Effective 3D Object Detection and Regression Using Probabilistic Segmentation Features in CT Images

Le Lu   Jinbo Bi   Matthias Wolf   Marcos Salganicoff

CAD & Knowledge Solutions, Siemens Medical Solutions, Inc.

Malvern, PA 19355, USA

le-lu@siemens.com

Abstract

3D object detection and importance regression/ranking are at the core for semantically interpreting 3D medical images of computer aided diagnosis (CAD). In this paper, we propose effective image segmentation features and a novel multiple instance regression method for solving the above challenges. We perform supervised learning based segmentation algorithm on numerous lesion candidates (as 3D VOIs: Volumes Of Interest in CT images) which can be true or false. By assessing the statistical properties in the joint space of segmentation output (e.g., a 3D class-specific probability map or cloud), and original image appearance, 57 descriptive features in six subgroups are derived. The new feature set shows excellent performance on effectively classifying ambiguous positive and negative VOIs, for our CAD system of detecting colonic polyps using CT images. The proposed regression model on our segmentation derived features behaves as a robust object (polyp) size/importance estimator and ranking module with high reliability, which is critical for automatic clinical reporting and cancer staging. Extensive evaluation is executed on a large clinical dataset of 770 CT scans from 12 medical sites for validation, with the best state-of-the-art results.

1. Introduction

Robust, highly sensitive 3D object (e.g., cancer tumors) detection and its importance staging/ranking are the key computer vision components to develop a semantically useful tool for computer aided diagnosis (CAD). We propose a stratified learning framework including (supervised) object-specific image segmentation, segmentation feature extraction, robust object classification and importance regression (taking into account the segmentation ambiguity and uncertainty). Segmentation has been extensively explored for various medical imaging purpose [27, 15, 29, 8, 22, 11, 24], but explicit descriptive feature extraction and analysis on statistically characterizing segmentation outputs, for object classification and robust staging of cancer (as estimate of object key attributes), has not been much studied. In this paper,

we focus on finding and analyzing polyps[1], the precursors of colon cancer, but the proposed method can be extended to other medical imaging applications (e.g., lung nodule detection), or generic object segmentation and detection tasks in 3D range, LIDAR or spatial-temporal volumetric data.

Our algorithm use object-specific Figure-Ground segmentation results (i.e., probabilistic 3D map or clouds) as input for separately encoding or wrapping the object-image information for classification and analysis. Explicit or implicit figure-ground segmentation [18, 6, 10, 20, 22, 12] has shown of being capable to improve object recognition and detection problems in both computer vision and medical imaging. Particularly, the seminal work [10] develops a seamlessly interleaved segmentation and detection process to iteratively refine both segmentation and detection in a loop. [22] follows the similar "interleaved-looping" strategy and apply to lung nodule detection in 3D CT images. In our work, however, segmentation process and segmentation feature extraction, detection and regression steps are separate, consecutive building blocks in a flow-sequence.

Giving a per-voxel soft foreground/background segmentation map (as each pixel with the object-class posterior probability $[0, 1]$ assigned by the segmentor; close to 1 means more foreground class) and the original image, we try to classify if the segmented foreground is the object to be detected, in the joint image and segmentation probability space. As shown in Fig. 1, our algorithmic flow is a straightforward process, with no loops needed. In other words, the proposed segmentation feature extraction, classification and regression techniques are not fully dependent on a specific segmentation algorithm tuned for a specific object class (different from interleaved approaches [10, 20, 22]). The segmentation process and output probability map must be sensitive to different object classes (e.g., colon polyp, lung nodule, emphysema, or other diseases/tumors) or sub-categories (e.g., sessile, pedunculated,

---

[1]Colon cancer is the second leading cause of cancer death in western countries, but it is one of the most preventable of cancers because doctors can identify and remove its precursor known as a polyp. 3D Computed Tomography Colonography (CTC) has emerged as a powerful screening tool for polyp detection and the field of computer aided detection (CAD) of colonic polyps in CTC is highly intensive [2].
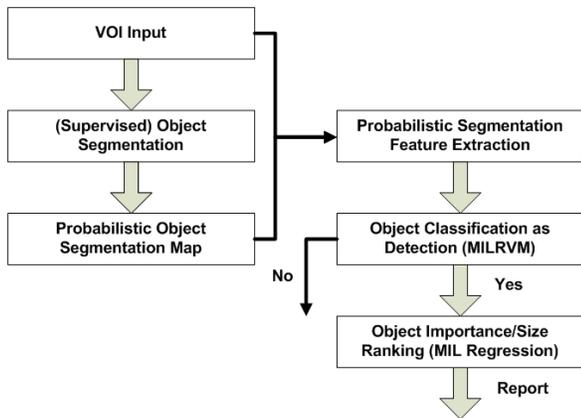
Figure 1. Flow-chart of the staged object/polyp segmentation, classification/detection, and size/importance regression process.

flat polyps; and solid, partial solid and GGO nodules), while our method is not, but applicable, as a more general segmentation+appearance feature extraction and classification/regression process.

Polyp segmentation in CT images is defined as identifying and isolating polyp voxels from non-polyp contexts (air, soft-tissue, colonic wall and tagging material), using a 3D mask in volumetric space, or a closed 1D curve boundary on 3D colonic surface. It can be heuristic, non-probabilistic [15, 29, 8, 24], or data-driven learned and probabilistic [12]. In [12], a compositional polyp segmentation framework is proposed (locating possible polyp tips; finding inside/outside polyp surface voxels; and optimizing polyp boundaries), by supervisedly learning the medical experts' knowledge as the annotated polyp boundary curves in a database. It also generates per-voxel polyp class probability for each segmented polyp voxel. Examples of the boundary based segmentation and probability maps on polyps with different shape morphologies, and under various surface contexts, are shown in Fig. 2. We choose and implement [12] as our object/polyp segmentor, due to its good segmentation accuracy, generality over polyp variations, and intrinsic probability output which represents the segmentation uncertainty and ambiguity. The non-probabilistic anatomical object segmentation approaches in medical imaging can give heuristically defined, hard segmentation sizes for analyzed lesions. We exploit 57 segmentation descriptive features in six groups of cues (in the joint probability and image space): polyp shapeness and dimension statistics, segmentation posterior probability statistics, Multi-resolution segmentation boundary smoothness, spatially-banded probability and area statistics, 3D Ellipsoid based shape descriptors, and multiscale intensity histogram features. The proposed feature extraction approach from segments of probability and image appearance is relevant with generic feature design [9, 19, 14] for other computer vision problems.

