

Semi-supervised Nasopharyngeal Carcinoma Lesion Extraction from Magnetic Resonance Images Using Online Spectral Clustering with a Learned Metric

Wei Huang¹, Kap Luk Chan¹, Yan Gao¹, Jiayin Zhou², and Vincent Chong²

¹ School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore

² School of Medicine, National University of Singapore, Singapore

Abstract. In this paper, we consider the extraction of nasopharyngeal carcinoma lesion from MR images as a region segmentation problem. We propose a semi-supervised segmentation approach to segment the lesion in two steps. First, a metric is learned in a supervised fashion, which maximizes the separation between two groups of pixels (tumor or non-tumor) with minimal user interaction. Second, the learned metric is used to complete extraction of tumor region in an unsupervised fashion. Several experiments were conducted to evaluate the performance of similar methods with learned metrics for grouping or classifying pixels to form the tumor region. It is observed that the spectral clustering-based method performs well and the performance is comparable or marginally better than the discriminative SVM-based method.

1 Introduction

Nasopharyngeal carcinoma (NPC) lesion is an oral cancer developed in nasopharynx, which is different from other tumors in its occurrence, cause and clinical behavior [1]. Extracting NPC lesion region from Magnetic Resonance (MR) images for volumetric analysis is important to its treatment planning and prognosis. In order to help clinicians reduce the heavy workload of inspecting numerous MR images and delineating lesion manually, various segmentation methods have been proposed during the last decade or so [2], [3], [4], [5]. Most MR images segmentation methods can be categorized into *contour-based methods* and *region-based methods*. In contour-based methods [2], [3], active contours have been popularly used to evolve an deformable curve to the boundary of tumor. The evolving curve is driven by internal or external forces and additional constraints depending on characteristics of MR images. Although these methods require minimal user interaction to draw an initial contour inside or outside tumor regions, they require users to specify and adjust numerous parameters. For example, in the method proposed by Li et al. [3], there are totally 6 parameters with respect to various energy functional terms, constraints, etc. These parameters need to be

adjusted properly for different MR images in order to ensure correct convergence of the evolving curves.

Region-based methods consider the tumor segmentation problem as a region segmentation problem. In order to discern tumor region from other non-tumor tissue regions, image pixels are classified into different tissue classes or clustered into different groups according to certain similarity criteria [4],[5]. Finding a proper metric encoding the notion of similarity is often of great importance to these methods. Since MR images data often have large varying statistical properties across scans, either making an assumption about the metrics beforehand or learning metrics based on training images and applying it to other MR scans may not match the characteristics of the two sets of images well. Therefore, it is necessary to learn different metrics for different MR images.

In this paper, we focus on learning of a metric that can be used for separating tumor regions from non-tumor regions in MR images based on *spectral clustering* [6],[7] in a semi-supervised fashion. The paper is organized as follows. In section 2, we learn a metric to measure the pairwise similarity of pixels by a supervised spectral clustering algorithm. In section 3, we elaborate how to make use of the out-of-sample extension algorithm to complete MR image segmentation. In section 4, we report several experiments and demonstrate the performance of our method and compare it with results of other similar methods. In section 5, we give the conclusion of this work.

2 Metrics Learning for NPC Lesion Extraction

2.1 A Spatially Weighted Metric

We construct a metric that reflects the pairwise similarity $d(\mathbf{x}_i, \mathbf{x}_j)$ between i -th and j -th pixel in a MR image as follows.

$$\begin{aligned} d(\mathbf{x}_i, \mathbf{x}_j) &= K(\mathbf{x}_i, \mathbf{x}_j) \cdot G(\mathbf{x}_i, \mathbf{x}_j) \\ &= \exp\left(-\frac{\|p_i - p_j\|^2}{\sigma_p^2}\right) \cdot \exp\left(-(s_i - s_j)^T \mathbf{A} (s_i - s_j)\right) \end{aligned} \quad (1)$$

