Learning A Multi-size Patch-based Hybrid Kernel Machine Ensemble for Abnormal Region Detection in Colonoscopic Images

Peng Li[†], Kap Luk Chan[‡] and S. M. Krishnan
Biomedical Engineering Research Center, Nanyang Technological University
Research Techno Plaza, 50 Nanyang Drive, XFrontiers Block, 6th Storey, 637553, Singapore
[†] lipeng@pmail.ntu.edu.sg, [‡] eklchan@ntu.edu.sg

Abstract

When detecting abnormalities in colonoscopic images, the location, shape and size of the abnormal regions in the image are unknown and vary across images. It is difficult to determine the appropriate patch-size for patch-based approach. So multi-size patches are used simultaneously to represent the image regions and an ensemble is constructed in which each classifier handles one patch size. The combination of classifiers trained using multiple-size patches can recognize abnormal regions more effectively than only using single-size patches. The classification of the image patches can be performed using a discriminative binary Support Vector Machine (SVM) or a recognition-based one-class SVM. Integration of the two types of SVMs is expected to further improve abnormal region detection. Experimental results show the good performance of our proposed ensemble.

1. Introduction

Colonoscopy is a minimal invasive procedure for screening the colon and rectum for early signs of colorectal cancer or other diseases, and diagnosing the causes of unexplained changes in the bowel. The analysis of colonoscopic images for clinical diagnosis of abnormalities relies on the expertise of medical experts, which need years of training to acquire. It is thus meaningful to develop a computer-assisted technique to help the process for screening of these potentially fatal diseases. Previous research on colonoscopic image analysis focuses on discriminating tumors from normal tissues. Wang *et al* employed a self-organizing map for segmenting colonoscopic images [13]. Maroulis *et al* developed a detection system for colorectal lesions in endoscopic video using multi-layer neural network [7]. Recently Karkanis *et al* employed linear discriminant analysis to detect tumors in endoscopic video [3]. However, few work have been done to discriminate normal tissues from different kinds of abnormalities in colonoscopic images, which is more significant for screening purpose. In fact, many categories of abnormalities can be seen in colonoscopic images, such as polyps, tumors, inflammation, bleeding, ulceration and diverticula etc.(Figure 1) and they show large variations. The abnormal regions usually do not occupy the whole image and vary in location, color, size and shape, which add difficulties to the abnormal region detection in colonoscopic images.

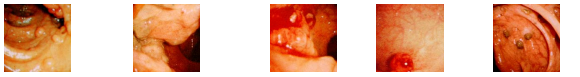

Figure 1. Colonoscopic images with abnormal regions.

Abnormal region detection in colonoscopic images can be regarded as a perceptual image segmentation problem. Although many methods have been developed for segmenting medical images, such as thresholding, region growing and classification-based methods etc.[9], the partitioning of the pixels in an image in most of these methods is based on low level cues, such as illumination, color, texture etc, and the resulted segmentation often disagree with the way human-beings partition the images. Recently, there are some attempts to segment the images by incorporating higher level knowledge. For example, image segmentation is regarded as a graph partitioning problem and a criterion called "Normalized Cuts" for segmenting the graph is proposed [12]. In [10], the image pixels are grouped into "superpixels" which are roughly homogenous in shape and size. Superpixels are grouped into different segments using a classifier. The superpixels provide a higher level cue for segmentation which can produce better segmentation than using single pixels alone. In [8], the entire image was used as an extra source of information to help resolve local ambiguities for object detection and scene recognition. Jojic proposed an intermediate appearance and shape model

between pixel and the whole image called "epitomes" [2] which is a miniature, condensed version of an image. They used some square patches as epitomes. The patch-based approach seems to be a good solution for image segmentation, but choosing the size of the patches is still an open problem. Karkanis investigated patch-based approach for polyp and tumor detection in colonoscopic images [3]. Several patch sizes were tried separately and the one with the least detection error was selected to detect polyps and tumors in the images. However, the sizes of the polyps and tumors are different and the shapes of the tumors are often irregular. Therefore, using single-size patches may not be suitable for all types of abnormalities.

