the radius of a sphere, which includes all transformed feature vectors in the high-dimensional feature space. The kernel used in this

Fig. 1 An illustration of support vector machine classifier with a separating hyperplane (solid line) together with its two parallel hyperplanes (dashed lines), as well as samples from two classes in a simple 2-D feature space

segmentation method is the Gaussian radius basis kernel. Hence, parameters (i.e., Gaussian width) in the applied SVM-based segmentation method are also determined algorithmically, not empirically. Training data for the kernel learning is sampled in the same way as depicted in Section “User Interaction”. With the learned result, a SVM classifier is constructed. Pixels in delineated ROIs of MRI images can be classified into tumor and non-tumor groups therein, to accomplish the NPC segmentation task.

Experiments, Statistical Analysis, and Discussion

Data Description and Implementation of the Introduced Segmentation Methods

The study was approved by the institutional review board of National University Hospital regarding the rights and welfare protection of human subjects in research, and informed consent was obtained for all subjects. The proposed segmentation methods have been evaluated using 253 different pairs of MRI slices containing NPC tumor lesion acquired from a 1.5 T MRI scanner (Signa, GE Medical Systems) from 40 patients. The low-level feature vector s in Eq. 1 is composed of normalized intensities from both the T1W (or T2W) image and its corresponding CET1W images with fat suppression. The reason to utilize MRI slice pairs, rather than single MRI slice in this study, is to follow the way in which radiologists manually trace NPC in their clinical practice by utilizing both T1W (or T2W) and CET1W images. The variation among radiologists (i.e., inter-variability) is simulated by creating three different ROIs for each pair of MRI slices. Three training sample sets for each ROI are further generated following the adopted sampling strategy to test the stability of algorithms. Hence, totally 2,277 ($253 \times 3 \times 3$) sets of data from 253 pairs of tumor-bearing MRI slices are generated. In this way, a total of 2,277 segmentation results can be obtained from each segmentation method for experimental evaluation of NPC segmentation performance.

For the proposed clustering-based segmentation method, parameters A and σ_p^2 are set as an identity matrix and 0.1, respectively, as initializations in Table 1. For both the proposed classification- and clustering-based methods, totally 90 pixels are sampled to construct the training set (input S in Table 1) in consideration of both the efficiency of implementation in clinical practice and the effectiveness of tumor segmentation methods. Among them, 45 samples are positive training samples taken within ROI while the other 45 samples are negative training samples taken outside of ROI. For the negative training samples, five stratum of candidates are utilized in the applied stratified random sampling ($L=5$ in Eq. 3).

Segmentation Methods for Comparison

To demonstrate the segmentation capability of the two introduced methods, two other similar region-based segmentation methods are also implemented for performance comparison, including *support vector data description* (denoted as “SVDD”) [29] of one-class classification, and a baseline method in which the same steps as the proposed clustering-based method are performed but without the learning phase (denoted as “Baseline”). For Baseline method, parameters A and σ_p^2 are set as the same values as their initializations used in the proposed clustering-based method. The reason to include Baseline in comparison here is to demonstrate the learning effect of the proposed methods. For SVDD, it is a type of one-class classification methods. The main idea of SVDD is to construct a spherically shaped decision boundary via a set of *support vectors* [28] around objects belonging to one class, thus the classification purpose can be fulfilled [30–32]. Its sphere is achieved from the following optimization problem:

$$E(R, \alpha, \varepsilon) = R^2 + C \sum_i \varepsilon_i, \quad \text{s.t.} \|x_i - \alpha\|^2 \leq R^2 + \varepsilon_i, \quad \varepsilon_i \geq 0, \forall i \quad (7)$$

where a sphere with radius R locates at center α ; ε ($\varepsilon_i \geq 0, \forall i$) is a slack variable and C is a parameter controlling the trade-off between the volume of the sphere and description errors. In this study, Gaussian radius basis kernel is also used as the kernel function. To our best knowledge, there is no existing learning method reported for SVDD’s parameters learning, thus an empirical setting of Gaussian width 0.5 and a rejection ratio 0.1 is used for the optimal tumor segmentation performance in its implementation.

