

Adaptive kernel based multiple kernel learning for computer-aided polyp detection in CT colonography

MING MA^{*,†}, LIHONG LI[‡], HAO HAN[†], YIFAN HU[†],
XIANFENG GU^{*}, AND ZHENGRONG LIANG^{*,†}

Computer-aided detection (CAD) of colonic polyps, as a second reader for computed tomographic colonography (CTC) screening, has earned extensive research interest over the past decades. False positive (FP) reduction in the CAD system plays a crucial role in detecting the polyps. To improve the performance of FP reduction and better assist the physician's diagnosis, we propose an adaptive kernel based multiple kernel learning (MKL) method for CAD of colonic polyps, called AK-MKL. This method builds a more adaptive synthesized classifier by incorporating an adaptive kernel into a set of predefined base kernels for better performance in differentiating true polyps from FPs, which is implemented by learning an optimal combination of a collection of those kernel-based classifiers. Performance evaluation for the presented AK-MKL method was performed on a CTC database, consisting of 25 patients with 50 CT scans. In terms of the AUC (area under the curve of receiver operating characteristic) and accuracy merits, the experimental results showed that our AK-MKL method achieves better performance, compared with two other different methods, i.e., one classifier based on support vector machine (SVM) with only one adaptive kernel (AK-SVM) and the other one based on multiple kernel learning only (MKL).

1. Introduction

Colorectal cancer, as pointed out in American Cancer Society annual report in 2013 [1], is the third most common malignancy and the second leading cause of cancer-related deaths in the United States. Appearing as bulbous protrusions from the colon wall into the lumen, colonic polyps bear a high

^{*}The author is with Department of Computer Science, Stony Brook University.

[†]The author is with Department of Radiology, Stony Brook University.

[‡]The author is with Department of Engineering Science and Physics, City University of New York at College of Staten Island.

probability (greater than 90%) of transformation into the cancer, although such a long latency period as estimated 5 to 15 years is usually required for this malignant transformation. Thus, early detection and removal of colonic polyps can diminish the risk of colon cancer [34, 7, 15].

Traditionally, optic colonoscopy (OC) acts as a gold standard, being routinely employed in clinic practice for colonic polyp detection. However, OC is not ideal in terms of safety, cost and patient compliance. Computed tomographic colonography (CTC) or virtual colonoscopy (VC), as a new minimally-invasive screening technique for detection of colonic polyps by scanning cleansed and air-distended colon using computed tomography, has shown the potential to become a mass screening modality [36, 10, 18].

In spite of the less invasiveness, VC still suffers from the disadvantages of time-consumingness and proneness to perceptual errors [9]. To address these difficulties, computer-aided detection (CAD) was proposed to automatically detect colonic polyps on CTC scans [6]. Moreover, CAD serves as an initial examination tool to detect polyp candidates. Radiologists can focus on a small portion of regions where colonic polyp candidates are detected, thus reducing interpretation time of radiologists and augmenting the diagnostic performance in polyp detection. As a result, detection of colonic polyps turns out to be more accurate and efficient using VC in conjunction with CAD.

A variety of CAD methods for polyps have been under investigation over the past decades. Summers et al. [29, 30] proposed the use of some geometrical features, such as mean, principle and Gaussian curvatures, to perform the detection of colonic polyp candidates. Yoshida and Nappi [38, 37, 23, 22] improved the polyp detection performance by using the local shape index and curvedness, in comparison with mean and Gaussian curvature features. Paik et al. [24, 25, 26] presented contour normal method and surface overlap method for detection of polyp candidates. Kiss et al. [12, 13] utilized the combination of surface normal and sphere fitting method to detect the polyp candidates, where sphere fitting acted as an important step for discerning polyps from haustral folds. Wang et al. [33] introduced a global curvature shape descriptor for colon wall and extracted internal geometrical, morphological and textural features for detection of colonic polyps. Suzuki et al. [31] utilized a massive-training artificial neural network for distinction between polyps and rectal tubes in polyp detection. Kim et al. [11] applied a Hessian matrix-based algorithm for polyp detection. Li et al. [17] suggested a novel detection method using wavelet-based features generated from two dimensional (2D) projection images, and also optimized colonic polyp detection by evolving the Pareto front [16]. Zhu et al. [42] augmented the detection specificity using features extracted from both the gray and color projection

images. Liu et al. [20] improved the sensitivity of small polyp detection by utilizing image interpolation method for curvature estimation. Mang et al. [21] evaluated the stand-alone performance of a polyp detection method with and without fecal tagging, and showed no significant performance differences. Wang et al. [32] provided a novel polyp detection scheme based on decomposition of complex colon into uniform broken parts using second order derivatives of the volumetric image.

After detecting initial candidates of colonic polyps, false positive (FP) reduction plays a crucial role in a CAD system, where a variety of features extracted from initial polyp candidates (IPCs) are fed into a classifier to reduce FPs, including geometric, density, texture and projection features. These different types of features are highly heterogeneous. Thus, multiple kernel learning (MKL) methods [2, 27, 28, 8], which aim at tackling real classification applications with multiple and heterogeneous data sources, can be employed at the FP reduction stage in our CAD pipeline. The main idea of our paper is that we present a novel kernel-based classification method called AK-MKL for improving the performance of FP reduction in the CAD pipeline. We first construct an adaptive kernel using the training dataset based on the boosting paradigm in [4]. This adaptive kernel is then incorporated into a set of predefined base kernels as a kernel pool for the purpose of learning an optimal combination of a collection of those kernel-based classifiers, under the framework of multiple kernel learning in [35]. The combination of the adaptive kernel learned from training dataset with the well-known predefined base kernels shall generate a more adaptive and strong classifier for better differentiating true polyps from FPs. To validate the feasibility and effectiveness of the proposed method, we evaluate the performance of our AK-MKL method in the experiments on a CTC dataset.

The novelties of our proposed method in this paper are as follows: 1) This paper is the first attempt in introducing the boosting paradigm-based adaptive kernel construction in the application of classification for FP reduction in CAD of colonic polyps in CT colonography. 2) Secondly, our paper is also the first attempt to apply the newly presented multiple kernel learning method to the classification problem in medical imaging, in particular, the false positive reduction in computer-aided detection of colorectal polyps in CT colonography. 3) In the perspective of novelty of the method, we propose a novel AK-MKL method which builds a more adaptive classifier by incorporating an adaptive kernel into a set of predefined base kernels for better discriminative performance in differentiating true polyps from false positives in the stage of FP reduction. 4) We further validate the feasibility and effectiveness of our AK-MKL method in the experiments, demonstrating the

better performance of our method in comparison with two other different methods.

The remainder of this paper is organized as follows. Section II presents our CAD pipeline and the multiple kernel learning with adaptive kernel method for computer-aided detection of polyps. Experimental results on evaluation of our proposed AK-MKL method is reported in Section III, followed by discussion and conclusion in Section IV.