Our framework does not require perfectly accurate object-level segmentation which can be very challenging. Since the ultimate goal is for object detection (i.e., detecting a cancer VOI and staging), not on improving segmentation accuracy, we focus on deriving and fusing informative or

discriminative statistics as features, drawn from a noisy segmentation output. Though [12] is the state-of-the-art polyp segmentor, spurious polyp-class probability responses are still visible in some polyp VOIs, in Fig. 2. Furthermore, it performance on negative VOIs (non-polyp) is uncontrolled because supervised segmentation classifiers can not be trained from negative populations (There is no annotation or definition on how to segment negative anatomic structures, unless used for bootstrapping.). Thus image segmentation is primarily used as a probing process to find observations on positive/negative populations, for object classifier and regressor.

On the other hand, our classifier (MILRVM: multiple instance relevence vector machine) and regressor (section 3) can be trained to learn and handle the bottom-up segmentation/feature noise or bias, in a hierarchical learning sense. For example, not directly (semantically) relevant features (e.g., segmentation boundary smoothness) can be integrated to improve the estimation of object size (for staging) in regression, as they may be statistically correlated. In a similar spirit, [19, 14] use image regional features, or Gestalt grouping features to classify edges as object or non-object boundary classes in natural images.

Our contribution are three folds. First, we propose a set of probabilistic segmentation (PSM) features, serving as an intermediate-level object representation for object recognition and attribute regression purposes[2]. Second, we exploit a new probabilistic ridge regression model on object (polyp) size estimation which can handle the multiple instance setting, essential for CAD problems[4]; and a soft-gating classification framework over MILRVM for polyp detection. Third, the validity of new feature set, regression model and gating classification architecture is demonstrated on a comprehensive clinical dataset of 770 CT scans, collected from 12 hospitals in US, Europe and Asia.

2. Features & Algorithm

In this section, we first briefly review the polyp segmentation algorithm [12], for self-contained content. Then the detailed derivation of 57 probabilistic segmentation (PSM) features in six subgroups, ranging from statistics of dimensions, shapeness, gloabl-local probability distributions to boundary smoothness and multiscale histogram, is presented.

2.1. Supervised Probabilistic Polyp Segmentation

VOI Proposal: To generate polyp-like candidates (VOIs) from a 3D CT scan, various heuristics on symmetric, spherical or semi-spherical curvature pattern aggregation [15, 29, 8, 24] can be exploited on colon surface. Though many polyps are not spherical in a global object shape sense, they may still have some spherical local surface patches as parts, to be identified by Candidate Generation (CG) procedure. Then VOIs proposed by CG will

---

[2]Previous work mainly explore on directly observable, low-level intensity, texture and curvature features [15, 29, 8, 24, 22, 25, 27] in medical imaging.
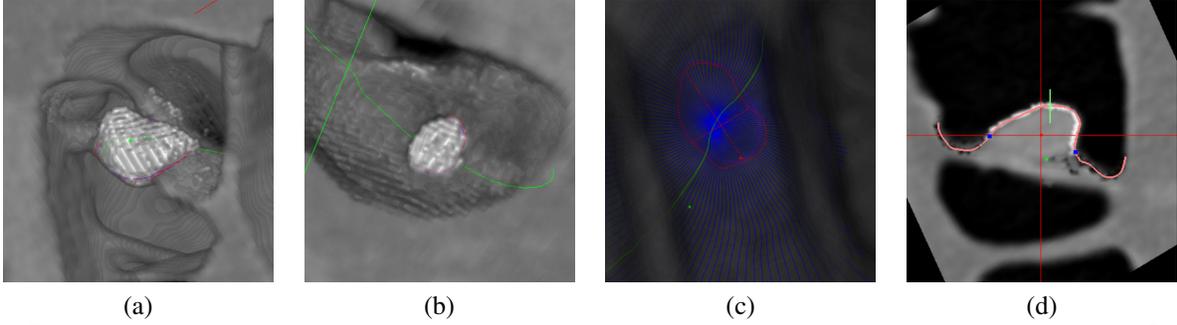
| (a) | (b) | (c) | (d) |

Figure 2. Illustrative examples of polyp-class voxel labeling probability responses and contour boundary segmentation on 3D colonic surface in CT images, using [12]. Polyps can have various shape morphologies and ambiguous surface/intensity contexts thus spurious polyp-class probability responses are still visible in all above segmentation maps. (a,b,c) are 3D colon volumetric renderings where surface voxels with higher polyp posterior probabilities are enhanced by higher intensities, and the purple and red-colored curves presents the radiologist annotated or computer segmented polyp boundary contours, respectively. The green or blue curves are the radial axes of polar coordinates [12]. (d) is a 2D projection view and the orange curve is the current polar coordinates axis, with two blue dots marking the polyp boundary by segmentation algorithm. (a) a $15.0mm$ irregular-shaped polyp on a colon fold; (b) a $9.1mm$ sessile polyp on colon wall; (c) a $7.6mm$ lobulated flat polyp on colon wall; (d) a $13.8mm$ pedunculated polyp on a colonic fold.

be inputed into the pipeline in Fig. 1. Our detection by classification is also a rare-event cascade detection [26], as the number of VOIs are normally two orders of magnitude more than the true object (polyp). To achieve high sensitivity on detecting positives, CG cascade produces a large portion of negatives ($> 99\%$, triggered by a large variety of colonic anatomical structures resembling polyp-like shapes, e.g., haustral folds, Ileo-Cecal Valve), which makes the object classification task challenging. For Lung nodel CG, 3D multiscale Gaussian and DOG (Derivatives of Gaussian) filters [22] can be used to pick round-like structures through template matching.