where, σ_p is a scalar and \mathbf{A} is a full matrix; p_i is the spatial location of i -th pixel, and s_i is the feature vector based on low-level image features of i -th pixel. Obviously, the term $K(\mathbf{x}_i, \mathbf{x}_j)$ of Eq. (1) is a Gaussian RBF (Radial Basis Function) with a spatial localization emphasis that decreases with the increase of the distance between i -th and j -th pixel increases. Therefore, two spatially nearby pixels will have more influence on the measure of similarity than two pixels that are far apart. Our metric is different from those used in previous related works. For *N-cuts* [8], Shi et al. did not include the explicit form of the term \mathbf{K} in their metric and assumed the matrix \mathbf{A} as a diagonal matrix, ignoring the correlation among features. For *Ng-Jordan-Weiss* [6], they only utilized the term \mathbf{K} in their metric, but p_i is a vector in a d -dimensional space.

To ensure Eq.(1) be a metric, it should meet the requirements of non-negativity and triangle inequality. Therefore, the matrix \mathbf{A} must be positive semi-definite,

Table 1. Algorithm of metric learning with spectral clustering

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

Algorithm:

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

$$\min J(\mathbf{A}, \sigma_p^2) = \frac{1}{2} \left\| \mathbf{X}\mathbf{X}^T - \mathbf{X}_{part}\mathbf{X}_{part}^T \right\|_F^2 \quad (2)$$

$$s.t. \mathbf{A} \succeq \mathbf{0}$$

where, $\mathbf{X}_{part} = \mathbf{C}^{-1/2} \mathbf{E}(\mathbf{E}^T \mathbf{C} \mathbf{E})^{-1/2} \mathbf{B}$. \mathbf{E} is an indicator matrix of data set partitions and \mathbf{B} is an arbitrary orthogonal matrix. Applying $\mathbf{A} \succeq \mathbf{0}$ as a constraint to the optimization problem and find optimal solutions of \mathbf{A} and σ_p^2 through gradient descent algorithm.

Output: Learned parameters \mathbf{A} and σ_p^2 of the metric.

i.e. $\mathbf{A} \succeq \mathbf{0}$. The aim of metric learning is to find proper parameters of the metric, so that data from different groups can be well separated. The task of learning such a metric involves finding proper \mathbf{A} and σ_p^2 in our case. Hence, parameters in the metric is determined algorithmically, not empirically.

2.2 Metric Learning with Spectral Clustering

Spectral clustering [6] is often considered as an approximate solution to the *graph-cut* problem [9], which achieves partitioning of images by cutting weak links between graph nodes to separate images into various segments. Our metric learning method is performed with a spectral clustering algorithm [6]. The procedure of our metric learning algorithm based on that in [7] is presented in Table 1.

The original metric learning algorithm with spectral clustering is for the Gaussian width σ^2 of a simple Gaussian RBF metric in [7]. Since the metric in Eq.(1) is more complicated, simply applying the learning algorithm in [7] is unreasonable, and \mathbf{A} in Eq.(1) has its own constraint to be satisfied, viz. $\mathbf{A} \succeq \mathbf{0}$. Therefore, we take this constraint into consideration when learning parameters of the metric via a minimization of function in Eq.(2). Through a gradient decent algorithm with constraint in Eq.(2), optimum solutions for parameters in Eq.(1) are found. Hence, a learned metric is obtained.

2.3 User Interaction

In this work, we allow clinicians to draw their own *Region of Interests* (ROI), in which they assume NPC lesion will 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. The reason for us to do so is to utilize the prior knowledge about NPC lesion used by clinicians. By using such a ROI, we can also avoid the influence of tissues with numerous capillary vessels (e.g. nasal cavity), which share similar visual properties as NPC lesion regions after patients are injected with contrast agents for lesion enhancement. Points inside the enclosed region are sampled as positive training samples and points outside are sampled as negative training samples for constructing input S in Table 1. In our implementations, positive samples often share similar intensities and are chosen from the center area of ROI. However, negative samples are usually highly variable in intensity and more difficult to sample. A simple random sampling strategy can result in under-representation of some sub-populations (stratums) of negative points, making negative samples less representative. It will bias our learning and deteriorate the final segmentation results. Therefore, we apply *stratified random sampling* strategy [10] in sampling negative points. The optimal stratum allocation scheme used is *Neyman allocation* [10], which allocates relatively large number of negative samples if one stratum contains a larger fraction of points with largely varying intensities.