Evaluation Criteria

After obtaining NPC segmentation results by various methods, they are quantitatively evaluated based on two criteria as well as being compared with NPC ground truth manually delineated by our senior radiologists by consensus. The two criteria are positive predictive value (PPV) and sensitivity (SEN) [33]. PPV is used to evaluate the proportion of correctly extracted NPC regions. The definition of PPV is as $PPV = TP / (TP + FP)$, where TP is the number of *true positives* (i.e., tumor points correctly detected) and FP is the number of *false positives* (i.e., non-tumor points wrongly determined as tumor points). For SEN, it is defined as $SEN = TP / GT$, where GT is the ground truth of NPC (i.e., NPC tumor region delineated by radiologists). The reason to take both criteria into consideration is that either of the two measures can be biased under certain situations. An illustration is demonstrated in Fig. 2. In this figure, regions enclosed via dashed lines denote segmented results, while regions enclosed by solid lines represent its corresponding ground truth. PPV

Fig. 2 An illustration of biased situations for PPV (*left*) and SEN (*right*)

can be biased by the situation of under-segmentation (left in Fig. 2), in which the segmentation result is only a tiny portion of the whole NPC region. Unfortunately, PPV value calculated in this situation is high (nearly 1, which is caused by low TPs and $FPs \approx 0$). On the other hand, SEN can be biased in the situation of over-segmentation (right in Fig. 2), in which TPs is large due to an overlapping of the NPC ground truth with imprecisely segmented regions. The SEN value calculated for this situation is also high (nearly 1 caused by $TPs \approx GT$). Thus, in order to have a balanced assessment of segmentation results, both criteria

are adopted for the quantitative evaluation and analysis in this study. Segmentation methods with both high PPV and high SEN outcomes are superior.

Experimental Results and Statistical Analysis

Experimental Results All four segmentation methods are applied in segmenting tumor lesion based on 2,277 sets of data, and NPC segmentation results from some examples are shown in Fig. 3. All NPC segmentation results have been post-processed by *morphological opening* and *closing* operations with a disk-shaped structuring element of radius 1 to remove isolated points, holes and thin line structures in identified NPC regions [34]. It can be observed that both the SVM-based and the clustering-based segmentation methods achieve more similar results towards the ground truth compared with the other methods. The box-and-whisker plots in Fig. 4, which are drawn based on calculated PPV and SEN results from all segmentation results, also substantiate this observation. In each box, a horizontal line is drawn across the box at the *median*; the *upper* and *lower quartiles* are lines above and below it. A vertical dashed line is drawn up from the upper quartile and down from the lower quartile to their most

Fig. 3 NPC segmentation results from six pairs of MRI slices by all compared methods. (*Rows: first*—T1W (or T2W) MRI slices, *second*—CET1W MRI slices with fat suppression, *third*—NPC lesion ground truth manually traced by senior radiologists, *fourth*—NPC segmentation results by SVDD-based method, *fifth*—NPC segmentation results by SVM-based method, *sixth*—NPC segmentation results by Baseline; *seventh*—NPC segmentation results by the clustering-based method)

Fig. 4 **a** Box-and-whisker plot of PPV among all four compared methods. **b** Box-and-whisker plot of SEN among all four compared methods

extreme data points, which are within a distance of 1.5 interquartile range [20]. Each data point beyond ends of the vertical line is marked via a plus sign. For SVDD, it tends to extract a small portion of the NPC ground truth, since it only uses positive samples with high intensities. The small partial segmentation outcomes compared with NPC ground truth result in low TPs and FPs ≈ 0 , which can bring about large PPV and low SEN. This conclusion can be substantiated by the high whisker of PPV box of SVDD in Fig. 4a and the corresponding low whisker of SEN box of SVDD in Fig. 4b.

For Baseline, ranges between upper quartiles and lower quartiles of its two boxes in Fig. 4 are larger than corresponding ones of others, suggesting that pre-defined parameters of A and σ_p^2 may work well for some cases, but not for all. Hence, this method is not as stable and robust as the proposed methods with learning for diverse cases. Hence, the necessity of learning can be demonstrated. For SVM-based and clustering-based methods, it is not easy to conclude from Fig. 4 solely about which method is better. Thus, a comprehensive statistical test is applied to further verify whether the two methods are really comparable, from the statistical point of view.