2. Materials and methods

2.1. CAD pipeline

A CAD scheme can automatically detect potential colonic polyps such that radiologists' interpretation time is substantially reduced and the diagnostic performance largely improved. The schematic diagram of the CAD pipeline for the detection of colonic is given in Figure 1, where four stages are summarized [39, 41, 40].

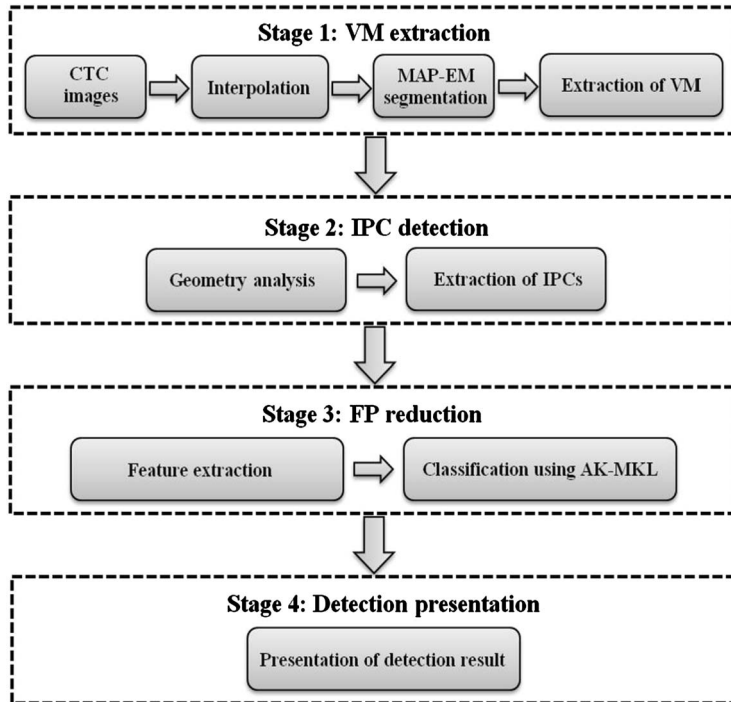


Figure 1: Schematic diagram of CAD pipeline for polyp detection.

In stage 1, a volumetric mucosa (VM) extraction is performed by the maximum-a-posterior expectation-maximization (MAP-EM) image segmentation algorithm [19], which solves the partial volume segmentation problem by estimating the mixture proportion of each tissue type within each voxel of CTC images. Mathematically, the MAP-EM algorithm aims to iteratively maximize the following objective function:

$$(1) \quad Q(\theta|\theta^{(n)}) = E_{Y=\sum_k X_k} [\ln(p(\mathbf{X}|\theta)p(\mathbf{Z}))|\mathbf{Y}, \theta^{(n)}]$$

Suppose there are k types of tissues in the volume of interest for VM segmentation, the image intensity \mathbf{Y} of each voxel consists of the contributions \mathbf{X}_k from each tissue type and thus we have $\mathbf{Y} = \sum_k \mathbf{X}_k$. Let μ_k and ν_k be the mean and variance of the intensity of tissue type k , \mathbf{Z}_{ik} be the mixture proportion of tissue type k in the i -th voxel, then $\theta = \{\mu_k, \nu_k, \mathbf{Z}_{ik}\}$ is the vector of to-be-estimated parameters. Let $p(\mathbf{X}|\theta)$ be the likelihood on CT image intensities and $p(\mathbf{Z})$ represent the prior of mixture proportions, the Q function actually models the conditional expectation of the posterior probability of the CT image intensity and tissue mixture proportions.

In practice, the interpolation between corresponding pixels on the adjacent slices of CTC images is required for the generation of 3D isotropic voxels, prior to the MAP-EM segmentation. Figure 2 illustrates the process of MAP-EM segmentation for VM extraction. One representative CTC image is shown in Figure 2(a), and the image after electronically cleansing of the tagged colonic residual is given in Figure 2(b). Electronic colon cleans-

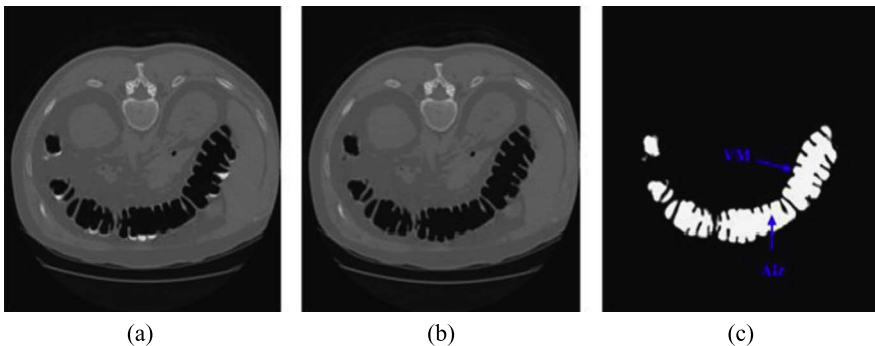


Figure 2: MAP-EM segmentation for VM extraction. (a) A representative slice of abdominal CT image with tagged fecal material. (b) The resulting abdominal CT image with tagged fecal material cleansed electronically. (c) The cleaned colon lumen slice with arrows indicating VM and air.

ing is an emerging technique developed to segment the colon lumen from a patient’s abdominal CTC images by removing the residue stool and fluid tagged using contrast materials. As the region of interest (ROI) for polyp detection, the VM extracted by the MAP-EM segmentation algorithm is shown in Figure 2(c). The colon lumen within the VM layer is shown in Figure 2(c) as well. Stage 2 shows that the VM is analyzed through geometric computation, and the resulting measurement is then utilized to extract the IPCs. FP reduction in stage 3 is our main contribution in this paper, where various features are extracted and fed into the classifier constructed using the proposed AK-MKL method for the purpose of FP reduction. In stage 4, the results of detection are displayed in the colon visualization system in order to assist the physician’s diagnosis.

2.2. Adaptive kernel based multiple kernel learning method

2.2.1. Overview of AK-MKL method The proposed multiple kernel learning with adaptive kernel (AK-MKL) method in this paper is essentially aimed to learn an optimal combination of a group of base classifiers, each of which is constructed using a specific kernel. The kernel pool comprises two parts. One part is a set of predefined radial basis function (RBF) kernels, and the other part is an adaptive kernel constructed using a boosting paradigm. The key of our method is to incorporate the adaptive kernel with the predefined base kernels for the purpose of building a synthesized strong classifier.

The flow chart of AK-MKL method is presented in Figure 3. The details of this method are presented in both Algorithm 1 and 2.