A hierarchical, three-staged supervised learning architecture is employed for polyp segmentation [12] where only true polyp VOIs are considered. First the polyp tip is estimated by labeling and grouping the polyp-tip probability by scanning all surface voxels using a trained classifier. A surface polar coordinate system is then fitted centering the located tip as its origin and sampling the surrounding colonic surface with a set of spatially-evenly distributed radial curves/axes. Second the polyp boundary learning is performed through two-layered stacked learning: scanning a small $7 \times 7 \times 7$ cubic volumetric window centered at each voxel on each polar axis to give its polyp-class probability $\wp$,; followed by running a 1D curve parser on the sequence of $\{\wp\}$ along each ith polar axis to determine the cutting boundary point $\hat{B}_i$ with bi-partitioning confidence ρ_i, to separate the interior portion of polyp and exterior portion of non-polyp region. Third the polyp boundary contour (1D surface curve) is formed by connecting the boundary points according to the axis proximity of polar coordinates.

The segmented polyp voxels are represented as a set of $\hat{\mathbb{S}}\{v_{ij}\} = \{v_{ij}\} : (PC(v_{ij}) \leq \hat{B}_i)$. The remaining polyp dimensioning process of finding three sizes ($Diameter_i, i = 1, 2, 3$) considering $\hat{\mathbb{S}}$ as a volumetric mass, can be referred from [12]. Note that [12] and our implementation are both trained on $200 \sim 300$ clean-prep polyps (easier for radiologists to annotate polyp segmen-

tation boundaries than tagging-prep polyps with possible artifact-coatings). However, we apply the trained segmentor on both positive and negative VOIs (as false positives from CG process) for detection purpose, in more challenging but more clinically important and popular tagging-prep datasets.

2.2. Probabilistic Segmentation (PSM) Features

Our probabilistic segmentation (PSM) features compute various types of intermediate-level statistics, capturing the polyp segmentation process uncertainty and confidence distributions in 3D space. This "descriptor of segments of probability+appearance" aggregates on low level, per-voxel polyp-class labeling probability map and the original object image appearance. There is no need to set hard thresholds, apart from previous work [15, 29, 8]. We expect that PSM features can effectively discriminate true objects (polyps) out of various candidates in CG. The estimated polyp dimensions ($Diameter_i, i = 1, 2, 3$) [12], are denoted as (D_1, D_2, D_3) for conciseness. Each classified polyp voxel of $\hat{\mathbb{S}}\{v_{ij}\} = \{v_{ij}\} : (PC(v_{ij}) \leq \hat{B}_i)$ is represented as a tuple of ($\wp_{ij}, I_{ij}, (x_{ij}, y_{ij}, z_{ij})$) where $\wp_{ij}$ is the polyp interior posterior probability, and $I_{ij} = I(x_{ij}, y_{ij}, z_{ij})$ and (x_{ij}, y_{ij}, z_{ij}) are the intensity value of v_{ij} and its 3D volumetric coordinates, respectively. We compute the following *six subgroups of PSM features* as follows.

Statistics of polyp dimensions:(9) This group of 9 size related features include three diameters (D_1, D_2, D_3) [12] and three other composed features: ($D_1 \times D_2 \times D_3$) to approximate the polyp volume; ($D_1 \times D_2$) to describe the area of 2D polyp base on colon surface; D_3/D_1 to indicate the flatness or plateness of polyp as a simple *height-length ratio*; and three nonlinearly expanded features ($D_2/D_1, D_1/D_2, D_1/D_3$) to better fit into the linear classifier[3] MILRVM we will use. We can also incorporate the shapeness descriptor for faint or light blob-like, light

---

[3]This is analogical to the nonlinear feature expansion/mapping in nonlinear support vector machines.

tubular-like and faint plate-like structures in medical images [25, 22] or ballness, stickness and plateness features [5] in Space-Time shape action modeling, as functions of (D_1, D_2, D_3). We leave this as future work.

Statistics of polyp probabilities $\{\wp\}$**:(3)** This feature group computes three overall confidence of $\{\wp_{ij}\}$ on the final segmented polyp surface $\hat{\mathbb{S}}$ (negative VOIs may have empty or non-empty $\hat{\mathbb{S}}$). $ProbSum$ is the sum of polyp-class posterior probabilities $\sum_{i,j} \wp_{ij}$ within segmentation $v_{ij} \in \hat{\mathbb{S}}\{v_{ij}\}$, similar to the data-model fidelity energy term in conditional random field formulation; $Area$ is the area of the segmented polyp surface as $count(\hat{\mathbb{S}}\{v_{ij}\})$ in voxel counts; and $ProbAvg$ is the averaged polyp probability or confidence $ProbSum/count(\hat{\mathbb{S}}\{v_{ij}\})$.

Multi-resolution polyp boundary smoothness:(5) Given the estimated polyp boundary points $\hat{B}_i, i = 1, 2, ..., 120$ with parsing confidence ρ_i in polar coordinates, we can measure the segmentation boundary smoothness term by

$$\text{BS}(inter) = \frac{\sum \|\hat{B}_i - \hat{B}_{i+inter}\| \times (\rho_i \times \rho_{i+inter})}{\sum (\rho_i \times \rho_{i+inter})} \quad (1)$$

where $\sum_{i=1,2,...,120-inter}$ and $inter$ is a parameter which controls how far the two boundary points are compared for smoothness. Multi-resolution smoothness is implemented by varying $inter = 2^i$ ($i = 0, 1, 2, 3, 4$) exponentially. True polyps are expected to have smoother boundaries whereas false polyps may return irregularly-rugged boundaries. This is inspired by Gestalt Perception Law that a good object-level segmentation should have smooth boundaries in natural images [19, 14].

