This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2017.2734329, IEEE Journal of Biomedical and Health Informatics

# Automatic Polyp Detection via A Novel Unified Bottom-up and Top-down Saliency Approach

Yixuan Yuan, *Student Member, IEEE*, Dengwang Li, *Member, IEEE*, Max Q.-H. Meng, *Fellow, IEEE*

***Abstract*—In this paper, we propose a novel automatic computer-aided method to detect polyps for colonoscopy videos. To capture perceptually and semantically meaningful salient polyp regions, we first segment images into multi-level superpixels. Each level corresponds to different sizes of superpixels. Rather than adopting hand-designed features to describe these superpixels in images, we employ sparse autoencoder (SAE) to learn discriminative features in an unsupervised way. Then a novel unified bottom-up and top-down saliency method is proposed to detect polyps. In the first stage, we propose a weak bottom-up (WBU) saliency map by fusing the contrast based saliency and object-center based saliency together. The contrast based saliency map highlights image parts that show different appearances compared with surrounding areas while the object-center based saliency map emphasizes the center of the salient object. In the second stage, a strong classifier with Multiple Kernel Boosting (MKB) is learned to calculate the strong top-down (STD) saliency map based on samples directly from the obtained multi-level WBU saliency maps. We finally integrate these two stage saliency maps from all levels together to highlight polyps. Experiment results achieve 0.818 recall for saliency calculation, validating the effectiveness of our method. Extensive experiments on public polyp datasets demonstrate that the proposed saliency algorithm performs better compared with state-of-the-art saliency methods to detect polyps.**

***Index Terms*—Saliency detection for polyp, Weak bottom-up (WBU) saliency, Contrast based saliency, Object-center based saliency, Strong top-down (STD) saliency.**

## I. Introduction

Colon cancer is a second leading cause of cancer deaths worldwide. The American Cancer Society estimates that 95,520 new cases of colon cancer will be added and 50,260 deaths will be caused from colon cancer in 2017 [1]. Since colon cancer usually begins many years earlier as small polyps, therefore it is important for clinicians to detect polyps in advance automatically and remove them before they deteriorate to cancer cells.

Colonoscopy is an important and widely used standard medical procedure in hospitals to detect polyps. Many works have been undertaken on automatic polyp detection. Generally, these approaches can be summarized into three main categories: shape-based methods, active contour models and machine learning methods. Shape-based approaches aim to search specific shapes that polyps commonly have in colonoscopy frames. These methods use low-level image features such as gradient, edges or valley information to obtain curvature values of boundaries [2]–[6]. Active contour models use a deformable model and let the contour evolve to minimize a given energy function [7]–[9]. But active contour methods usually depend on user interaction to supply an initial contour or set of parameters for weighting energy terms. Supervised learning methods are widely used in polyp detection. These methods first define specific features to describe polyps, and then use well-trained classifiers to classify spatial regions into abnormal polyp or normal tissues [10]–[13]. However, this kind of method requires large, diverse training datasets to ensure good detection performance, and such datasets are usually difficult to generate in medical domain.

Clinicians are able to distinguish polyps from the normal surrounding region because of their different characteristics. To model this visual process during automatic analysis, we argue that visual saliency computations can be used to detect polyps. The saliency models [14]–[22] have been applied for object detection and achieved promising segmentation results in natural images. Given an input image, the salient objects have high saliency values, while the background has lower values. The saliency method is advantageous for polyp detection since it eliminates the need for the training step of machine learning methods with large amounts of labeled images and the need for predefined parameters or user interaction that is usually required for the active contour models.

The saliency methods are usually divided into two categories: bottom-up and top-down approaches. The bottom-up saliency models are data-driven and rely on some predefined assumptions. The most common assumption is contrast, which means that salient objects are supposed to present various appearances compared with background. Itti et al. [14] calculated the intensity, color and orientation feature maps through evaluating corresponding center-surround differences between various Gaussian pyramid and oriented pyramid scales. Then the final saliency map was measured by unifying these feature maps together. Seo and Milanfar [23] utilized local regression kernels as features, and measured the similarity between a location and its neighboring regions to obtain the saliency map. Goferman et al. [24] proposed context aware saliency method. In this method, they calculated local and global saliencies by evaluating the similarity between each given patch and the surrounding patches locally and globally.

The top-down saliency models introduce prior knowledge acquired through supervised learning to detect image saliency. For instance, Jiang et al. [19] assumed the saliency detection as

Yixuan Yuan is with the Department of Radiation Oncology, Stanford University, Stanford, CA, USA (email: yxyuan@stanford.edu).