2.2.2. Construction of the adaptive kernel Some of the conventional implementations of the classification methods in previous applications only pre-select a kernel function (e.g., RBF kernels), ignoring available training data. These kind of pre-selection approaches fail to provide an optimal kernel function which is adaptive to the specific data. Because of the adaptive nature of the adaptive kernel constructed by learning from the training data, it is expected that by introducing the adaptive kernel into our method, we shall improve the classification for polyp detection.

The adaptive kernel in our AK-MKL method is constructed by an iterative learning process where the boosting paradigm is applied to combine the primitive kernels in consideration of the training data. In the following, the basic kernel method is firstly outlined, and then details of the construction of the adaptive kernel are presented.

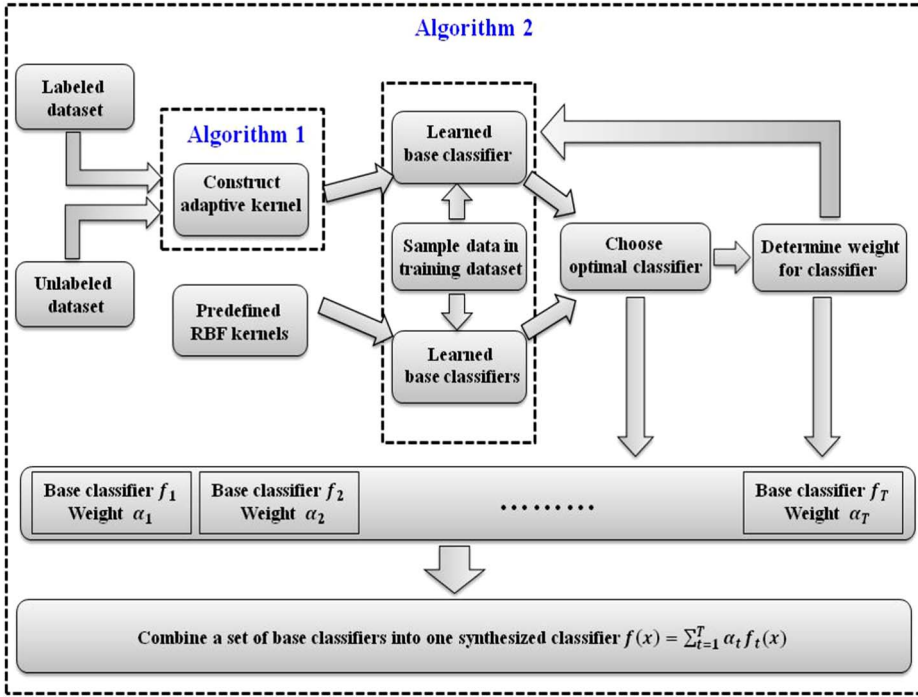


Figure 3: Flow chart of our proposed AK-MKL algorithm.

For a linear classifier, a hyperplane can be used to linearly separate the data. Given a set of labeled data $\mathbb{D} = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \{+1, -1\}, i = 1, \dots, n\}$ where x_i is a d -dimensional vector of features and y_i is the class label, the hyperplane can be represented as

$$(2) \quad \mathbf{w}^T \mathbf{x} + b = 0$$

where \mathbf{w} denotes the normal vector to the hyperplane. To linearly separate the data, two hyperplanes are selected as the boundary of “margin”, which are expressed as follows:

$$(3) \quad \mathbf{w}^T \mathbf{x} + b = 1$$

$$(4) \quad \mathbf{w}^T \mathbf{x} + b = -1$$

Then we aim to maximize the margin $\frac{2}{\sqrt{\mathbf{w}^T \mathbf{w}}}$, i.e., the distance between two hyperplanes above, which is equivalent to minimizing $\frac{\sqrt{\mathbf{w}^T \mathbf{w}}}{2}$. Hence, we need to solve a quadratic programming optimization problem:

$$(5) \quad \min_{\mathbf{w}, b} \frac{\sqrt{\mathbf{w}^T \mathbf{w}}}{2}$$

subject to $y_i(\mathbf{w}^T x_i + b) \geq 1, \forall i = 1, \dots, n$

However, we cannot guarantee that all the data are correctly labelled, so we introduce slack variables ϵ_i . The quadratic programming optimization problems then becomes:

$$(6) \quad \min_{\mathbf{w}, b, \epsilon} \frac{\sqrt{\mathbf{w}^T \mathbf{w}}}{2} + C \sum_i \epsilon_i$$

subject to $y_i(\mathbf{w}^T x_i + b) \geq 1 - \epsilon_i, \forall i = 1, \dots, n$

By introducing Lagrange multipliers α , the constrained optimization problem in (5) can be expressed as

$$(7) \quad \max_{\alpha_i \geq 0} \min_{\mathbf{w}, b} \frac{\sqrt{\mathbf{w}^T \mathbf{w}}}{2} - \sum_i \alpha_i (1 - y_i(\mathbf{w}^T x_i + b))$$

When taking into consideration the slack variables ϵ_i , we impose constraints on the Lagrange multipliers α such that $0 \leq \alpha_i \leq C$. Let $\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{\sqrt{\mathbf{w}^T \mathbf{w}}}{2} - \sum_i \alpha_i (1 - y_i(\mathbf{w}^T x_i + b))$, we solve this optimization problem by setting $\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0$ and $\frac{\partial \mathcal{L}}{\partial b} = 0$ and then have $\mathbf{w} = \sum_i \alpha_i y_i x_i$ and $\sum_i \alpha_i y_i = 0$ respectively. After plugging $\mathbf{w} = \sum_i \alpha_i y_i x_i$ and simplifying, we have

$$(8) \quad \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Finally, we have constrained optimization problem for linear classifier:

$$(9) \quad \max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

subject to $\sum_i \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$

where C is an error allowance parameter.

However, data is usually non-linear and cannot be linearly separable. Kernel method is such a technique that could non-linearly map the data in \mathbb{R}^d space to a high-dimensional feature space \mathbb{F} where they are linearly separable. Formally, let x_i and x_j denote two data in \mathbb{R}^d space respectively and $\phi : \mathbb{R}^d \rightarrow \mathbb{F}$ be a mapping into feature space \mathbb{F} . The kernel function is defined as an inner product in feature space

$$(10) \quad K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

By replacing $x_i^T x_j$ with the kernel function $K(x_i, x_j)$ in (9), we have the constrained optimization problem for non-linear classifier based on the kernel function:

$$(11) \quad \begin{aligned} & \max_{\alpha_i \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ & \text{subject to } \sum_i \alpha_i y_i = 0 \text{ and } 0 \leq \alpha_i \leq C \end{aligned}$$

The corresponding classifier is then expressed as:

$$(12) \quad f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x)$$

Intuitively, a kernel function is a means for similarity measurement between a pair of data. Two common kernels applied for classification in various applications are polynomial kernels and radial basis function (RBF) kernels, which are given as follows respectively:

$$(13) \quad K(x_i, x_j) = (x_i^T x_j + 1)^p$$

$$(14) \quad K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

where the parameter p is the degree of the polynomial kernel and σ is the width of RBF kernel.