We propose to represent the image regions in colonoscopic images using multi-size patches simultaneously. Multiple-size patches provide multiple level cues on the image regions. At least some among all the patch sizes can better characterize the image region. Represented as multi-size patches, abnormal region detection becomes a binary classification problem of discriminating the patches of normal regions from those abnormal ones. The class label of each pixel can be obtained using the ensemble of multiple classifiers based on different patch sizes.

An ensemble is also known as a mixture of experts, classifier fusion and combination of multiple classifiers, etc [5]. It is a mechanism to combine a set of classifiers so that the resulted ensemble has superior classification performances than using each individual classifier only. The necessary condition for the success of an ensemble is that the outputs of individual classifiers to the same inputs must be diverse [6]. Since we use multi-size patches to construct an ensemble of classifiers for abnormal region detection, these multi-size patches produce different feature sets for the ensemble. We may further improve the performance of the ensemble if the diversity can be increased.

The performance of the ensemble depends on the individual classifiers used. On one hand, the binary classification problem for discriminating the normal patches from those abnormal ones can be solved using a discriminative model, such as binary Support Vector Classifier (2-SVC)[1]. On the other hand, abnormal region detection in colnoscopic images can also be treated as a concept learning problem. Many patterns for each category of abnormalities have to be collected for training a good classifier, which means the concept "abnormal" is not easy to learn. The normal patterns show smaller variation than those of abnormal ones and they are much easier to be collected, which means the concept "normal" is easier to learn. Therefore, the concept "normal" can be learned using a one-class classifier, such as ν–Support Vector Classifier (ν-SVC) [11]. Trained using only the data from one class, the goal of ν-SVC is to find a decision boundary around the training data – called targets. ν-SVC is a non-discriminative recognition-based model since it tries to estimate the support of the target samples rather than for discrimination purpose. Exploiting the different natures of the two types of SVMs, combination of the two types of kernel machines is expected to produce higher diversity to the ensemble, which may help further improve the classification. Experimental results show that our multi-size patch-based hybrid kernel machine ensemble outperforms that of only using single-size patches for the abnormal region detection in colonoscopic images.

2. Image region representation based on multi-size patches

As illustrated in Figure 1, abnormal regions in colonoscopic images vary in location, shape, color and size. The representation of these regions is very important. Patch-based approach is chosen and the task of abnormal region segmentation becomes a binary classification problem. Each image can be cropped into a set of image patches which can be categorized to abnormal class or normal class by a classifier. The abnormal regions can thus be segmented from the normal ones. However, small patch size cannot capture sufficient information of the image regions and often lead to large detection error. Large patch size contains more information about the image regions that match its size and achieve better detection, but fail to represent smaller regions. It is very difficult to determine the appropriate patch-size to use. So we represent these regions using multi-size patches simultaneously. Multi-size patches provide multiple-level representation of the image contents of an image region. At least some among all patch sizes can better characterize the image region. Hence, we propose to construct an ensemble in which each classifier handles one patch size. The combination of classifiers trained using multiple-size patches can recognize the image region more effectively than using single-size patches only.

3. Learning an ensemble for integrating the detection results based on multi-size patches

In this study, five ensembles were investigated for the abnormal region detection in colonoscopic images, including averaging, product, majority voting, decision template and double layer classification [4] [5]. Let $C_i(x) = C_{i1}(x), C_{i2}(x), \cdots, C_{ik}(x)$ be a set of individual classifiers, called an ensemble, each of which gets an input feature vector $x = [x_1, x_2, \cdots, x_d]^T$ and assigns it to a class label y_i from $Y = \{-1, +1\}$, the goal of the ensemble is to find the a class label L_{en} for x based on the outputs of k classifiers $C_1(x), C_2(x), \cdots, C_k(x)$ corresponding to label $L_1(x), L_2(x), \cdots, L_k(x)$. $C_i(x)$ is often an estimate of the posterior probability $P(y_i|x)$. The five ensembles are

described below.

1. Averaging (AVG):

$$L_{en}(x) = \arg\max_j (\sum_{i=1}^{k} \frac{C_{ji}(x)}{k}) \quad (1)$$

where $j = -1, +1$. It calculates the average of the outputs of k individual classifiers and assigns an input x the class label with the largest posterior probability.