Dengwang Li is with the Shandong Province Key Laboratory of Medical Physics and Image Processing Technology, Institute of Biomedical Sciences, School of Physics and Electronics, Shandong Normal University, Jinan, China (email: dengwang@sdnu.edu.cn).

Max Q.-H. Meng is with the Department of Electronic Engineering, The Chinese University of Hong Kong, N.T., Hong Kong SAR, China (email: qhmeng@ee.cuhk.edu.hk).

a regression problem, and proposed a novel saliency detection method though learning latent relationships among numerous image features extracted from training samples and ground truth labels. Kanan et al. [20] constructed a saliency model by training a support vector machine (SVM) from learned natural image features. Lin et al. [25] considered three types of features (appearing frequency features, image information and the pixel location of features) and proposed a computational saliency method using the Bayesian probability theory and machine learning techniques.

While these two kinds of saliency methods achieve good performance to highlight important information in the image, they usually suffer from following four limitations for polyp detection. First, the existing saliency models usually compute saliency maps heuristically from low-level hand-crafted features. But the hand-crafted features usually could not be generalized to different images. The features that obtain good performance in natural image may not be suitable for polyp saliency calculation. Second, the bottom-up saliency maps normally tend to highlight object boundaries and fail to detect the whole target region uniformly [26], which could not achieve satisfactory results for polyp detection in colonoscopy videos. Thirdly, top-down saliency models conduct the saliency calculation based on large training samples with manual labels. In the medical image analysis field, labeling the huge number of images is tedious and time-consuming for clinicians. Therefore the large and diverse medical training datasets are not available, which makes the fully automatic polyp saliency calculation impossible. Moreover, although many saliency models have been developed recently, the saliency based polyp estimation for colonoscopy videos is not available yet. The existing saliency calculation methods designed for natural images could not achieve good performance for polyp images.

To deal with these difficulties, we develop a novel unified bottom-up and top-down saliency map to highlight polyps in colonoscopy images. The workflow is showed in Fig. 1. We first preprocess images to remove light spots, and utilize multi-level superpixel representation [27] to segment images. Rather than adopting hand-designed features to represent superpixel features as in previous works [2]–[6], we employ a kind of automatic feature learning method: sparse autoencoder (SAE) to learn powerful representation from calculated color, texture and shape features in an unsupervised way. We then propose a weak bottom-up saliency map (WBU) by fusing the contrast based saliency and object-center based saliency together. The contrast based saliency map is calculated by the distinctiveness between a certain image superpixel and the other regions while the novel object-center based saliency map is estimated by key points to predict the center of the polyp region. A set of training samples is collected from the obtained multi-level WBU saliency maps, where positive samples are selected from the salient objects while negative samples are pertaining to the image background. We thus propose to calculate a strong top-down (STD) saliency map by learning a solid classifier with Multiple Kernel Boosting (MKB) algorithm [28] from collected samples. Since the bottom-up saliency model detects local structure details and the top-down saliency model tends to highlight the global shape, these two models are fused together to generate the final polyp saliency map.

Our main contributions can be summarized in the following five aspects.

1) We propose a novel two-stage saliency method to detect polyps in colonoscopy images. Instead of using the traditional detection methods, such as shape-based methods, active contour models and machine learning methods, to detect polyps, we are the first to utilize saliency information to highlight polyps as far as we know.
2) Instead of directly using traditional hand-crafted features to calculate image saliency, we introduce an efficient feature learning method, named SAE, to learn high-level superpixel characterization from the hand-crafted features. Thus our approach is essentially different from numbers of existing saliency methods that depend on low-level image features such as color, texture, and shape cues.
3) A effective WBU saliency model is proposed with the contrast based saliency and object-center based saliency. The contrast based saliency highlights the distinct regions in images compared with surrounding areas and the object-center based saliency introduces a novel coarse object detection method to emphasize polyps. Thus the proposed bottom-up saliency model could highlight the whole saliency object accurately.
4) We introduce a novel STD saliency map for polyps with MKB by learning training samples selected from the multi-level WBU saliency maps. In this way, our method eliminates the limitation of existing top-down saliency models, which need large labeled samples. Moreover, we restrict this saliency learning process to a single image since we just used the selected superpixels from WBU saliency maps within a given image as training samples to calculate the corresponding polyp saliency. Therefore, our proposed method avoids the heavy computational cost.
5) Extensive experiments on the standard polyp image datasets illustrate that our proposed saliency method detects polyps well and outperforms state-of-the-art approaches significantly. We further evaluate each component of our model carefully and analyze the corresponding contribution to the saliency performance.