In most classification applications, the kernels are predefined, for example, a RBF kernel is selected before learning starts. This pre-selection approach, however, may fail to offer a desirable kernel which is adaptive to a specific classification application. In our presented AK-MKL method, a key kernel used in classification is constructed by learning the empirical training data, rather than by the pre-selecting approach. The experts' input as the labeling information is employed to construct an adaptive kernel with respect to the specific data.

The adaptive kernel in our AK-MKL method is constructed using the boosting paradigm [4]. The idea behind this strategy is, in an iterative learning fashion, to construct a target adaptive kernel \widehat{K} which is a weighted combination of a set of primitive kernels $\{K_{\mu,t} | t = 1, 2, \dots, T\}$ w.r.t. a vector μ , written as:

$$(15) \quad \widehat{K}(x_i, x_j) = \sum_t \alpha_t K_{\mu,t}(x_i, x_j)$$

In each iteration t of the boosting process, a primitive kernel $K_{\mu,t}$ is learned using generalized eigen vector decomposition, given the labeled training data set \mathbb{D} , unlabeled testing dataset $\widetilde{\mathbb{D}}$ and a distribution P over the labeled training dataset. Then the target adaptive kernel \widehat{K} is composed of a set of

Algorithm 1 Construction of the adaptive kernel**Input:** Labeled dataset: $\mathbb{D} = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \{+1, -1\}, i = 1, \dots, n\}$,unlabeled dataset: $\tilde{\mathbb{D}} = \{\tilde{x}_i | \tilde{x}_i \in \mathbb{R}^d, i = 1, \dots, \tilde{n}\}$ **Output:** Adaptive kernel \hat{K} 1: $\hat{K} = 0$ 2: **for** $t = 1, 2, \dots, T$ **do**3: Compute distribution P_t over \mathbb{D} : $P_t(i, j) = \frac{1}{1 + \exp(-y_i y_j K(x_i, x_j))}$ 4: compute matrices $B \in \mathbb{R}^{n \times \tilde{n}}$ with $B_{i,r} = x_i^T \tilde{x}_r$, $W \in \mathbb{R}^{n \times n}$ with $W_{i,j} = P_t(i, j) y_i y_j$ and $M \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$ with $M_{r,s} = \tilde{x}_r^T \tilde{x}_s$ 5: Using generalized eigen vector decomposition, solve $B^T W B v = \lambda M v$ and find the generalized eigenvector $v \in \mathbb{R}^n$ corresponding to the largest eigenvalue λ 6: Set vector $\mu = \frac{\sum_r v_r \tilde{x}_r}{\|\sum_r v_r \tilde{x}_r\|}$ 7: Primitive kernel $K_{\mu,t}$ is acquired by $K_{\mu,t} = \mu \mu^T$ 8: Compute $l_t^+ = \{(i, j) | y_i y_j K_{\mu,t}(x_i, x_j) > 0\}$
 $W_t^+ = \sum_{(i,j) \in l_t^+} P_t(i, j) |K_{\mu,t}(x_i, x_j)|$
 $l_t^- = \{(i, j) | y_i y_j K_{\mu,t}(x_i, x_j) < 0\}$
 $W_t^- = \sum_{(i,j) \in l_t^-} P_t(i, j) |K_{\mu,t}(x_i, x_j)|$ 9: Update $\hat{K} = \hat{K} + \frac{1}{2} \ln\left(\frac{W_t^+}{W_t^-}\right) K_{\mu,t}$ 10: **end for**

weighted primitive kernels $\{\alpha_t K_{\mu,t} | t = 1, 2, \dots, T\}$, where the labeled training dataset is employed to evaluate the learned kernel $K_{\mu,t}$ and to determine its corresponding weight α_t . In the following, we give the description of the algorithm for the adaptive kernel construction using the boosting paradigm.

In the above algorithm, step (1) initializes the kernel \hat{K} by assigning zeros matrix, followed by T iterations of boosting process. During each iteration, a distribution P_t is computed in step (3). This distribution is essentially a logarithmic alignment loss [5, 14] which bounds the empirical classification error. Steps (4) to (7) aim to compute the current primitive kernel $K_{\mu,t}$ with respect to the current distribution P_t . The resulting primitive kernel $K_{\mu,t}$ is the outer-product between the vector μ and itself, written as $K_{\mu,t} = \mu \mu^T$ with norm constraint $\|\mu\| = 1$, and the kernel function for a pair of data x_i and x_j is expressed as $K_{\mu}(x_i, x_j) = x_i^T \mu \mu^T x_j$. To evaluate how good a primitive kernel K_{μ} is, a score for the kernel K_{μ} w.r.t. the current distribution P_t is defined:

$$(16) \quad \text{score}(K_{\mu}) = \sum_{i,j} P(i, j) y_i y_j K_{\mu}(x_i, x_j)$$

The larger the score is, the better primitive kernel K_{μ} fits the training data.

The optimization problem of maximizing score of K_μ can be solved by the generalized eigen vector decomposition (i.e. step (5)). At the completion of the primitive kernel generation, step (8) groups all pairs of data into two sets with index pairs l_t^+ and l_t^- respectively, one being the positive labeled pairs where a pair of data x_i and x_j have the same label, and the other one being the negative labeled pairs where they have the opposite label. Using these two sets of index pairs, the weight α_t for the primitive kernel $K_{\mu,t}$ in the weighted combination $\hat{K}(x_i, x_j) = \sum_t \alpha_t K_{\mu,t}(x_i, x_j)$ which denotes the target adaptive kernel is expressed as $\alpha_t = \frac{1}{2} \ln\left(\frac{W_t^+}{W_t^-}\right)$ in step (9). Thus, combining each weighted primitive kernel which is generated in each iteration of boosting process will eventually lead to an adaptive kernel \hat{K} .

2.2.3. Boosting-based multiple kernel learning Multiple kernel learning (MKL) is a technique using multiple kernel functions to learn the optimal combination of kernels or classifiers for classification tasks. In this paper, we propose a novel kernel-based classification method by combining the predefined base kernels with the adaptive kernels to build a synthesized strong classifier for FP reduction in CAD of colonic polyps, which is new and a breakthrough in the presented paper.

The idea behind the presented AK-MKL method is that the boosting strategy is applied to learn a synthesized classifier by a weighted combination of multiple base classifiers. Each of such base classifier is generated using an optimal kernel selected among a set of kernels in each iteration of the learning process. The pool of kernels which are used in our AK-MKL method encompass not only several predefined RBF kernels but also the adaptive kernel constructed using the aforementioned Algorithm 1. In what follows, our proposed AK-MKL method is detailed in Algorithm 2.