Statistical Analysis A statistical test made up of one-way analysis of variance (ANOVA) followed by post-hoc *multiple comparison tests* [20] is utilized for statistical analysis.

In one-way ANOVA, PPV and SEN results from all methods are compared to test a hypothesis (H_0) that, PPV/SEN means of various methods are equivalent, against the general alternative that these means cannot be all the same. P value is used here as an indicator to reveal whether H_0 holds or not. In this study, P values calculated from PPV and SEN results from all methods are nearly 0, which is a strong indication that all methods cannot share the same PPV or SEN means. Therefore, the next step is to conduct more detailed paired comparison. The reason to do so is because the alternative against H_0 is too general.

Information about which method is superior cannot be perceived by one-way ANOVA alone. Therefore, tests that can provide such information are needed and they are multiple comparison tests.

There are two kinds of evaluation results after applying multiple comparison tests on PPV and SEN outcomes between SVM-based and clustering-based methods. One is the estimated mean difference, which is a *single-value estimator* of PPV/SEN mean difference between the two compared methods. Another is a 95 % confidence interval (CI), which is a special form of *interval estimator* for a parameter (i.e., PPV/SEN mean difference in our study). Generally speaking, instead of estimating the parameter by a single value, CI provides an estimated range, which is likely to include the estimated parameter. To be specific, for PPV comparison, the clustering-based method is 0.1061 higher than SVM, and the PPV mean difference is likely to fall within a 95 % CI between 0.0946 and 0.1175. Since the upper and lower bounds of the CI are both positive (resulting from clustering-SVM), it gives a strong indication (>95 %) that the clustering-based method is superior to the SVM-based method in terms of PPV from statistical point of view. For SEN comparison, the analysis is similar. The clustering-based method is 0.0601 lower than the SVM-based method. The SEN mean difference is likely to fall within a 95 % CI between -0.0748 and -0.0453 . Since the upper and lower bounds of the CI are both negative (resulting from clustering-SVM), it gives a strong indication (>95 %) that the SVM-based method is superior to the clustering-based method in terms of SEN from statistical point of view. To sum up, based on the above statistical analysis, the clustering-based method achieves comparable performance with the SVM-based method based on segmentation results from the 2,277 sets of data in terms of both PPV and SEN measures from the statistical point of view.

Fig. 5 NPC segmentation results after applying different sampling strategies for negative training data in six pairs (columns) of cases. (Rows: *first*—T1W (or T2W) MRI slices, *second*—NPC lesion ground truth, *third*—ROI for each case, *fourth*, *sixth*, *eighth*—scatter plots of negative samples by applying a bi-ased random sampling strategy, the random sampling strategy, and the stratified sampling strategy, respectively, *fifth*, *seventh*, *ninth*—corresponding NPC segmentation results after adopting a biased sampling strategy, the random sampling strategy, and the stratified sampling strategy in the proposed clustering-based method, respectively)

Discussion

In Section “[User Interaction](#)”, stratified random sampling strategy is introduced when sampling negative training pixels, in order to have more representative training data for the learning procedure. In this section, to demonstrate the

effectiveness of the applied stratified sampling strategy, it is compared with two other sampling strategies, i.e., a *bi-ased random sampling strategy* and the *random sampling strategy*. For biased random sampling strategy, negative samples are randomly chosen from pixels with large values of intensity outside ROI (from ones which are almost as

Table 2 PPV and SEN OF NPC segmentation results obtained by the proposed clustering-based method with different sampling strategies for negative samples

Sampling strategy	Measures	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Biased random sampling	PPV	0.9789	0.9469	0.9325	0.8525	0.8408	0.9075
	SEN	0.5562	0.7665	0.4872	0.9623	0.9738	0.9671
Random sampling	PPV	0.8806	0.8947	0.8730	0.9570	0.7216	0.8318
	SEN	0.7877	0.7899	0.8026	0.9546	0.9938	0.9824
Stratified random sampling	PPV	0.9345	0.8799	0.8079	0.8525	0.8408	0.8845
	SEN	0.8550	0.9697	0.9776	0.9632	0.9738	0.9711

bright as the NPC lesion, after patients are injected with contrast agents for tumor lesion enhancement), and it reflects an extreme case of under-representing negative samples (i.e., only negative samples with high intensities are sampled). Scatter plots of negative samples obtained by this strategy in several cases are shown in the fourth row of Fig. 5. For random sampling strategy, negative samples are chosen randomly outside ROI. Scatter plots of negative samples obtained by this strategy in several cases are shown in the sixth row of Fig. 5.