Spatially banded probability and area statistics:(15) To capture the finer spatial granularity statistics, we divide the segmented polyp surface into K "circular belts", by uniformly shrinking the polyp boundary contour inwards from outskirt. It forms equal-width banded zones which share the same geometric centroid as polyp tip and are proportional to the segmented polyp boundary, in a similar spirit to different contour levels on topographical height map [28] of 3D surfaces, or log-polar spatial histogram in shape context [1]. For implementation, we compute the band number as

$$\text{BN}(v_{ij}) = ceil(v_{ij}/(B_i/5)) \quad (2)$$

where $ceil(\bullet)$ is a function to return the next bigger or equal integer of any given input, and contribute v_{ij} only to the band zone $\text{BN}(v_{ij})$ for features of probability sums $\{ProbSum_k\}$, areas $\{Area_k\}$ and averaged probabilities $\{ProbAvg_k\}$. Without loss of generality, we set $K = 5$ with 0th band at the core and 4th close to boundary, gradually. Band histogram pooling can also be implemented by general point-to-center geodesic distance binning, thus independent of polar coordinates [12].

3D Ellipsoid Shape Descriptor:(9) For the 3D voxel mass of $\hat{\mathbb{S}}\{v_{ij}\}$ per VOI, we first estimate its centroid and covariance matrix in *volumetric* coordinates using (x_{ij}, y_{ij}, z_{ij}).

$$[\bar{x}, \bar{y}, \bar{z}] = \frac{\sum_{i,j}[x_{ij}, y_{ij}, z_{ij}] \times \wp_{ij}}{\sum_{i,j} \wp_{ij}} \quad (3)$$

$$\text{CoMat} = \frac{\sum_{i,j}(\Delta X)^T(\Delta X) \times \wp_{ij}}{\sum_{i,j} \wp_{ij}} \quad (4)$$

where $\Delta X = ([x_{ij}, y_{ij}, z_{ij}] - [\bar{x}, \bar{y}, \bar{z}])$ Then, Singular Value Decomposition (SVD) is employed to calculate three Eigen-values of $CoMat$: R_1, R_2, R_3 that geometrically maps to the three radii if fitting the mass of $\hat{\mathbb{S}}$ as an ellipsoid. For comparison, D_1, D_2, D_3 are computed directly from $\hat{\mathbb{S}}$. The covariance matrix $CoMat$ models the *3D volumetric spatial distribution* of underlying polyp segmentation, including and beneath the colonic surface. Unique to regular Ellipsoid fitting, $\wp_{ij}$ is used as a weight factor in Eq. 3,4, to reflect per-voxel segmentation confidence or uncertainty. Assuming $R_1 \geq R_2 \geq R_3$, six other features $(R_1 \times R_2 \times R_3, R_1 \times R_2, R_1/R_2, R_2/R_1, R_1/R_3, R_3/R_1)$ are also computed for feature expansion purpose.

Multiscale Intensity Histogram Features:(16) By using the $\wp_{ij}$-weighted covariance matrix $CoMat$, we search all voxels $\{v\}$ of volumetric coordinates $[v_x, v_y, v_z]$, within the $\sqrt{L}$ *Mahalanobis* distance, originating from the ellipsoid centroid $[\bar{x}, \bar{y}, \bar{z}]$, i.e.,

$$\text{MHD}(v) = (\Delta V)(\text{CoMat})^{-1}(\Delta V)^T \leq L = 8 \quad (5)$$

where $\Delta V = [v_x, v_y, v_z] - [\bar{x}, \bar{y}, \bar{z}]$. L is set as $2, 4, 6, 8$, corresponding to fitted 3D object (polyp) ellipsoids with multiple spatial scales, while keeping their radii aspect ratios. A domain-knowledge based CT intensity binning of $[0, 350); [350, 950); [950, 1100); [1100, 2^{12}]$ is used to construct an intensity histogram $IH_k, k = 0, 1, 2, 3$ of voxels for each ellipsoid, where $[0, 350)$ stands for air, $[350, 950)$ for soft tissue, $[950, 1100)$ for fat and $[1100, 2^{12}]$ for bone or tagging materials. Polyps are mostly composed of soft tissues. A total of 16 features are calculated (4 bins by 4 scales) to model the intensity patterns of the object and its multiscale contexts. Similar analogy is applicable to analyzing other tumors in CT imaging. Deriving histograms in the image *Gradient* domain, such as 3D HOG (Histogram of Oriented Gradients [3]) features, is another option, but how to compactly parametrize 2D orientation histograms and achieve 3D rotation-invariance for effective object detection [23] is non-trivial. This is left for future work.

3. Probabilistic Multiple Instance Regression

In this section, we describe a robust regression model on object (polyp) size estimation for cancer lesion staging[4]. Many probabilistic segmentation features are statistically correlated in terms of dimensioning polyps, and they

---

[4]The lesion's true size is the clinical gold-standard indicating the lesion development, but very hard to estimate reliably. Staging critically affects the follow-up therapy treatment plan.

can be robustly fused through a statistical regression model, to estimate the polyp diameter more accurately than D_1 in [12]. Our method is developed and extended from ridge regression [7], by adding a probabilistic formulation and handling multiple instance learning. Suppose we have a set of samples (x_i, y_i), $i = 1, \cdots N$, where x_i is a d-dimensional vector and y_i is the size measured by radiologists. Denote X the feature matrix whose i-th row contains the features for the i-th data point, and y the label vector of N labels. The conventional linear ridge regression constructs a hyperplane-based function $w^T x$ to approximate the output y by minimizing the following loss function:

$$L_{RR}(w) = ||y - Xw||^2 + \lambda ||w||^2 \qquad (6)$$

where $|| \cdot ||$ denotes the 2-norm of a vector and $\lambda > 0$ is the regularization parameter. Here the first term is the squared loss of the output, the second term is the regularization term which penalizes the weight vector w with large norm, and parameter λ balances off the two terms. By zeroing the derivative of L with respect to w, it is not difficult to see that ridge regression has a closed-form solution $w = (X^T X + \lambda I)^{-1} X^T y$.