The rest of this paper is organized as follows. Section II reviews traditional feature extraction methods for polyps. Section III illustrates the image pre-processing step and feature extraction method of our proposed saliency model. The proposed first stage WBU and second stage STD saliency maps are illustrated in Section IV. The experimental saliency maps and extensive comparison results are illustrated and discussed in Section V. Finally, we draw conclusions in Section VI.

## II. Related Work with Feature Extraction

To characterize polyp information in colonoscopy images, three kinds of features (color, texture and shape) are usually utilized.

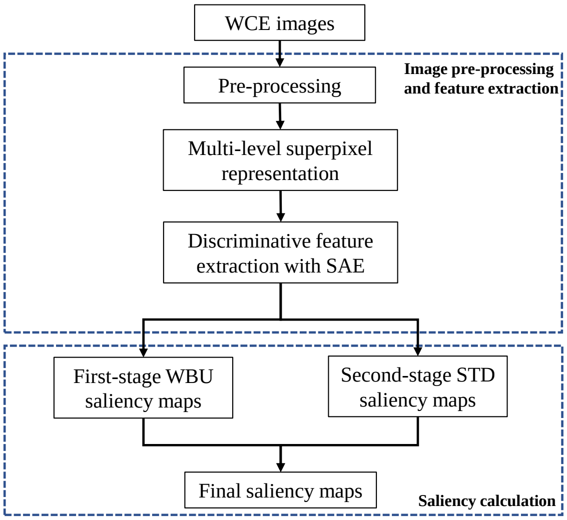

WCE images
Pre-processing
Image pre-processing and feature extraction
Multi-level superpixel representation
Discriminative feature extraction with SAE
First-stage WBU saliency maps
Second-stage STD saliency maps
Final saliency maps
Saliency calculation

Fig. 1. Workflow of our proposed saliency model for polyp images.

*1) Color feature:* To recognize the polyp region from normal tissues, we commonly extracted the color feature. The original polyp images is visualized in RGB color space, but different color spaces represent different image information. Choosing suitable color space is very important and convenient to highlight specific color information. The images in the method [4] are converted into LAB color space since this color space accurately stores color information and aids in processing color contrast relationships in polyp images. The color histogram and statistic information of color distribution are also usually used to characterize polyps.

*2) Texture feature:* Texture is another primary feature exploited by clinicians to recognize polyps from normal images. It usually includes the filter based features and the local binary pattern (LBP). The filter based features tend to remove noisy background parts and preserve important texture information. Good filters can highlight the polyp region and suppress the values in normal regions. The methods in [6] used discrete cosine transform to preprocess images. Yuji et al. [2] have utilized hessian filter to calculate texture information of polyp images. LBP descriptor [29] has been widely used to describe image texture features. In LBP, a local circular neighborhood is thresholded by the gray value of the central pixel and characterized by a LBP code. Then the statistical histograms within the image were generated as a texture descriptor. In method [10], LBP features were adopted in analysis of polyp images as the method of extracting texture feature.

*3) Shape feature:* Besides color and texture features, the polyps also show specific elliptic shape characteristics. Histogram of oriented gradients (HOG) [30] is a common shape feature that exploits the local gradient orientation information to characterize polyp images. The methods in [12], [31] utilized HOG to extract shape features for endoscopy images. Hafner et al. [11] utilized a combination of features, such as convex hull, perimeter, contrast feature to extract shape information.

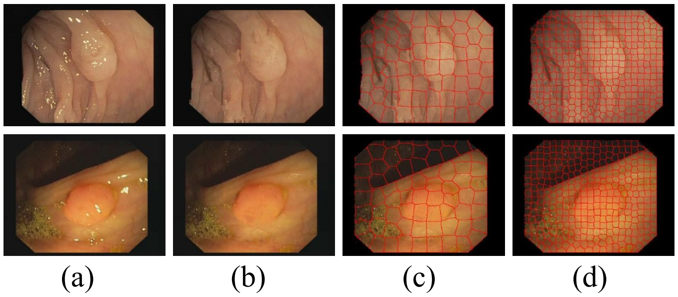


Fig. 2. Illustration of polyp images. (a) Original images, (b) Images after the preprocessing step, (c) Polyp images with 50 superpixels where the red lines define the superpixel boundaries, (d) Polyp images with 250 superpixels.