Several examples of segmentation results obtained by the proposed clustering-based segmentation methods with different sampling strategies for negative samples are illustrated in Fig. 5. It can be observed that NPC segmentation results obtained by incorporating stratified random sampling strategy (the ninth row) are often more similar than ones obtained by adopting biased random sampling strategy (the fifth row) and random sampling strategy (the seventh row), compared with their corresponding NPC ground truth (the second row). For negative samples obtained using biased random sampling strategy (the fifth row), regions with lower intensities are totally overlooked. The corresponding segmentation results are often partial portions of the ground truth (e.g., cases 1–3), which is similar as ones of SVDD since only bright pixels are utilized in its learning. This observation can be quantitatively measured by their high PPV and low SEN results, compared with ones obtained by adopting other sampling strategies in Table 2. For results achieved via random sampling strategy (seventh row), they are not stable. The overall performance heavily depends on whether the chosen negative samples can reflect characteristics of all negative data well. The comparison between PPV/SEN results in Table 2 also suggests that stratified random sampling strategy can often provide more favorable segmentation results.

Conclusion

In this paper, two region-based segmentation methods conducting NPC tumor extraction from MRI with learning are introduced. One method is a newly proposed generative clustering-based method, which is made up of two steps. For the supervised step, a spatially weighted metric-based similarity is proposed. Its parameters are learned via a supervised spectral clustering algorithm. For the unsupervised step, an unsupervised out-of-sample extension strategy is incorporated to finish the whole tumor segmentation. The other method is a discriminative classification-based method, in which non-linear SVM is incorporated as the classifier, and a radius-margin bound method is applied to learn the kernel's parameters. A large number of experiments have been conducted to evaluate the NPC segmentation performance. The superiority of adopting learning in region-based NPC tumor segmentation methods is

demonstrated, and the two methods achieves comparable segmentation performance from the statistical point of view.

The way that the segmentation performed in this study is also different from many existing work in which it requires the clinician to delineate a region of interest enclosing the tumor region. This human interaction is not too demanding towards clinicians as the accuracy of delineated lines is not important and yet it facilitates the learning procedure in the introduced methods, and achieves more stable and accurate segmentation. The main contribution of this study also includes taking automatic parameters learning into consideration when proposing NPC tumor segmentation methods for real clinical practice. For the clustering-based method, the technical novelty lies in the aspects of proposing a new spatial weighted metric-based similarity and a new way to learn its parameters via spectral clustering. The limitation of the current study includes that clinicians need to draw separated ROI on multiple tumor-bearing slices to realize tumor segmentation for one patient, if tumor is illustrated on several slices. However, it is already much easier for clinicians to draw multiple ROIs, compared with delineating tumor boundaries carefully on numerous slices. We will further try to alleviate the computer–user interaction workload in our future study.

References

1. Chang E, Adami H: The enigmatic epidemiology of nasopharyngeal carcinoma. *Canc Epidemiol Biomarkers Prev* 15:1765–1777, 2006
2. Parkin D, Bray F, Ferlay J, Pisani P: Global cancer statistics. *CA Cancer J Clin* 55:74–108, 2005
3. Chong V, Fan V, Khoo J: Nasopharyngeal carcinoma with intracranial spread: CT and MR characteristics. *J Comput Assist Tomogr* 20(4):563–569, 1996
4. Chua D, Sham J, Kwong D: Volumetric analysis of tumor extent in nasopharyngeal carcinoma and correlation with treatment outcome. *Int J Radiat Oncol Biol Phys* 39(3):711–719, 1997
5. Chen M, Chen T, Liu J, Chang C, Chie W: Better prediction of prognosis for patients with nasopharyngeal carcinoma using primary tumor volume. *Cancer* 100(10):2160–2166, 2004
6. Lee N, Xia P, Quivey J, Sultanem K, Poon I, Akazawa C, Akazawa P, Weinberg V, Fu K: Intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: an update of the UCSF experience. *Int J Radiat Oncol Biol Phys* 53(1):12–22, 2002
7. Hsu C, Lai P, Lee C, Huang W: Automated nasopharyngeal carcinoma detection with dynamic gadolinium-enhanced MR imaging. *Methods Inform Med* 40(4):331–337, 2001
8. Zhou J, Lim T, Chong V, Huang J: Segmentation and visualization of nasopharyngeal carcinoma. *Comput Biol Med* 33(5):407–424, 2003
9. Lee F, Yeung D, King A, Leung S, Ahuja A: Segmentation of nasopharyngeal carcinoma lesions in MR images. *Int J Radiat Oncol Biol Phys* 61(2):608–620, 2005
10. Li C, Kao C, Gore J, Ding Z: Implicit active contours driven by local binary fitting energy. *IEEE Conf Comput Vis Pattern Recognit* 1:1–7, 2007
11. Xu C, Prince J: Snake, shape, and gradient vector flow. *IEEE Trans Image Process* 7:359–369, 1998
12. Bullmore E, Brammer M, Rouleau G, Everitt B, Simmons A: Computerized brain tissue classification of magnetic resonance