3 Tumor Segmentation through Unsupervised Out-of-Sample Extension

The second step of our method is to perform MR image segmentation by spectral clustering on pixels other than those sampled for metric learning in the previous step. The main computational burden of spectral clustering is from the eigen-decomposition of the matrix \mathbf{L} of size $n \times n$, where n is the total number of pixels in a MR image. In the previous metric learning step, this is not a problem as the number of points in the labeled training sample set S are still small. If spectral clustering is used to group all points in MR images, the size of matrix \mathbf{L} will become extremely large and causes problems in eigen-decomposition. For this reason, we apply the *out-of-sample extension method* [11] in our problem. The aim of this method is to map points into the spectral domain directly using a mapping function. Therefore, the main computational burden of eigen-decomposition can be avoided.

For remaining points, the similarity is calculated with respect to each sample in the training set S via the learned metric. In our work, since we use prior knowledge to enclose the NPC lesion region inside ROI, the points outside the ROI should not be considered. Hence, we have a similarity matrix \mathbf{D} of size (number of remaining samples in ROI) \times (number of samples in S). Although the size of this matrix is smaller than if considering all remaining points in MR images, it is still not practical to apply eigen-decomposition directly. After normalization, the affinity matrix \mathbf{D} is given by

$$\mathbf{D}_{ij} = \frac{\mathbf{D}_{ij}}{\sqrt{\sum_{x \in S} \mathbf{D}_{ix} \sum_{x' \in S} \mathbf{D}_{jx'}}} \quad (3)$$

The spectral embedding of remaining points in ROI can be computed from

$$\mathbf{X}_{mapped} = \mathbf{D} \cdot \mathbf{X} \cdot \mathbf{V}^{-1} \quad (4)$$

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}_{mapped} , we can use some clustering algorithms, such as the *K-means* algorithm ($k = 2$, for tumor or non-tumor group), to extract the NPC lesion region unsupervisedly therein.

4 Experiments and Discussion

Our method has been evaluated with 6 pairs of MR images from 6 patients, which are acquired from a 1.5T MR scanner (Signa, GE Medical Systems) from different patients. The feature vector s in Eq. (1) is composed of normalized intensities from T1 Weighted (T1W) (or T2 Weighted T2W) images and their corresponding Contrast Enhanced T1 Weighted (CET1W) images with fat suppression. The reason to do so is to follow the way in which clinicians manually trace NPC lesion in their practice by utilizing both T1W (or T2W) and CET1W images. We simulate the variation among radiologists by creating 4 slightly different ROIs in each pair of MR scans (from a patient) for a total of 6 patients. We further generated 4 sample sets for each ROI to test algorithm stability. Hence, we have 96 sets of training data sampled from all the MR images, and we ran 96 times of the methods to obtain a total of 96 segmentation results for each method.

To demonstrate the performance of our method over other existing similar region-based methods, we conducted several experiments and set up a statistical test for quantitative analysis of results. Besides our method, there are three other methods for comparison, including binary-class *Support Vector Machine* (SVM), *Support Vector Data Description* (SVDD) [12] of one-class SVM, and a baseline method in which one performs the same steps as in our method but without metric learning (Baseline). For SVM, we utilized Gaussian RBF kernel and *radius-margin bound* method [13] for its parameter learning. For baseline method, \mathbf{A} and σ_p are set as an identity matrix and 0.1, respectively. For SVDD, a Gaussian RBF kernel with width 0.5 and a 0.1 rejection ratio are applied.

We evaluated the NPC extraction results based on two criteria with respect to NPC ground truth manually delineated by our radiologist, and they are *Positive Predictive Value* (PPV) [14] and *sensitivity*. PPV is used to evaluate the proportion of correctly extracted NPC regions. The definition of PPV is $PPV = \frac{TP_s}{TP_s + FP_s}$ [14], where TP_s is the number of *True Positives* (tumor points correctly detected) and FP_s is the number of *False Positives* (non-tumor points wrongly detected as tumor points). Sensitivity is defined as $sensitivity = \frac{TP_s}{RS}$, in which RS is the NPC ground truth. PPV can be biased by under-segmentation, in which the segmentation result is a tiny portion of NPC caused by small TP_s and $FP_s \simeq 0$.