The color, texture and shape features can be extracted from a whole image or cropped image patches to describe images. But these handcraft designed features, which achieved good performance in natural image, may not characterize the polyp images well [32]. Therefore, we are motivated to learn meaningful and discriminative features of polyps with effective feature learning based method.

## III. Image Pre-processing and Feature Extraction

This paper proposes a novel saliency method to outline polyp regions. To better characterize meaningful salient regions, we first preprocess image to remove the specular reflections and then segment the image into multi-level superpixels. The corresponding color, texture, shape and gradient features of superpixel regions for each level are extracted. Specifically, different from hand-crafted feature representation based methods, our method utilizes a kind of deep learning method: SAE model to learn discriminative high-level robust features.

### A. Image Preprocessing

As shown in Fig. 2(a), the specular reflections frequently exist in polyp images. They appear as bright spots on surfaces with high reflectivity and would produce artificial boundaries during the saliency calculation. Therefore, in this paper, we first preprocess images to reduce the influence of specular reflections.

*1) Specular Reflections Detection:* The first step of preprocessing is to identify the candidates of specular reflections. Since the specular reflections usually have small the saturation and large intensity values [4], we transformed the polyp image in RGB color space to HSI (Hue, Saturation, Intensity) space to obtain the corresponding saturation and intensity values of images. Then for each image pixel in HSI color space, if its saturation is smaller than threshold $t_1$ and its intensity larger than threshold $t_2$, we designate this pixel as a high-confidence pixel. $t_1 = 0.29, t_2 = 0.65$ are chosen based on the reference [4]. Next, a dilation with a 5-pixel circular structuring element is performed on these high-confidence pixels. In this way, we identify the specular reflections.

*2) Image Inpainting:* Instead of inpainting the specular reflections by averaging nonspecular pixels within 8-neighbors of the specular pixel [4], we utilize a novel dynamic search-based inpainting algorithm with an adaptive window [33]. The algorithm starts by looking into each detected specular pixel at a time and examining its immediate 8-neighbors window. If it appears that more than six pixels within the window contain information, i.e. non-specular, then those pixels will be averaged and the result is assigned to the specular pixel. Otherwise, the window is moved in all directions counter clockwise until enough information is found. The images with the preprocessing step are shown in Fig. 2(b). We find that our preprocessing step could remove the influence of specular reflections well.

### *B. Superpixel Extraction*

In this paper, we utilize superpixel-based method to represent image information since it offers more useful spatial structure information compared with pixel-level methods. Simple linear iterative clustering (SLIC) algorithm [27] is utilized to obtain superpixels. We further utilize a multi-level superpixel method to preserve clear image boundaries, which first segments an image into multiple superpixels for different levels, then fuses saliency results of different superpixel levels together. In this paper, we evaluate five superpixel level for the saliency calculation of polyps and the numbers of superpixels for each level are 50, 100, 150, 200, and 250, respectively. Fig. 2(c-d) demonstrates examples of superpixel representation for polyp images with superpixel number of 50 and 250.

### *C. Feature Extraction*

After segmenting a polyp image into several superpixels, we extract corresponding features to calculate the polyp saliency in colonoscopy images. As discussed in Section II, the color, texture, shape information can characterize various and diverse contextual structures well [34], we thus utilize the feature set of the combination of color histogram, LM filter features [35], LBP features, multi-orientation Histogram of Gradient (HOG) to represent image information since they all demonstrate good performance on polyp characterization. The four types of features for a given superpixel $i$ are defined ${x_i}^{color}$, ${x_i}^{LM}$, ${x_i}^{LBP}$, ${x_i}^{HOG}$.

However, these features are hand-crafted features, which may have limited characterization power and may be insufficient to capture high-level information of complex images. Therefore, in this paper, we propose to utilize a symmetrical neural network: SAE to learn discriminative high-level features from obtained low-level features in an unsupervised manner. The core idea of SAE is to minimize the reconstruction error between the input data at the encoding layer and its reconstruction at the decoding layer. The structure of SAE is illustrated in Fig. 3.