The regularization parameter λ is important for getting a good weight vector w. It is mostly tuned via a cross-validation procedure though there are some other ways of estimating λ in the ridge regression literature. Nevertheless, we present a probabilistic interpretation for the method and derive a principled way of adapting these parameters. Assume the output y_i follows a Gaussian distribution with mean $w^T x_i$ and variance σ^2, i.e., $y_i \sim \mathcal{N}(w^T x_i, \sigma^2)$, and the weight vector w satisfies a Gaussian prior distribution: $w \sim \mathcal{N}(0, I)$. Then the negative log-posterior density of w is exactly the $L_{RR}(w)$ as defined in Eq. 6, with $\lambda = \sigma^2$. One advantage of this interpretation is that one can optimize the regularization parameter $\lambda = \sigma^2$ by maximizing the marginal likelihood of the data, which is also called the evidence maximization [13] (or the type-II likelihood):

$$\log P(y|\sigma^2) = \quad -\frac{N}{2} \log 2\pi - \frac{1}{2} \log |XX^T + \sigma^2 I| \\ -\frac{1}{2} y^T (XX^T + \sigma^2 I)^{-1} y \qquad (7)$$

Alternatively, one can also derive an Expectation-Maximization algorithm, taking w as the missing data and σ^2 the model parameter. In this approach, we estimate the posterior distribution of w in the E-step, which is a Gaussian with

$$\begin{aligned} \mu_w &= (X^T X + \sigma^2 I)^{-1} X^T y, \\ C_w &= \sigma^2 (X^T X + \sigma^2 I)^{-1}. \end{aligned} \qquad (8)$$

Then in the M-step we maximize the "complete" log-likelihood with respect to σ^2, assuming the posterior of w as given in the E-step. This leads to the following update for σ^2:

$$\sigma^2 = \frac{1}{N} \left[||y - Xw||^2 + tr(XC_W X^T) \right] \qquad (9)$$

The final algorithm iterates the E-step and M-step until convergence. The posterior mean of w can be used to make predictions for test samples, and we can also report the variances of these predictions (by considering the posterior co-variance of w).

In addition, the proposed probabilistic ridge regression approach makes use of the specific observation in polyp dimensioning that *a polyp may be represented by multiple CG candidates*. Hence all these candidates will be labeled with the same size annotated by the radiologist. However, the candidates may represent fundamentally different segmentations based on their different locations on the polyp. It is desirable to use the candidate whose segmentation is the closest to the actual layout of the polyp and hence its segmented size approximates the diameter of the polyp with the best accuracy, to make the size prediction. Mathematically, it can be translated into "finding the candidate whose estimated size $f(x) = w^T x$ differs the least from y, i.e., $\min |y - f(x)|$". Hence, the objective of ridge regression Eq. 6 can be revised to the following function:

$$\min_w \sum_j \min_{i \in I_j} (y_j - w^T x_i)^2 + \lambda ||w||^2 \qquad (10)$$

where I_j corresponds to the index set of candidates that point to the same polyp j. We use an *alternating* algorithm to optimize this objective. It initiates all candidates as in ridge regression to learn a weight vector w. Then at each iteration, only the candidates x_i which have the predicted size $w^T x_i$ the closest to y_j are selected to construct the next w, until algorithm reaches a fixed point. A general feature selection approach can also be devised for the multiple instance regression formulation when replacing $||w||^2$ by its 1-norm counterpart $||w||_1$. Minimizing the 1-norm regularization condition leads to sparse w which relies only on few features.

4. Experimental Results

We report extensive experimental evaluations on tasks of discriminative feature analysis, polyp classification/detection, and polyp size regression and ranking. A new two-layer, hierarchical soft-gating tree classifier on polyp detection is also discussed. Note that an unseen clinical database of 770 Tagged-prep CT scans collected from multiple medical sites is employed, with 239 patient-level and 416 volume-level polyps.No data volumes from the same patient are used for both training and testing. Only clean-prep polyps are used for polyp segmentor training.

Polyp Detection by Classification: The Fisher Discriminant (FD) Score of any given feature f is defined as follows

$$J(f) = \frac{(\bar{f}^+ - \bar{f}^-)^2}{\sigma^2(f^+) + \sigma^2(f^-)} \qquad (11)$$

where $\bar{f}^+$ and $\bar{f}^-$ denote the mean; $\sigma^2(f^+)$ and $\sigma^2(f^-)$ present the covariance of feature f's distribution on positive $\{f^+\}$ (e.g., object, polyp) and negative $\{f^-\}$ classes of populations. In general, FD score describes the two-class separatibility using a single feature f, or the discriminative power of feature. The FD scores of PSM features with

ranks are given in Table 4. The group of locally banded features is more discriminative than global-scale features, e.g., $ProbSum_k$ features rank higher than the original estimated D_2, D_3, D_1 polyp sizes [12] and $ProbSum$. Refer to Fig. 6 for the kernel density plots of the highest ranked $ProbSum_0$ on positive and negative populations.

PSM features combined with MILRVM classifier [16] achieve better detection rates under the same false positive (FP) rate per volume[5], than without using PSM features (but original feature set in CAD), in both tagging-prep training and testing datasets. The improvement on training dataset is similar with the testing dataset and only testing performance is reported below (118 patient-level or 222 volume-level polyps), due to space limit. This demonstrates good generality across different datasets by the combination of classifier and features. In testing, with PSM feature integrated, our CAD system improves the per-patient polyp detection sensitivity from 82.20% to 84.75%. Furthermore, it significantly improves the detection rates of two most difficult polyp subcategories: from 76.92% to 82.05% for large, and from 67.35% to 71.43% for flat polyps, with ~ 3.5 FPs per-volume. For details, refer to the FROC curves in Fig. 3.