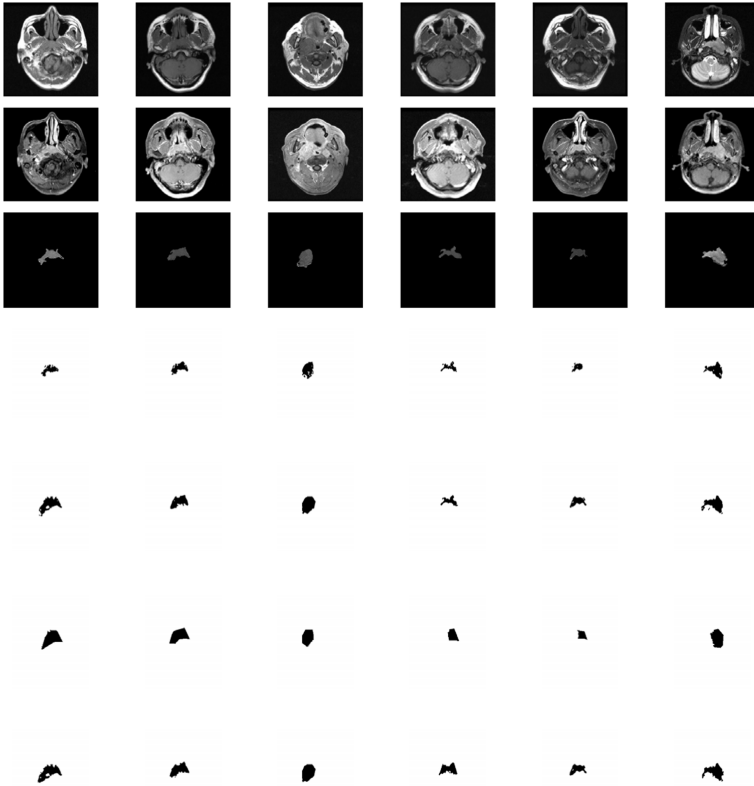


Fig. 1. (Rows from up to down) T1W (or T2W) MR images, CET1W MR images with fat suppression, NPC lesion ground truth manually traced by our radiologist, NPC extraction results by SVDD, NPC extraction results by SVM, NPC extraction results by Base line, NPC extraction results by our method.

Sensitivity can also be biased by over-segmentation, in which TPs is large due to overlapping of NPC ground truth with large segmented region. Thus, in order to have a more objective assessment, we adopt both measures for quantitative analysis. Some NPC extraction results based on the same training data and ROIs are shown in Figure 1. All extraction 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. We can easily see that SVM and our method achieve the most similar results compared to the ground truth. For SVDD, it tends to extract a small portion of the ground truth, since it only uses positive samples. The high whisker of SVDD PPV box in Figure 2 and the low whisker of SVDD sensitivity box in Figure 3 also substantiate this observation. For the baseline method, the ranges between upper quartiles and lower quartiles of both PPV and sensitivity boxes are large, which shows that predefined parameters work well for some sets, but not for others. For SVM and our method, a detailed statistical test is applied to further discern which

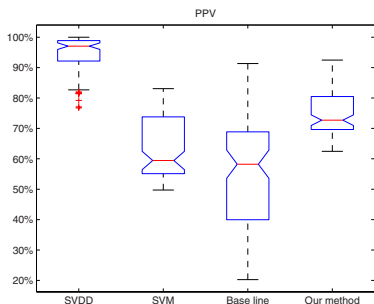


Fig. 2. PPV of four methods

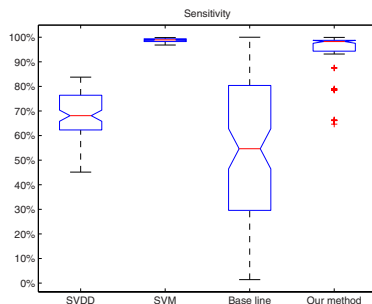


Fig. 3. Sensitivity of four methods

Table 2. Multiple comparison test results of PPV and sensitivity of SVM and our method