We take color descriptors (${x_i}^{color}$) as an example to illustrate the process of SAE. The right superscripts for different features are dismissed for simplicity. In the encoding step, an input vector $x_i \in \mathbb{R}^M (i = 1,...,N)$ is mapped to a hidden representation $h = f(W_1 x_i + b_1)$, where $W_1 \in \mathbb{R}^{K\times M}$ is a weight matrix, $b_1 \in \mathbb{R}^K$ represents the encoding bias. $M$ defines the feature dimension while $N$ is the number of image superpixels. $f(z)$ denotes the logistic sigmoid function $(1+exp(-z))^{-1}$. In the decoding step, the hidden representation $h \in \mathbb{R}^K$ is decoded using another linear matrix $W_2 \in \mathbb{R}^{K\times M}$ as $\hat{x}_i = f(W_2^T h + b_2)$, where $b_2 \in \mathbb{R}^M$ is the decoding bias and $\hat{x}_i$ represents the reconstructed feature of $x_i$.

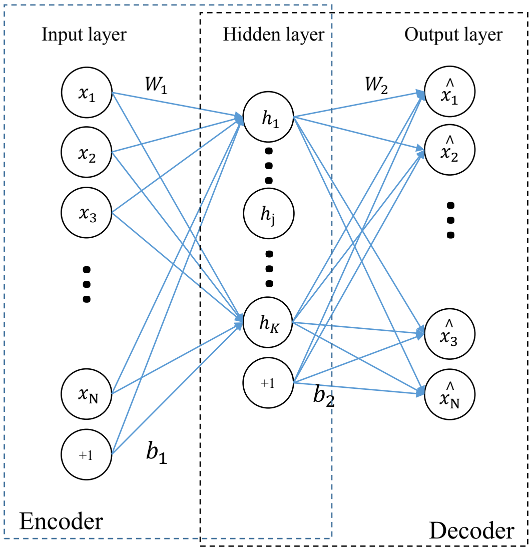


Fig. 3. Illustration of the workflow of the SAE model. N defines the number of superpixels while K denotes the feature dimension.

Then the objective of SAE is to minimize the following cost function with a sparsity constraint and learn parameters $W_1, W_2, b_1, b_2$ :

$$J_{SAE} = \frac{1}{2}\sum_{i=1}^{N}\|x_i - \hat{x}_i\|^2 + \frac{\alpha}{2}(\|W_1\|^2 + \|W_2\|^2) + \beta\sum_{j=1}^{K} KL(\rho\|\hat{\rho}); \quad (1)$$

where $\alpha$ defines the weight decay cost parameter, $\beta$ is the weight of the sparsity penalty.

The first term in Eq. (1) describes the discrepancy between input $x_i$ and reconstruction $\hat{x}_i$ over the entire data. The second term is a weight decay term, which tends to decrease the magnitude of the weight, and prevents from overfitting. The third term is a sparsity constraint based on Kullback-Leibler (KL) divergence [36] to enable sparse connections among layers in autoencoders. $K$ denotes the nodes number in hidden layer and controls the final feature representation dimension. The sparsity parameter $\rho$ is the target average activation of hidden units while $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}[h_j]$ defines the average activation over the training set. $KL()$ is used to measure the similarity between the desired and actual distributions as follows:

$$KL(\rho\|\hat{\rho}) = \rho log\frac{\rho}{\hat{\rho}} + (1-\rho)log\frac{1-\rho}{1-\hat{\rho}} \quad (2)$$

This penalty function has the property that $KL(\rho\|\hat{\rho}) = 0$ if $\rho = \hat{\rho}$, and otherwise, it increases monotonically as $\hat{\rho}$ diverges from $\rho$, which acts as the sparsity constraint.

By applying back-propagation method for the objective function Eq. (1), we obtain the parameters $W_1$ and $b_1$ to minimize the discrepancy among the input and reconstruction. Thus we can calculate the learned descriptive feature $h$. The SAE process could also be applied on LM filter features (${x_i}^{LM}$), LBP feature (${x_i}^{LBP}$) and HOG feature (${x_i}^{HOG}$). Thus, each superpixel of the image is represented by four discriminative descriptors ${h_i}^{Color}, {h_i}^{LM}, {h_i}^{LBP}, {h_i}^{HOG}$. These obtained features are learned directly from existing hand-crafted features, thus they can represent superpixel information well.

## IV. Saliency Calculation

In this paper, we propose a novel unified bottom-up and top-down saliency approach to detect polyp abnormalities. The first stage WBU saliency map is computed using the combination of contrast based saliency map among superpixels and object-center based saliency map. Then we collect a set of training samples for the second learning based saliency models. The positive samples are selected from saliency regions of previous WBU maps while negative ones are pertaining to the image background. With the collected training samples, we use multiple kernel boosting methods to learn strong saliency models and then apply this model on all superpixels to calculate the strong top-down (STD) saliency map. Finally, the proposed saliency map for polyp detection is obtained by unifying multi-level saliency maps.