Soft-gating Classification: We next evaluate the PSM features using a two-layer, hierarchical tree classifier architecture, called "soft-gating" framework in a Bayesian "divide-and-conquer" setting. We first train a basis classifier (as MILRVM) using large ($\geq 10mm$) polyps versus intermediate size ($6 \sim 9mm$) polyps. In runtime, this so-learned "size classifier" can assign weights $Pb(large)$, $Pb(interm)$ for each VOI input (including negatives) as its probabilities of being large or intermediate polyp class. Two weights are further normalized as $Pb(large) + Pb(interm) = 1$. Then all positive/negative training VOIs are passed into both the left and right branches of the soft-gating tree, with different weights of $Pb(large)$, or $Pb(interm)$ respectively. A modified MILRVM classifier is applied again on each leaf to learn $Pb_L(prob)$ and $Pb_R(prob)$, which can handle weighted training samples. Finally, the probability of being "Polyp" for any candidate VOI is obtained in a Bayesian manner

$$Pb(Polyp) = Pb(large) \times Pb_L(polyp|large) + \\ Pb(interm) \times Pb_R(polyp|interm) \quad (12)$$

, different from hard cascading [26]. This binary tree classification framework is illustrated in Figure 4 (a). Other gating strategies other than size gating (e.g., object shape morphology gating) can also be exploited.

We further investigate the effectiveness of PSM features using within the size-gating tree classifier. Five PSM features (D_1D_2; $Area$; $Area_3$; $D_1D_2D_3$; D_1) are selected with dominating weights for the size-gating branch MILRVM classifier for learning $Pb(large)$, $Pb(interm)$, out of total 10 selected features. PSM features also significantly contribute to compose the left $Pb_L(prob)$ and

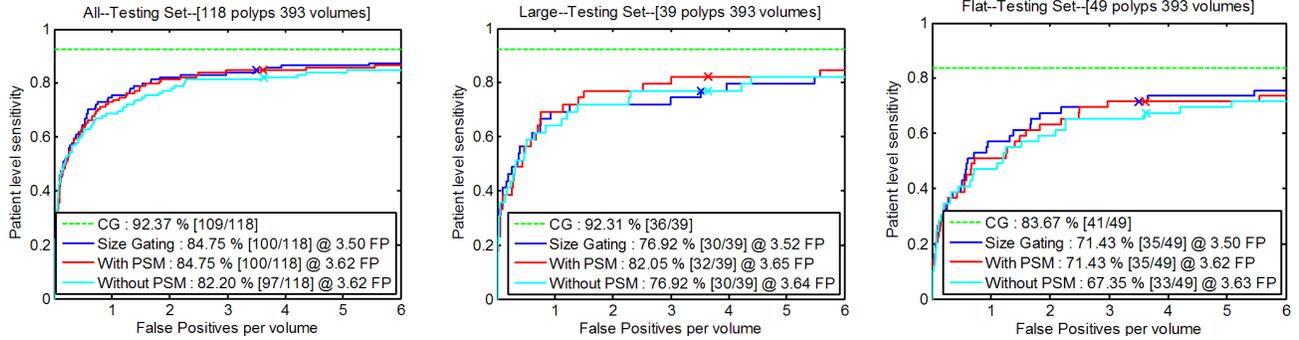| Feature	Rank	FDS	Feature	Rank	FDS
$ProbSum_0$	1	1.341	$ProbSum_3$	2	1.314
$ProbSum_2$	3	1.299	$ProbSum_1$	4	1.289
D_2	5	1.268	$ProbAvg_0$	6	1.172
$ProbSum_4$	7	1.167	D_3	9	1.132
$ProbAvg_1$	12	1.111	D_1	14	1.102
$ProbAvg_2$	15	1.092	$ProbSum$	17	0.977

Table 1. The FD score comparison of PSM features for polyp detection. Core band features are more discriminative than outer, boundary band features.

right $Pb_R(prob)$ leaf classifiers. MILRVM has an intrinsic feature selection and weighting mechanism on building classification margins/probabilities. Size based soft-gating achieves the best performance for our CAD system (than plain and shape-gating MILRVM), on an enhanced testing dataset of 412 CT volumes. The detection FROC curves of two subcategories of Sessile-Pedunculated (SP) and Flat polyps are shown in Fig. 4 (b,c). We obtain the detection rates of 95.0% and 88.0% for SP or Flat polyps respectively, under the same 3.36 FPs per-volume. The most competitive recent work [21] reports the detection sensitivity of 90% for SP, and $75 \sim 80\%$ for flat polyps, with 4.5 FPs per-scan on average[6]. [24] achieves 95% or 85% sensitivity rates, at 5 FPs per scan, on two 86 or 141 patient datasets, respectively.

Polyp Size Regression & Ranking: After classification, we first exploit the cross-correlation or Pearson's correlation coefficient between any single PSM feature and the annotated polyp size (in the range of $[3, 30]mm$) in the database. The correlation value is $\in [-1, 1]$, with higher value meaning better approximation. Among all ≥ 350 classification features (a superset of PSM), D_1 returns the highest correlation value as 0.6843 which validates good generality of the clean-prep trained polyp segmentation module for unseen data. For details, refer to Fig. 5 (a) where multiple instances per-polyp are shown. Second, the correlation coefficient of the estimated polyp size using our probabilistic regression model (section 3) and PSM features, improves to 0.8102 or 0.7174 for training or testing dataset respectively, as shown in Fig. 5 (b,c). The RMSEs also decreases from $4.3915mm$ (a) to $3.3072mm$ (b) and $3.6426mm$ (c), compared with using D_1 only. The final regressor is a linear combination of 16 weighted PSM features including D_1, $ProbSum$, boundary smoothness $BS(inter)$ and banded $Area_k$, $ProbSum_k$ features, *with automatic feature selection*. 185 volume-level polyps are used for training and 231 polyps for testing.
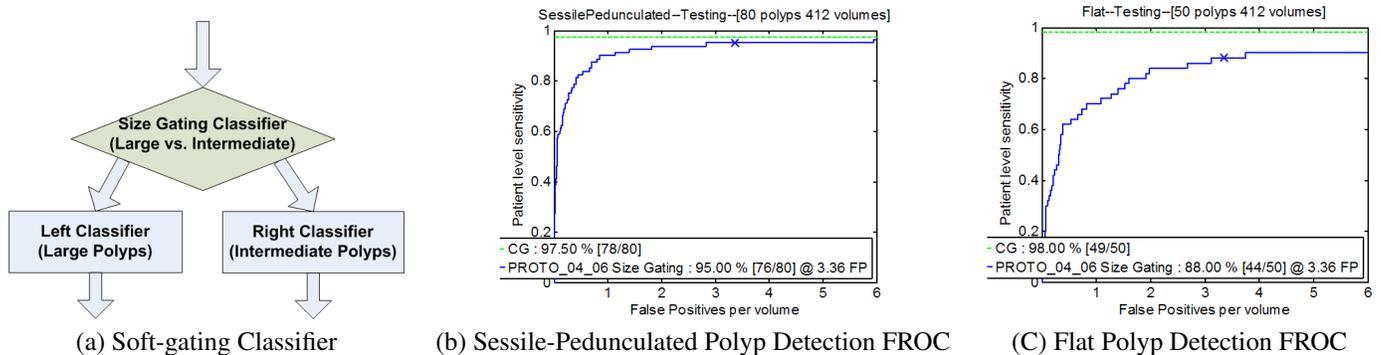
Finally, it is an important task for a useful CAD system to return detected objects/polyps in an order of their predicted importances. In training, we study the statistical rank consistency measurement of two lists, ranked by a PSM feature or the annotated polyp size respectively. A lot of group-wise nonparametric comparison statistics can be applied. Based on concordance index (CI) criterion [17] (as the probability