2. Product (PROD):

$$L_{en}(x) = \arg\max_j (\prod_{i=1}^{k} \frac{C_{ji(x)}}{k}) \quad (2)$$

where $j = -1, +1$. It calculates the product of the outputs of k individual classifiers and assigns an input x the class label with the largest posterior probability.

3. Majority voting (MV):

$$L_{en}(x) = \arg\max_j (Ct_j) \quad (3)$$

where $j = -1, +1$, and Ct_j is the count of individual classifiers that assign x to the class label j. The output of the ensemble by MV is the label assigned by most of the individual classifiers.

4. Decision template (DT): The decision template DT_j for class y_j is the average of the outputs of individual classifiers in the training set for class y_j [5]. The ensemble DT assigns an input x with the label given by the individual classifier whose Euclidean distance to the decision template DT_j is the smallest.

5. Double layer classification: Taking the output of individual classifiers $C_i(x)$ as input of a upper layer classifier which makes the final decision.

$$L_{en}(x) = F(C_1(x), C_2(x), \cdots, C_k(x)) \quad (4)$$

Linear discriminant classifier (LDC) is used as the upper-layer classifier with assumption of normally distributed classes [5].

4. Hybrid kernel machine ensemble

4.1. Binary Support Vector Classifiers

As a powerful discriminative classifier, 2-SVC is a classical SVM that has been increasingly used in many medical applications and has shown to achieve higher performance than traditional classifiers [1]. It has good generalization by finding an optimal separating hyperplane which minimizes the classification errors made on the training set while maximize the "margin" between different classes. Given a two-class training set $X = \{x_1, x_2 \cdots x_N\}$ (labeled as $y_i = \pm 1$) with N patterns, the data are mapped to another space where the data can be separated by an optimal separating hyperplane:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + \rho \quad (5)$$

where $K(\cdot)$ is a kernel function, ρ is a bias item, α_i ($i = 1, 2, \cdots, N$) are the solutions of a quadratic programming problem that finds the maximum margin.

4.2. ν – Support Vector Classifiers

ν– Support Vector Classifier is a kind of SVM [11]. The difference between ν-SVC and classical SVM is that only data from one class – targets, are used in the training. Without the use of the data from the other class – outliers, the target data are mapped into another space corresponding to a kernel function and separated from the origin with maximum margin by a hyperplane

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x) + \rho \quad (6)$$

For a new point, the value of $f(x)$ is $+1$ if it belongs to the targets. Otherwise the value of $f(x)$ is -1. ν-SVC tries to estimate the density of the target data. It is a recognition-based model whose goal is to represent the target data rather than to discriminate the targets from the outliers. Compared to 2-SVCs, the classification performance of ν-SVC may be a little inferior due to the absence of information from the outlier class, but it leads to a more compact classifier which not only needs less computation in training but also it is robust when the training data set is highly imbalanced.

4.3. Hybrid kernel machine ensemble

Since there are many kinds of abnormalities in colonoscopic images showing large variation, many patterns from abnormal regions in colonoscopic images have to be collected for training a reliable classifier and it is difficult to collect. This leads to an imbalanced data problem. One class – "normal" has many training samples and is easier to model, while the other class – "abnormal" is difficult to model because it has more diverse distributions than the normal class. Therefore, ν-SVC is very suitable for this problem. As a recognition-based model, ν-SVC tries to describe the target data rather than for discrimination purpose,

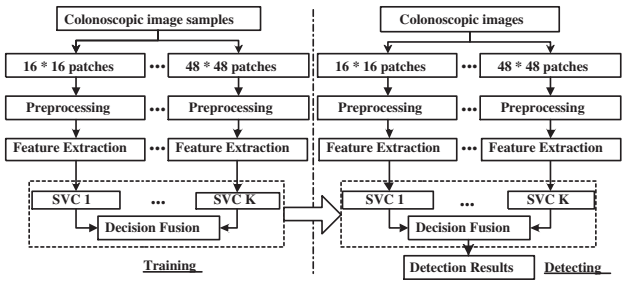

Figure 2. The flowchart of proposed method.

it can handle the problem of missing information. However, ν-SVC is often inferior to 2-SVC for discrimination purpose. This motivates us to combine these two types of kernel machines for this problem. We can build a set of 2-SVCs for the classification first, and use ν-SVCs to provide further decision information. The classification results of the two kernel machines can be aggregated using an ensemble as described in section 3. The different natures of the two types of SVMs adds more diversity to the ensemble, which may further improve the performance of the ensemble.