The first two steps in the above algorithm generate several predefined RBF kernels and an adaptive kernel constructed using Algorithm 1. Step (1) builds the commonly used RBF kernels which are predefined by assigning different values of width parameter. Step (2) constructs an adaptive kernel learned from training data. This combination of predefined kernels with adaptive kernel based on empirical data could be employed to learn a more adaptive classifier. In each iteration, only a proportion ρ of training data are sampled for learning in step (5). Then the optimal base classifier f_t with minimal classification error ϵ_t in current iteration t is selected among M learned classifiers, each of which is learned from a kernel function K_j , as shown in steps (7) to (14). Then, a weight α_t for the learned optimal base classifier f_t is computed according to the classification error ϵ_t in step (15). The subsequent steps (16) to (22) update the distribution $D_t(i)$ which

Algorithm 2 Multiple kernel learning with adaptive kernel

Input: Labeled dataset: $\mathbb{D} = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \{+1, -1\}, i = 1, \dots, n\}$,

unlabelled dataset: $\tilde{\mathbb{D}} = \{(\tilde{x}_i | \tilde{x}_i \in \mathbb{R}^d, i = 1, \dots, \tilde{n})\}$

number of kernels M , sample ratio ρ and number of boosting trials T

Output: The synthesized classifier $f(x)$

- 1: Generate $M - 1$ RBF kernels: K_1, K_2, \dots, K_{M-1}
 - 2: Call Algorithm 1 to construct an adaptive kernel K_M as one of the base kernels
 - 3: Initialize the distribution $D_1(i = \frac{1}{n})$ where $i = 1, 2, \dots, T$
 - 4: **for** $t = 1, 2, \dots, T$ **do**
 - 5: Sample $m = n\rho$ data in labeled training dataset \mathbb{D}
 - 6: Reset optimal base classifier f_t and minimal training error $\epsilon_t = 1$
 - 7: **for** $j = 1, 2, \dots, M$ **do**
 - 8: Learn a base classifier f_t^j using kernel K_j
 - 9: Compute training error $\epsilon_t^j = \sum_{i=1}^n D_t(i)(f_t^j(x_i) \neq y_i)$
 - 10: **if** $\epsilon_t^j < \epsilon_t$ **then**
 - 11: $\epsilon_t = \epsilon_t^j$
 - 12: $f_t = f_t^j$
 - 13: **end if**
 - 14: **end for**
 - 15: set weight $\alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t})$
 - 16: **for** $i=1, 2, \dots, n$ **do**
 - 17: **if** $f_t(x_i) = y_i$ **then**
 - 18: $D_{t+1}(i + 1) = \frac{D_t(i)e^{-\alpha_t}}{\sum_{k=1}^n D_t(i)}$
 - 19: **else**
 - 20: $D_{t+1}(i + 1) = \frac{D_t(i)e^{\alpha_t}}{\sum_{k=1}^n D_t(i)}$
 - 21: **end if**
 - 22: **end for**
 - 23: **end for**
 - 24: Construct synthesized classifier $f(x) = \sum_{t=1}^T \alpha_t f_t(x)$
-

features the importance of training data for learning. Finally, the target synthesized classifier is given by the combination of T base classifiers.

3. Results

3.1. CTC database

The performance of the presented AK-MKL method for polyp detection was evaluated on a CTC database from 25 patients with 50 CT scans. The colon was cleaned by standard pre-colonoscopy or barium enema bowel preparation. The stool and fluid were tagged using a single dose of 2% barium and

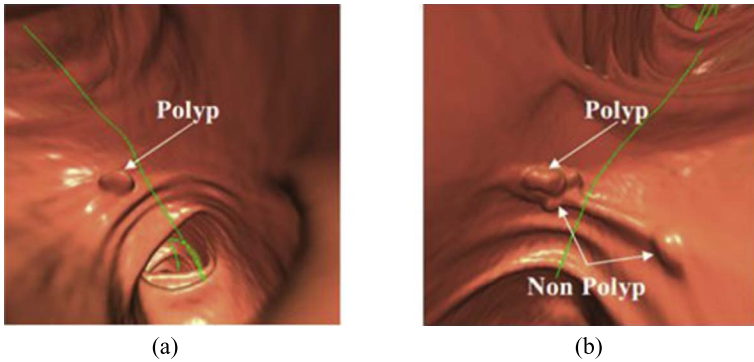


Figure 4: True polyps and non-polyps. (a) true polyps on colon wall. (b) true polyp and non-polyp on haustral fold.

diatrizoate. The 50 CT scans were acquired using the multi-slice CT scanners in helical mode with the following settings. The collimation was set to be 1.25 to 5.0 mm, the reconstruction intervals to be 1.25 to 5.0 mm, the modulated tube current to be 50 to 200 mAs and the tube voltages to be 80 to 120 kVp.

In this CTC database, 64 out of 786 IPCs are true polyps (shown in Figure 4(a)) confirmed by both optical colonoscopy (i.e. the golden standard in this study) and CTC expert reading. The remaining are FPs (shown in Figure 4(b)), such as fecal residue, normal tissue or image artifact, etc. The polyp size ranges from 6 to 22 mm, and totally 21 features were extracted for FP reduction.

3.2. Experimental settings

To evaluate the performance of our presented AK-MKL method, we compared it with another two classifiers in terms of AUC (area under the curve of receiver operating characteristics) and accuracy. One is the support vector machine (SVM) with only one adaptive kernel (abbreviated as AK-SVM) which is constructed using Algorithm 1, and the other one is the classifier based on multiple kernel learning only (abbreviated as MKL) which is learned on a set of 21 different RBF kernels without the adaptive kernel. The classifier based on our proposed AK-MKL method employs both RBF kernels and the adaptive kernel.

In terms of implementation of these methods, the LIBSVM library [3] was utilized in our experiments. The number of boosting trials was set to

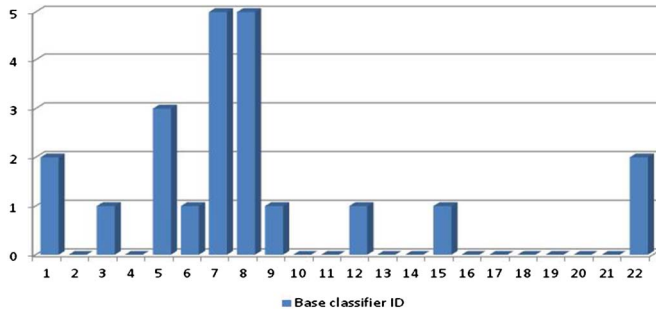


Figure 5: The number of times that each base classifier is chosen.

100 unless stated otherwise. To report the average performance, we ran the programs 100 times repeatedly. In the following, we report experimental results by evaluating the performances.