---

[5] We provide the patient-level sensitivity (i.e., if a polyp is detected in either view, it is counted as a true positive) as gold-standard in colon CAD and per-volume false positive rate where a single negative structure (e.g., a stool) can appear as two FPs in both Prone and Supine views.

[6] Note that we do not have access to the dataset in [21], but generally assume that our dataset is comparable to [21], as they both are in the order of hundreds of CT volumes and collected from a variety of clinical hispitals.

(a) FROC of All Polyps (Testing) (b) FROC of Large Polyps (Testing) (C) FROC of Flat Polyps (Testing)

Figure 3. Free Receiver Operating Characteristic (FROC) Curves of polyp detection, using MILRVM classifier [16] with and without PSM features, on the category of **ALL** actionable polyps (a), **Large** (39 patient-level or 75 volume-level) (b), and **Flat** (49 patient-level or 90 volume-level) (c) polyps, based on the study of the tagging-prep, testing dataset of 393 CTC scans. Green dot lines are the CG sensitivity upper bound and FROC curves are shown in Red (with PSM features), Blue (Size gating) and Cyan (without PSM).

(a) Soft-gating Classifier (b) Sessile-Pedunculated Polyp Detection FROC (C) Flat Polyp Detection FROC

Figure 4. Size-based soft-gating classification framework (a) and FROC curves of polyp detection using size-gating classification tree with PSM features incorporated, on two polyp subcategories of 80 (142 in volume-level) Sessile-Pedunculated (b) and 50 (89 in volume-level) flat polyps (c). Green dot lines are the CG sensitivity upper bound and FROC curves are shown in Blue.

of order concordance between the observed and predicted distributions), the multi-resolution segmentation boundary smoothness features BS(*inter*) surprisingly give the highest scores. A reasonable explanation is that segmentation on bigger polyps may produce more rugged boundaries because of larger expanding surface. Then PSM diameters, probability, area and banded statistics follow, with decreasing index values. Refer to table 4 for details.

Feature	Rank	CI	Feature	Rank	CI
$BS(1)$	1	0.870	$BS4$	2	0.861
D_1	3	0.782	D_2	6	0.750
D_1D_2	9	0.741	$Area_4$	9	0.718
D_3	11	0.714	$D_1D_2D_3$	12	0.713
$Area$	14	0.702	$ProbSum$	17	0.680

Table 2. Concordance Index of PSM features for polyp ranking.
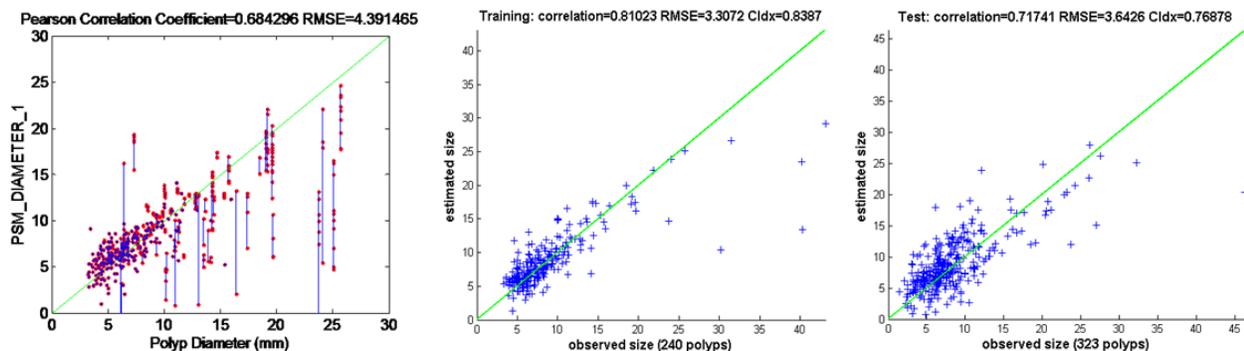
5. Conclusion

In this paper, we have presented a set of probabilistic segmentation (PSM) descriptive features, in the joint (object-specific) segmentation probability and image spatial space. The derived PSM features form an intermediate-level, object class descriptor using segmentation as a probing pro-

cess for object classification and detection. We demonstrate its validity in applications of colon polyp (object) detection, size regression and ranking, with the newly proposed Bayesian regression model and soft-gating classifier. The experiments are executed using a large multisite clinical dataset of 770 CT volumes. Our results statistically outperform the best other work reported [21, 24].