	Method 1	Method 2	Mean diff	Estimated mean diff	95% Confidence int.
PPV	SVM	Our method	-14.97%	-10.64%	-6.31%
Sensitivity	SVM	Our method	-2.22%	3.81%	9.83%

method has a superior performance over another from statistical point of view. We design a test of *analysis of variance* (ANOVA) followed by post-hoc *multiple comparison test* [10]. After performing two one-way ANOVA tests for results of PPV and sensitivity from SVM and our method respectively, the *p-values* are 0, which casts serious doubt on the null hypothesis that SVM and our method are with the same PPV and sensitivity means. Therefore, two more post-hoc multiple comparison tests are applied to test which method has superior statistical PPV and sensitivity means. Results of multiple comparison test for comparisons between the two methods are in listed in Table 2. The last three columns of the table indicate the mean difference being compared (mean of Method 1 minus mean of Method 2), the estimated mean difference and a 95% confidence interval for the mean difference, respectively. From the entries in the table, we can see that, the estimated mean of our method is 10.64% higher than SVM for PPV measure, but 3.81% lower than SVM for sensitivity measure. For the comparison of two methods by PPV, a 95% confidence interval for the mean difference is $[-14.97\%, -6.31\%]$. This result gives a strong indication that for over 95% sets of data, our method achieve better results than SVM based on PPV measure from statistical viewpoint. For the comparison of two methods by sensitivity, a 95% confidence interval for the mean difference is $[-2.22\%, 9.83\%]$. From this we can see that, SVM performs better than our method for some sets of data, but not for others. Therefore, based on the conclusions derived from statistical analysis of the two measures, our method can achieve comparable or marginally better performance than SVM for the majority of data from the statistical perspective with due consideration of non-linearity in the PPV measure.

5 Conclusion

In this paper, a semi-supervised spectral clustering method based on metric learning is proposed for the extraction of nasopharyngeal carcinoma lesion from MR images. We conducted several experiments to evaluate our methods and compared it with other similar methods. Experiments results show the superiority of adopting metric learning in our method. Also, our method achieves comparable or marginally better performance than discriminative SVM-based method according to our statistical analysis.

References

1. Chang, E., Adami, H.: The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiology Biomarkers and Prevention* 15, 1765–1777 (2006)
2. Xu, C., Prince, J.: Snake, shape, and gradient vector flow. *IEEE TIP* 7, 359–369 (1998)
3. Li, C., Kao, C., Gore, J., Ding, Z.: Implicit active contours driven by local binary fitting energy. In: *IEEE CVPR*, pp. 1–7 (2007)
4. 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)
5. Kapur, T., Grimson, W., Kikinis, R., Wells, W.: Enhanced spatial priors for segmentation of magnetic resonance imagery. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 457–468. Springer, Heidelberg (1998)
6. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: *NIPS*, pp. 64–72 (2002)
7. Bach, F., Jordan, M.: Learning spectral clustering, with application to speech separation. *JMLR* 7, 1963–2001 (2006)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE TPAMI* 22, 888–905 (2000)
9. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE TPAMI* 23, 1222–1239 (2001)
10. Rice, J.: *Mathematical Statistics and Data Analysis*, 2nd edn. (2007)
11. Bengio, Y., Paiement, J., Vincent, P., Delalleau, O., Roux, N., Ouimet, M.: Out-of-sample extension for lle, isomap, mds, eigenmaps, and spectral clustering. In: *NIPS*, pp. 857–863 (2003)
12. Tax, D., Duin, R.: Support vector data description. *Machine learning* 54, 45–66 (2004)
13. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing multiple parameters for support vector machines. *Machine learning* 46, 131–159 (2002)
14. Gunnarsson, R., Lanke, J.: The predictive value of microbiologic diagnostic tests if asymptomatic carriers are present. *Statistics in medicine* 21, 1773–1785 (2002)