3.3. Performance evaluation

Prior to the evaluation of the performance of our presented AK-MKL method, we first demonstrate its feasibility of combining a collection of 21 different RBF kernels with the adaptive kernel. In Figure 5, the number in the X-coordinate represents the ID of each base classifier, of which the first 21 classifiers are learned from various RBF kernels with different Gaussian kernel widths, and the last one (ID = 22) is constructed based on the adaptive kernel using Algorithm 1. Y-coordinate value denotes the number of times that each base classifier is selected in 22 boosting trials. As observed from Figure 5, the 22-th base classifier constructed with the adaptive kernel is chosen twice, which shows the strength of the adaptive kernel and its contribution for combination with other RBF kernels into an efficient and high-performance integrated classifier.

In order to evaluate the performance of the three classifiers based on AK-SVM, MKL and AK-MKL method, we firstly compare their performance in terms of AUC and accuracy w.r.t. different sample ratios (i.e., the proportion of training data to be sampled among all of training data, which ranges from 0.1 to 1). Table 1 shows the experiment results of AUC and accuracy for different methods. Figure 6 demonstrates the performance in terms of AUC. As observed from Figure 6, the AUC measurement of AK-SVM is not stable w.r.t. different sample ratios, while the performances of MKL and AK-MKL are improved with the growing sampling data size from the training dataset. Although AK-SMV outperforms MKL and AK-MKL when

Table 1: Evaluation of three classifiers in terms of AUC and accuracy w.r.t. different sample ratio

Sample ratio	AK-SVM	AK-SVM	MKL	MKL	AK-MKL	AK-MKL
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
0.1	0.9324	0.9183	0.9158	0.9756	0.9203	0.9763
0.2	0.9229	0.9181	0.9323	0.9778	0.9415	0.9782
0.3	0.9256	0.9178	0.9515	0.9792	0.9572	0.9794
0.4	0.9291	0.9175	0.9676	0.9796	0.9723	0.9799
0.5	0.9442	0.9186	0.9784	0.9807	0.9857	0.9815
0.6	0.9530	0.9183	0.9811	0.9821	0.9845	0.9822
0.7	0.9232	0.9180	0.9881	0.9828	0.9904	0.9832
0.8	0.9195	0.9175	0.9897	0.9832	0.9930	0.9837
0.9	0.9192	0.9171	0.9954	0.9839	0.9966	0.9840
1.0	0.9203	0.9186	0.9970	0.9835	0.9988	0.9836

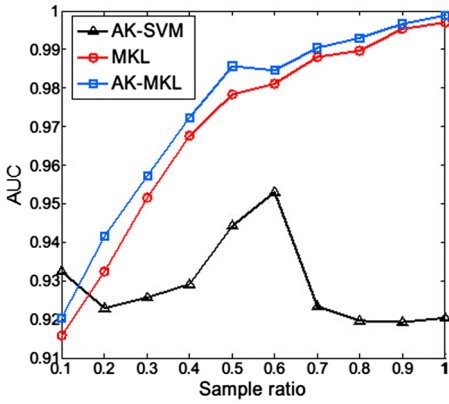


Figure 6: Comparison of three algorithms in terms of AUC.

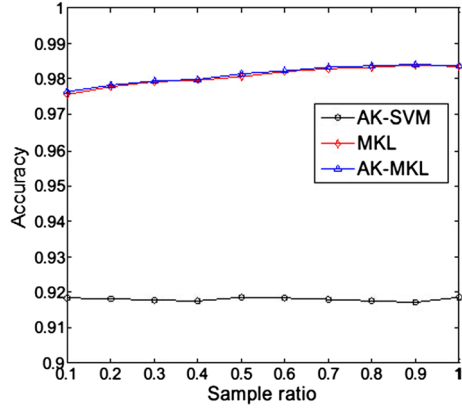


Figure 7: Comparison of three algorithms in terms of accuracy.

the sample ratio is around less than 0.15, the overall performance of MKL and AK-MKL achieve better AUC measurement (as seen when the sample ratio is greater than 0.15). Moreover, Figure 6 shows that the AK-MKL has better performance than MKL in terms of AUC measurement.

The three methods were also evaluated on their performance in accuracy, and the experimental results are presented in Figure 7 and Figure 8. As shown in Figure 7, MKL and AK-MKL attain better performance than AK-SVM, and AK-MKL and MKL are comparable to each other. When focusing on MKL and AK-MKL in the zoomed-in view shown in Figure 8, AK-MKL still outperforms MKL to some degree.

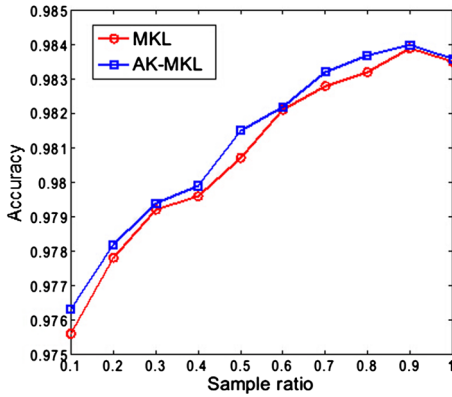


Figure 8: Accuracy comparison of two algorithms in a zoomed-in view.

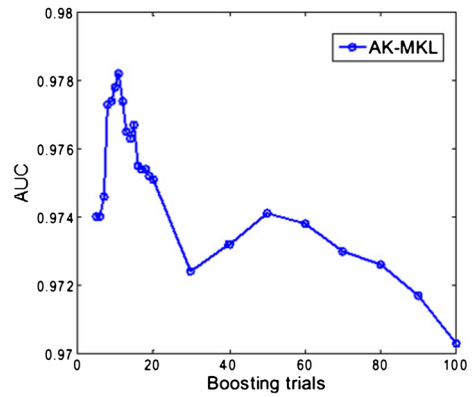


Figure 9: Performance of AK-MKL w.r.t. number of boosting trials.

To investigate the influence of the number of boosting trials on our proposed AK-MKL method, experiment of AUC evaluation was carried out for AK-MKL method when sample ratio is set to 0.4 and the corresponding results are shown in Table 2. Figure 9 demonstrates that increasing the number of boosting trials leads to the improvement of performance prior to reaching the optimal performance, and then the performance falls down with the increase in the number of boosting trials. At the scenario of 11 boosting trials, the optimal performance of AK-MKL in a current experimental setting is achieved.

4. Discussion and conclusion

In this paper, we proposed a novel AK-MKL method for achieving better classification performance in FP reduction stage of CAD of colorectal polyps. The novel adaptiveness of this method from the adaptive kernel constructed using boosting paradigm and an optimal combination of a set of base classifiers by incorporating the adaptive kernel into the predefined RBF kernel pool promotes our method to achieve better performance in terms of AUC and accuracy, compared with two other methods, such as AK-SVM and MKL.