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–521, 2002. 1052

[2] L. Bogoni, P. Cathier, and et al. Cad for colonography: A tool to address a growing need. *The British Journal of Radiology*, 78:57–62, 2005. 1049

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. IEEE CVPR*, pages 886–893, 2005. 1052

[4] G. Fung, M. Dundar, B. Krishnapuram, and B. Rao. Multiple instance learning for computer aided diagnosis. In *NIPS*, pages 425–432, 2006. 1050

[5] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:2247–2253, 2007. 1052

[6] C. Gu, J. Lim, P. Arbelaez, and J. Malik. Recognition using regions. In *Proc. IEEE CVPR*, 2009. 1049

[7] A. Hoerl and R. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):5567, 1970. 1053

(a) $Diameter_1$ correlation (b) PSM regression size correlation (Training) (c) PSM regression size correlation (Testing)

Figure 5. Comparison of correlation between the polyp segmentation diameter D_1 [12] and annotated polyp size in (a), versus correlations using the sizes from our Probabilistic regression model, in training (b) and testing datasets (c).
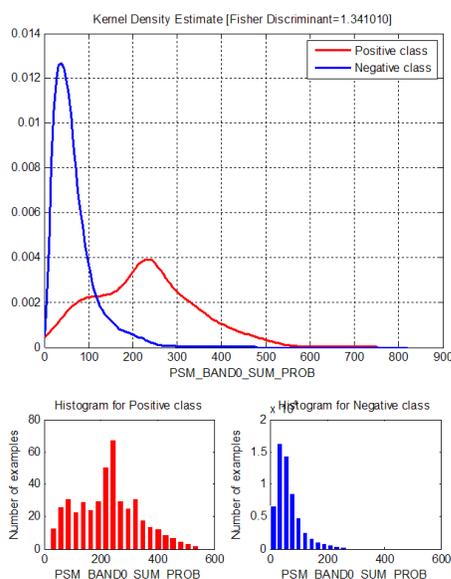

Figure 6. Local-scale banded feature $SumProb_0$ has the highest Fisher Discriminant (FD) Score of 1.3410, by evaluating a set of ≥ 350 classification features including PSM features. As comparison, the second polyp dimension D_2 [12] has the score of 1.2684 (higher than $D_1; D_3$), measuring the whole object-scale segmentation. A subgroup of "local", spatially banded features can outperform the discriminativeness of "global" object size features.

[8] A. Jerebko, S. Lakare, P. Cathier, S. Periaswamy, and L. Bogoni. Symmetric curvature patterns for colonic polyp detection. In *MIC-CAI*, pages 169–176, 2006. 1049, 1050, 1051

[9] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR*, 2006. 1050

[10] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *IJCV*, 2008. 1049

[11] H. Ling, S. Zhou, Y. Zheng, and et al. Hierarchical, learning-based automatic liver segmentation. In *Proc. IEEE CVPR*, 2008. 1049

[12] L. Lu, A. Barbu, M. Wolf, J. Liang, L. Bogoni, M. Salganicoff, and D. Comaniciu. Accurate polyp segmentation for 3d ct colonography using multi-staged probabilistic binary learning and compositional model. In *IEEE CVPR*, 2008. 1049, 1050, 1051, 1052, 1053, 1054, 1056

[13] D. MacKay. Ensemble learning and evidence maximization. In *NIPS*, 1995. 1053

[14] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color and texture cues. *IEEE Trans. PAMI*, 26(5): 530-549, 2004. 1050, 1052

[15] D. Paik and et al. Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical ct. *IEEE Trans. Med. Imag.*, page 23(6):661675, 2004. 1049, 1050, 1051

[16] V. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. Rao. Bayesian multiple instance learning: automatic feature selection and inductive transfer. In *ICML*, pages 808–815, 2008. 1054, 1055

[17] V. Raykar, H. Steck, B. Krishnapuram, C. Dehing-Oberije, and P. Lambin. On ranking in survival analysis: Bounds on the concordance index. In *NIPS*, pages 1209–1216, 2007. 1054

[18] X. Ren and C. Gu. Figure-ground segmentation improves handled object recognition in egocentric video. In *Proc. IEEE CVPR*, pages 886–893, 2010. 1049

[19] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proc. IEEE ICCV*, pages 10–17, 2003. 1050, 1052

[20] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling appearance, shape and context. *Int. J. Comp. Vis.*, 2009. 1049

[21] G. Slabaugh, X. Yang, X. Ye, R. Boyes, and G. Beddoe. A robust and fast system for ctc computer-aided detection of colorectal lesions. *Algorithms*, 3(1):21–43, 2010. 1054, 1055

[22] Y. Tao, L. Lu, M. Dewan, and et al. Multi-level ground glass opacity detection and segmentation in ct lung images. In *Proc. MICCAI*, pages 715–723, 2009. 1049, 1050, 1051, 1052

[23] Z. Tu, X. Zhou, L. Bogoni, A. Barbu, and D. Comaniciu. Probabilistic 3d polyp detection in ct images: The role of sample alignment. In *CVPR*, pages 1544–1551, 2006. 1052

[24] V. van Ravesteijn, C. van Wijk, F. Vos, R. Truyen, J. Peters, J. Stoker, and L. van Vliet. Computer aided detection of polyps in ct colonography using logistic regression. *IEEE Trans. on Med. Imag.*, 2010. 1049, 1050, 1054, 1055

[25] E. M. van Rikxoort, B. van Ginneken, M. Klik, and M. Prokop. Supervised enhancement filters: Application to fissure detection in chest ct scans. *IEEE Trans. Med. Imag.*, 27:1–10, 2008. 1050, 1052

[26] J. Wu, J. Rehg, and M. Mullin. Learning a rare event detection cascade by direct feature selection. In *NIPS*, 2003. 1051, 1054

[27] Y. Xu, M. Sonka, and et al. Mdct-based 3-d texture classification of emphysema and early smoking related lung pathologies. *IEEE Trans. Med. Imag.*, 25:464–475, 2006. 1049, 1050

[28] J. Yao, J. Li, and R. Summers:. Employing topographical height map in colonic polyp measurement and false positive reduction. *Pattern Recognition*, 42(6):1029–1040, 2009. 1052

[29] J. Yao, M. Miller, M. Franaszek, and R. Summers. Colonic polyp segmentation in ct colongraphy-based on fuzzy clustering and deformable models. *IEEE Trans on Medical Imaging*, 23(11):1344–1352, 2004. 1049, 1050, 1051