5. Implementation

Figure 2 illustrates the flowchart of the proposed method for abnormal region detection in colonoscopic images. The details are as follows.

5.1. Image preprocessing and feature extraction

The original endoscopic images obtained by the endoscopy system are RGB images with a resolution of 256×256 pixels. The RGB images are transformed into CIE-Lab color space to analyze the color and luminance components separately. The resulted images are then cropped into fixed sizes of patches. The neighboring patches overlap by 50% to each other to ensure no abnormal region is missed. Here we investigated 3 patch sizes for abnormal region detection, namely, 48×48, 32×32 and 16×16 (pixels). Features can be extracted from these image patches for classification. The features extracted here include the means and standard deviations of the absolute value of the approximate and detail coefficients from a two-level Discrete Wavelet Transform decomposition of the image patches in the 3 channels of CIE-Lab color space. Other features include 1-dimensional histograms of luminance L and 2-dimensional histograms of the component a and b in CIE-Lab color space. The number of bins of the histogram is chosen as 16 for 1-D histograms and 64 for 2-D histograms. Altogether 128 features are extracted, giving rise to a 128-D feature vector. Then the feature vectors are used to form the data set for classification.

5.2. Learning an ensemble for abnormal region detection in colonoscopic images

Using labeled image patches from normal and abnormal regions in colonoscopic images, a set of individual 2-SVCs and ν-SVCs is trained, each for a fixed patch size. An ensemble is constructed based on these individual classifiers. Given a new colonoscopic images, it is first cropped into image patches as described in section 5.1 and then categorized as normal or abnormal using the trained SVCs. Using overlapped image patches, each pixel in the patch can be classified as normal or abnormal by a SVC corresponding to the patch size. Thus each pixel in the original image has at least one label. If a pixel gets different labels, then the label of the patch that has larger confidence is chosen as the label of that pixel. The final label of each pixel in the whole image is assigned by the ensemble of decisions made by these SVMs. The abnormal regions can thus be segmented from those normal ones in the original images.

6. Experimental results and discussions

In the experiment, 46 colonoscopic images with multiple categories of abnormal regions and 12 normal ones are used. The numbers of collected image patches for training of 48×48, 32×32 and 16×16 (pixels) patches are 2002, 2090 and 2126 respectively. The pixels in the original image are manually labeled as the ground truth for comparison. The patches containing mostly abnormal region were labeled as a negative sample, otherwise, a positive one. A leave-one-out experiment was performed for the detection. In each round, one of the colonoscopic images was selected for testing and the patches from other 57 images were used for training. The experiment was repeated 58 times, the detected results were compared to the ground truth image and the average value of the total 58 results was taken as the final result.

The evaluation criteria are specificity (SPE), sensitivity (SEN) and average classification rate (AVR). Where SPE is the fraction of normal regions detected among all the normal regions, SEN is the abnormal regions detected among all the abnormal regions and AVR is the weighted average of SPE and SEN.

$$AVR = \lambda SPE + (1 - \lambda) SEN \tag{7}$$

where λ can be tuned to emphasize SPE or SEN. We used $\lambda = 0.5$ here, so that we treated SPE and SEN equally important.

The detection results of using single patch size and ν-SVC or 2-SVC individually are shown in Table 1. The

detection results of the ensembles are shown in Table 2. In Table 2, $S1 \sim S3$ are the ensemble results of using single-size patch (1 for 48×48, 2 for 32×32, 3 for 16×16) learned by 2-SVC and ν-SVC. AA and AB are the results of ensemble using all 3 patch sizes classified by 2-SVC and ν-SVC, respectively. Columns $A + n$ ($n = 1, 2, 3$) are the results of ensembles using 3 patch sizes learned by 2-SVCs and single patch size n learned by ν-SVC(s). In columns $A+xy$ ($x, y = 1, 2, 3$), x and y are 2 patch sizes are learned by ν-SVC(s). Column ALL are the ensemble results using all 3 patch sizes and both 2-SVC and ν-SVC. The detection results of 4 colonoscopic images are illustrated in Figure 3. Note that we also list the result of an ensemble called oracle, which assigns a correct label to the pattern if any of the single SVCs assigns a correct label [5]. In fact, it is an upper bound which an ensemble can reach. It can shed some light on the diversity of the ensembles.