In Figure 6, we observed that the AK-SMV outperforms MKL and AK-MKL when a small sampled portion of training data is used. This mainly results from the fact that a single learned adaptive kernel from the empirical training data enables the corresponding classifier to become more effective in classification, whereas a more complicated combination of a set of kernels

Table 2: Evaluation of AK-MKL in terms of AUC w.r.t. different boosting trials

Boosting trials	AK-MKL AUC
5	0.9740
5	0.9740
6	0.9740
7	0.9746
8	0.9773
9	0.9774
10	0.9778
11	0.9782
12	0.9774
13	0.9765
14	0.9763
15	0.9767
16	0.9775
17	0.9754
18	0.9754
19	0.9752
20	0.9751
30	0.9724
40	0.9732
50	0.9741
60	0.9738
70	0.9730
80	0.9726
90	0.9717
100	0.9703

may lead to the declining of the performance. On the contrary, when there is a large amount of sampled training data, the performance of AK-SVM with the single kernel learned from training data may decline due to redundancy of the training data. MKL and AK-MKL, however, can achieve good performance by an optimal combination of various kernels which fit the data well.

In examining how the number of boosting trials imposes the influence on the performance of the presented AK-MKL method, the observation from Figure 9 comes to show that there usually exists an optimal performance with respect to a well-selected number of boosting trials. Generally, the number of boosting trials can be chosen between 5 and 100 in practice.

We have developed a novel classification method for FP reduction in CAD of colorectal polyps based on multiple kernel learning with adaptive

kernel method, called AK-MKL. Experiment results demonstrated the feasibility and effectiveness of our presented AK-MKL method. For our future research, we will investigate the ensemble classifier based on the support vector machine (SVM), random forest (RF) and linear discriminant analysis (LDA), for achieving better classification performance to improve the FP reduction in CAD of colorectal polyps.

Acknowledgements

Ming Ma and Xianfeng Gu would like to acknowledge the founding sources NSF DMS-1418255, NSF DMS-1221339, AFOSR FA9550-10-1-0294. The research of Lihong Li, Hao Han, Yifan Hu and Zhengrong Liang was supported in part by the National Institutes of Health/National Cancer Institute under Grant #CA082402 and #CA143111.

References

- [1] American Cancer Society (2013). Cancer Facts & Figures. *American Cancer Society*.
- [2] F. Bach, G. Lanckriet, and M. I. Jordan (2004). Multiple kernel learning, conic duality, and the SMO algorithm. *Proc. of International Conference on Machine Learning*. **6**.
- [3] C. H. Chang and C. J. Lin (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. **2(27)** 1–27.
- [4] K. Crammer, J. Keshet, and Y. Singer (2002). Kernel design using boosting. *Advances in Neural Information Processing Systems*. **1** 537–544.
- [5] N. Cristianini, A. Elisseeff, J. Shawe-Taylor, and J. Kandla (2001). On kernel target alignment. *Advances in Neural Information Processing Systems*. **2** 4.
- [6] S. M. Frentz and R. M. Summers (2006). Current status of CT colonography. *Academic Radiology*. **13(12)** 1517–1531.
- [7] T. M. Gluecker, C. D. Johnson, W. S. Harmsen, K. P. Offord, A. M. Harris, L. A. Wilson, and D. A. Ahlquist (2003). Colorectal cancer screening with CT colonography, colonoscopy, and double-contrast barium enema examination: prospective assessment of patient perceptions and preferences. *Radiology*. **227(2)** 378–384.

- [8] M. Gönen and E. Alpaydin (2011). Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12** 2211–2268. [MR2825425](#)
- [9] C. D. Johnson and A. H. Dachman (2000). CT colonography: the next colon screening examination. *Radiology.* **216** 331–341.
- [10] D. Johnson, W. Harmsen, L. Wilson, R. MacCarty, T. Welch, D. Ilstrup, and D. Ahlqist (2003). Prospective blinded evaluation of CT colonoscopy for screen detection of colorectal polyps. *Gastroenterology.* **125(2)** 311–319.
- [11] S. H. Kim, J. M. Lee, J. G. Lee, and B. I. Choi (2007). Computer-aided detection of colonic polyps at CT colonography using a Hessian matrix-based algorithm: preliminary study. *Gastrointestinal Imaging.* **189** 41–51.
- [12] G. Kiss, J. Cleynenbreugel, M. Thomeer, P. Suetens, and G. Marchal (2002). Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods. *European Radiology.* **12(1)** 77–81.
- [13] G. Kiss, J. Cleynenbreugel, M. Thomeer, G. Marchal, and P. Suetens (2002). Computer aided diagnosis for virtual colonography by geometrical model fitting. *Proc. of 7th International Workshop on Vision, Modeling, and Visualization.* 27–34.
- [14] G. Lanckriet, N. Cristianini, P. Bartlett, L. ElGhaoui, and M. Jordan (2002). Learning the kernel matrix with semi-definite programming. *Proc. of the 19th International Conference on Machine Learning.* **5** 27–72. [MR2247973](#)
- [15] B. Levin, D. A. Lieberman, and B. McFarland (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps: a joint guideline from the American Cancer Society, the US multi-society task force on colorectal cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians.* **58(3)** 130–160.
- [16] J. Li, A. Huang, J. Yao, J. Liu, R. Van Uitert, N. Petrick, and R. M. Summers (2009). Optimizaing computer-aided colonic polyp detection for CT colonography by evolving the pareto front. *Medical Physics.* **36** 201–212.
- [17] J. Li, R. Van Uitert, J. Yao, and R. M. Summers (2008). Wavelet method for CT colonography computer-aided polyp detection. *Medical Physics.* **35** 3527.