- images: a new approach to the problem of partial volume artifact. *NeuroImage* 2:133–147, 1995
13. Kapur T, Grimson W, Kikinis R, Wells W: Enhanced spatial priors for segmentation of magnetic resonance imagery, *International Conference on Medical Image Computing and Computer Assisted Intervention*. LNCS 1496:457–468, 1998
 14. Li C, Xu C, Gui C, Fox M: Level set evolution without re-initialization: a new variational formulation. *IEEE Conf Comput Vis Pattern Recognit* 1:430–436, 2005
 15. Osher S, Paragios N: *Geometric level set methods in imaging, vision, and graphics*. Springer, Heidelberg, 2003
 16. Shi J, Malik J: Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Mach Intell* 22:888–905, 2000
 17. Ng A, Jordan M, Weiss Y: “On spectral clustering: analysis and an algorithm,” *Advances in neural information processing systems*, pp. 64–72, 2002
 18. Boykov Y, Veksler O, Zabih R: Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Anal Mach Intell* 23:1222–1239, 2001
 19. Bach F, Jordan M: Learning spectral clustering, with application to speech separation. *J Mach Learn Res* 7:1963–2001, 2006
 20. Rice J: *Mathematical statistics and data analysis*, 2nd edition, Duxbury Press, Belmont, CA, 2007
 21. Bengio Y, Païement J, Vincent P, Delalleau O, Roux N, Ouimet M: “Out-of-sample Extension for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering,” *Advances in Neural Information Processing Systems*, pp. 857–863, 2003
 22. Hsu C, Chang C, Lin C: A practical guide to support vector classification, <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, 2010
 23. Chapelle O, Vapnik V: “Model selection for support vector machines,” *Advances in neural information processing systems*, pp. 230–236, 1999
 24. Vapnik V, Chapelle O: “Bounds on error expectation for support vector machines,” *Neural Comput* 12(9):2013–2036, 2000
 25. Jaakkola T, Haussler D: “Probabilistic kernel regression methods,” *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics*, pp. 1–9, 1999
 26. Oppé M, Winther O: “Gaussian processes and SVM: Mean field and leave-one-out,” *Advances in large margin classifier*, pp. 311–326, 2000
 27. Vapnik V: *Statistical learning theory*, Wiley, 1998
 28. Vapnik V: *The nature of statistical learning theory*, Springer press, New York, 1995
 29. Tax D, Duin R: Support vector data description. *Mach Learn* 54:45–66, 2004
 30. Tax D, Duin R: Support vector domain description. *Pattern Recognit Lett* 20:1191–1199, 1999
 31. Tax D, Duin R: “Data domain description using support vectors,” *Proceedings of the European Symposium on Artificial Neural Networks*, pp. 251–257, 1999
 32. Tax D: “One-class classification; Concept-learning in the absence of counter-examples,” *Ph.D thesis, Delft University of Technology*, 2001
 33. Gunnarsson R, Lanke J: The predictive value of microbiologic diagnostic tests if asymptomatic carriers are present. *Stat Med* 21:1773–1785, 2002
 34. Gonzalez R, Wintz P, Woods R: *Digital image processing*, 2nd edition, Prentice Hall Press, New Jersey, 2002