6.1. ν-SVC vs 2-SVC using single patch size

In Table 1, we observe that 2-SVCs outperform ν-SVC in all the cases which agrees with the postulate that discriminative models are superior to that of recognition-based models. 2-SVCs achieved AVR around 74%, while ν-SVCs achieved only 55%. The ν-SVCs have a very high SPE, but almost completely fail for SEN. This may be resulted that the training set size used for ν-SVC was too small and it also suffered from the curse of dimensionality. Compared to ν-SVCs, 2-SVC have higher SEN while much less SPE, which may be good for add more diversity to the ensembles. The best AVR is 74.5% which was achieved using patches of size 16×16.

6.2. Multi-size patch ensemble of ν-SVCs or 2-SVCs

Columns 5 and 6 of Table 2 illustrate the detection results of 3 patch size ensembles using ν-SVCs or 2-SVCs separately. Obviously, all the ensembles outperforms that of the best SVMs using single patch size, which supports our claim that multi-size patch-based SVM ensemble can achieve more precise abnormal region detection in colonoscopic images. Due to the poor performance of individual ν-SVCs, the improvement of their ensemble is limited although there are still some. But the ensemble of 2-SVCs is quite effective in improving the classification.

6.3. Ensemble of SVMs using single-size patches

Columns 2 to 4 of Table 2 shows the detection results of the ensemble of a ν-SVC and a 2-SVC based on single-size patches. Only DT and LDC achieved AVR comparable to the best single classifier and the performance of other ensembles did not outperform the best single classifier. This may be due to the fact that the ν-SVC and 2-SVC are trained using the same features, less diversity can be introduced into the ensemble which limit the performance of this scheme.

Table 1. Results of abnormal region detection based on single-size patches.

Patch size	Classifier	AVR	SPE	SEN
48×48	2-SVC	0.744	0.675	0.813
48×48	ν-SVC	0.539	0.991	0.088
32×32	2-SVC	0.738	0.675	0.802
32×32	ν-SVC	0.546	0.998	0.094
16×16	2-SVC	0.745	0.668	0.822
16×16	ν-SVC	0.538	0.946	0.094

6.4. Ensemble of 2-SVCs using all 3 patch sizes and several ν-SVC(s)

Columns 7 to 13 of Table 2 illustrates the detection results of the ensemble of 2-SVCs using all 3 patch sizes and 1 (or 2) ν-SVC(s) trained using 1 (or 2) patch size(s). Most of the ensembles shows improvement over the best single SVM based on single-size patches. The performance of LDC, AVG and MV outperforms others. Figure 3 illustrates the result of the ensemble of 2-SVCs using all 3 patch sizes and a ν-SVC(s) trained using patches with size of 48×48. Obviously, the detection results by the ensemble is closer to the ground truth compared to those using single-size patches.

6.5. Observation from the oracle

In the last row of Table 2, we observe that the AVR of ORA (diversity) is only 86.8% for 2-SVCs and 56.9% for ν-SVCs, with multi-size patches trained using 2-SVCs or ν-SVCs only. The AVR of the ORA of using one patch size only and combining 2-SVC and ν-SVC is increased to about 90%, which shows that it adds some diversity but the increase is not very significant. The detection results of oracle 2 in Figure 3 is using the ensembles of three 2-SVCs using all 3 patch sizes and a ν-SVC using 48×48 patches and those of oracle 1 in Figure 3 is using the ensembles of three 2-SVCs trained by all 3 patch sizes. Obviously, the results of oracle 2 is more similar to the ground truth than those of oracle 1. Therefore, the ensembles of 2-SVCs using all 3 patch sizes and one or more ν-SVC(s) significantly improved the diversity, which increased the AVR of ORA to more than 95%. It supports our claim that the multi-size patch-based hybrid SVM ensemble produces higher diversity.