- [18] Z. Liang and R. Richards (2010). Virtual colonoscopy v.s. optical colonoscopy. *Expert Opinion on Medical Diagnostics Journal*. **4(2)** 149–158.
- [19] Z. Liang and S. Wang (2009). An EM approach to MAP solution of segmenting tissue mixtures: a numerical analysis. *IEEE Transactions on Medical Imaging*. **28(2)** 297–310.
- [20] J. Liu, S. Kabadi, R. Van Uitert, N. Petrick, R. Deriche, and R. M. Summers (2011). Improved computer-aided detection of small polyps in CT colonography using interpolation for curvature estimation. *Medical Physics*. **38(7)** 4276–4284.
- [21] T. Mang, L. Bogoni, M. Salganicoff, M. Wolf, V. Raykar, M. Macari, P. J. Pickhardt, F. Iafrate, A. Laghi, M. Weber, M. E. Baker, H. Ringl, C. J. Herold, and A. Graser (2012). Computer-aided detection of colorectal polyps in CT colonography with and without fecal tagging: a stand-alone evaluation. *Invest Radiology*. **47(2)** 99–108.
- [22] J. Nappi, H. Frimmel, A. Dachman, and H. Yoshida (2004). Computerized detection of colorectal masses in CT colonography based on fuzzy merging and wall-thickening analysis. *Medical Physics*. **31(4)** 860–872.
- [23] J. Nappi and H. Yoshida (2002). Automated detection of polyps with CT colonography: evaluation of volumetric features for reduction of false-positive findings. *Academic Radiology*. **9(4)** 386–397.
- [24] D. S. Paik, C. F. Beaulieu, C. Jeffrey, J. Yee, A. Steinauer-Gebauer, and S. Napel (2000). Computer aided detection of polyps in CT colonography: method and free-response ROC evaluation of performance. *Radiology*. **217(P)** 370.
- [25] D. S. Paik, C. F. Beaulieu, A. Mani, R. Prokesch, J. Lee, and S. Napel (2001). Evaluation of computer-aided detection in CT colonography: potential applicability to a screening population. *Radiology*. **221(P)** 332.
- [26] D. S. Paik, C. F. Beaulieu, G. D. Rubin, B. Acar, R. B. Jeffrey, J. Yee, J. Dey, and S. Napel (2004). Surface normal overlap: a computer-aided detection algorithm, with application to colonic polyps and lung nodules in helical CT. *IEEE Transactions on Medical Imaging*. **23** 661–675.
- [27] S. Sonnenburg, G. Ratsch, and C. Schafer (2005). A general and efficient multiple kernel learning algorithm. *NIPS*. **18** 1275–1282.

- [28] S. Sonnenburg, G. Ratsch, C. Schafer, and B. Scholkopf (2006). Large scale multiple kernel learning, *Journal of Machine Learning Research. Proc. of International Conference on Machine Learning.* **7** 1531–1565. [MR2274416](#)
- [29] R. M. Summers, C. D. Johnson, L. M. Pusanik, J. D. Malley, A. M. Youssef, and J. E. Reed (2001). Automated polyp detection at CT colonography: feasibility assessment in a human population. *Radiology.* **219** 51–59.
- [30] R. M. Summers, J. Yao, P. J. Franaszek, M. Pickhardt, I. Bitter, D. Brickman, V. Krishna, and J. R. Choi (2005). Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology.* **129** 1832–1844.
- [31] K. Suzuki, H. Yoshida, and J. Näppi (2006). Massive-training artificial neural network (MTANN) for reduction of false positives in computer-aided detection of polyps: suppression of rectal tubes. *Medical Physics.* **33** 3814–3824.
- [32] H. Wang, L. Li, H. Peng, H. Han, B. Song, Y. Wang, X. Gu, and Z. Liang (2013). A novel computer aided detection scheme for colonic polyps based on the structure decomposition. *Abdominal Imaging. Computation and Clinical Applications.* 63–72.
- [33] Z. Wang, Z. Liang, L. Li, X. Li, B. Li, J. Anderson, and D. Harrington (2005). Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy. *Medical Physics.* **32** 3602–3616.
- [34] S. J. Winawer, R. H. Fletcher, L. Miller, F. Godlee, M. H. Stolar, C. D. Mulrow, S. H. Woolf, S. N. Glick, T. G. Ganiats, J. H. Bond, L. Rosen, J. Zapka, S. G. Olsen, F. M. Giardiello, J. E. Sisk, R. Van Antwerp, C. Brown-Davis, D. A. Marciniak, and R. J. Mayer (1997). Colorectal cancer screening: clinical guidelines and rationale. *Gastroenterology.* **112** 594–642.
- [35] H. Xia and C. H. Hoi (2013). MKBoost: a framework of multiple kernel boosting. *IEEE Transactions on Knowledge and Data Engineering.* **25(7)** 1574–1586.
- [36] J. Yee, G. Akerkar, R. Hung, A. Steinauer-Gebauer, S. Wall, and K. McQuaid (2001). Colorectal neoplasia: performance characteristics of CT colonoscopy for detection in 300 patients. *CA: A Cancer Journal for Clinicians.* **219(2)** 685–692.

- [37] H. Yoshida, Y. Masutani, P. Maceneaney, D. Rubin, and A. Dachman (2002). Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: pilot study. *Radiology*. **222(2)** 327–336.
- [38] H. Yoshida and J. Nappi (2001). Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. *IEEE Transactions on Medical Imaging*. **20(12)** 1261–1274.
- [39] H. Zhu, C. Duan, P. Pickhardt, S. Wang, and Z. Liang (2009). Computer-aided detection of colonic polyps with level set-based adaptive convolution in volumetric mucosa to advance CT colonography toward a screening modality. *Cancer Manage Res*. **1** 1–13.
- [40] H. Zhu, Y. Fan, and Z. Liang (2011). Improved curvature estimation for shape analysis in computer-aided detection of colonic polyps, virtual colonoscopy and abdominal imaging. *Computational Challenges and Clinical Opportunities*. 9–14.
- [41] H. Zhu, Z. Liang, P. J. Pickhardt, M. A. Barish, J. You, Y. Fan, H. Lu, E. J. Posniak, J. R. Richards, and H. L. Cohen (2010). Increasing computer-aided detection specificity by projection features for CT colonography. *Medical Physics*. **37** 1468–1481.
- [42] H. Zhu, Z. Liang, P. J. Pickhardt, and R. M. Summers (2010). Increasing computer aided detection specificity by projection features for CT colonography. *Medical Physics*. **37** 1468–1481.

MING MA
DEPARTMENT OF COMPUTER SCIENCE
STONY BROOK UNIVERSITY
STONY BROOK, NY
USA
E-mail address: minma@cs.stonybrook.edu

LIHONG LI
DEPARTMENT OF ENGINEERING SCIENCE AND PHYSICS
CITY UNIVERSITY OF NEW YORK
AT COLLEGE OF STATEN ISLAND
STATEN ISLAND, NY
USA
E-mail address: lihong.li@csi.cuny.edu

HAO HAN

DEPARTMENT OF RADIOLOGY
STONY BROOK UNIVERSITY
STONY BROOK, NY
USA

E-mail address: haohan@mil.sunysb.edu

YIFAN HU

DEPARTMENT OF RADIOLOGY
STONY BROOK UNIVERSITY
STONY BROOK, NY
USA

E-mail address: yifanhu@mil.sunysb.edu

XIANFENG GU

DEPARTMENT OF COMPUTER SCIENCE
STONY BROOK UNIVERSITY
STONY BROOK, NY
USA

E-mail address: gu@cs.sunysb.edu

ZHENGRONG LIANG

DEPARTMENT OF RADIOLOGY
STONY BROOK UNIVERSITY
STONY BROOK, NY
USA

E-mail address: jerome.liang@sunysb.edu

RECEIVED 29 APRIL 2015