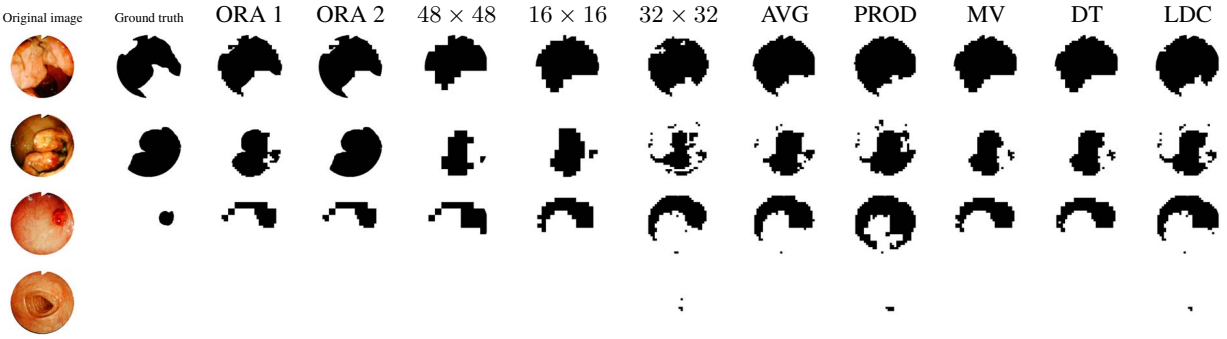

Figure 3. Detection results on 4 colonoscopic images. The regions in black are detected as abnormal.

Table 2. Abnormal region detection results (in terms of AVR) using different ensemble schemes.

Fusion rules	S1	S2	S3	AA	AB	A+1	A+2	A+3	A+12	A+13	A+23	ALL
AVG	0.551	0.556	0.563	0.751	**0.551**	**0.761**	**0.761**	**0.754**	**0.769**	**0.765**	**0.762**	0.565
PROD	0.551	0.556	0.563	0.745	0.535	**0.768**	0.659	0.730	0.598	0.656	0.572	0.548
MV	-	-	-	0.751	**0.551**	0.751	0.751	**0.751**	**0.765**	**0.765**	**0.765**	0.704
DT	**0.744**	**0.738**	**0.745**	0.753	0.538	0.753	0.753	**0.753**	0.753	0.753	0.753	0.751
LDC	**0.746**	**0.741**	**0.745**	**0.763**	0.533	**0.765**	**0.764**	**0.756**	**0.765**	**0.763**	**0.763**	**0.764**
Oracle	0.903	0.900	0.899	0.868	0.569	0.953	0.953	0.953	0.953	0.953	0.953	0.955

7. Conclusions

A new multi-size patch-based hybrid kernel machine ensemble is proposed for abnormal region detection in colonoscopic images. Exploiting the multi-size patch-based image region representation and complementary nature of the two types of SVMs, the proposed ensemble can produce better detection results than that of using single-size patch only. Among the five fusion rules, LDC-based double layer classification outperforms the others. Experimental results show the good performance improvement of our proposed ensemble.

References

[1] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

[2] N. Jojic, B. Frey, and A. Kannan. Epitomic analysis of appearance and shape. In *Proceedings of ICCV 2003*, pages 34–41, Nice, France, October 2003.

[3] S. Karkanis, D. Iakovidis, D. Maroulis, D. Karras, and M. Tzivras. Computer aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. on ITB*, 7(3):141–152, September 2003.

[4] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[5] L. Kuncheva, J. Bezdek, and R. Duin. Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition*, 34(2):299–314, 2001.

[6] L. I. Kuncheva and C. J. Whitaker. Measures of diversity in classifier ensembles. *Machine Learning*, 51(2):181–207, May 2003.

[7] D. E. Maroulis, D. K. Iakovidis, S. A. Karkanis, and D. A. Karras. Cold: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer Methods and Programs in Biomedicine*, 70(2):151–166, Febrary 2003.

[8] K. Murphy, A. Torralba, and W. T. Freeman. Using the forest to see the trees: a graphical model relating features, objects and scenes. In *16th NIPS*, 2004.

[9] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[10] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings of ICCV 2003*, pages 10–17, Nice, France, October 2003.

[11] B. Scholkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

[12] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, August 2000.

[13] P. Wang, S. M. Krishnan, C. Kugean, and M. P. Tjoa. Classification of endoscopic images based on texture and neural network. In *23rd Ann. Inter. Conf. of IEEE/EMBS*, pages 3691–3695, 2001.