### A. First Stage WBU Saliency Map

*1) Contrast based Saliency Map:* It is reported that a salient region shows different features compared with its neighboring regions [14], [24]. In our previous saliency study [35], we calculate the saliency value of a superpixel based on its contrast to the nearby superpixels. The inspiring saliency results demonstrate the effectiveness of our method. Therefore, the saliency of a superpixel is computed as the summation of its feature distances to all other superpixels, weighted by their spatial distances. The contrast based saliency $S_c^l$ of the $l$-th superpixel level is defined as

$$S_c^l(y,z) = \sum_{j=1, j\neq i}^{N} \frac{d_f(i,j)}{1 + d_{location}(i,j)}, \tag{3}$$

where $d_{location}(i,j)$ is the spatial Euclidean distance between the center of superpixel $i$ and the center of superpixel $j$. $N$ is the number of superpixels for a given colonoscopy image. $d_f(i,j)$ defines the Euclidean distance between the superpixel feature vectors $h_i$ and $h_j$. Since each superpixel in our paper is represented by four kinds of features ($h_i^{Color}$, $h_i^{HOG}$, $h_i^{LBP}$, $h_i^{LM}$), thus $d_f(i,j)$ is calculated by the sum of Euclidean distances among these four features between superpixel $i$ and superpixel $j$. The entry $(y,z)$ for the contrast based saliency $S_c^l$ represents pixels within the $i$th superpixel region. The saliency values of pixels in the same superpixel region are set to be same.

According to Eq. (3), a superpixel with different feature vectors compared with the other superpixels is assigned with a higher value. Those distant superpixels have smaller influences on the saliency value of the evaluated superpixel.

*2) Object-center based Saliency Map:* As shown in Fig. 4(b), the contrast based saliency map often incorrectly detects some background superpixels. Therefore, it is significant to introduce some principles to alleviate this problem. Our previous research utilized center prior to assign higher saliency values to regions near the image center [35]. But this prior information becomes invalid when the salient objects are located around the boundary of images. To deal with this problem, we propose to estimate the probable location of the salient region by coarsely enclosing key points within an image and then utilize the centroid of the estimated objects to get the object-center based saliency map.

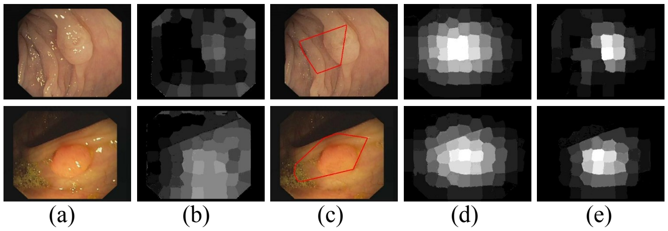


Fig. 4. Illustration of WBU saliency maps. (a) Original polyp images, (b) Contrast based saliency maps, (c) Estimated salient object boundary, (d) Object-center based saliency maps, (e) First stage WBU saliency maps.

Since key points represent most important points in the images and contain rich local information, thus the region near these points could characterize an image well. In our implementation, we utilize the color Harris point detector to compute key points and then enclose these points to provide a coarse polyp region estimation.

Given the center of the coarse polyp region $(y_o, z_o)$, the object-center based saliency $S_{oc}^l$ of $l$-th superpixel level is defined as:

$$S_{oc}^l(y,z) = exp(-\frac{\|y_c - y_o\|^2}{2\sigma_y^2} - \frac{\|z_c - z_o\|^2}{2\sigma_z^2}); \tag{4}$$

where $y_c$ and $z_c$ define the mean horizontal and vertical positions of the superpixel $i$, $\sigma_y^2$ and $\sigma_z^2$ denote the horizontal and vertical variances. In our experiment, we use a centered anisotropic Gaussian distribution to model the object-center prior map. We set $\sigma_y^2 = \sigma_z^2$ and these two parameters are calculated with the constraint that saliency values should be normalized to [0,1]. Since the boundary estimates rough locations of the salient objects, the object-center based saliency map could detect the polyps well.

*3) Integration of Saliency Map:* With the obtained contrast based saliency ($S_c^l$) and object-center based saliency ($S_{oc}^l$), we calculate the first stage WBU saliency map $S_1^l$ for superpixel level $l$ as follows,

$$S_1^l = S_c^l \times S_{oc}^l; \tag{5}$$

With this formula, we find that only the regions with higher values in both of these two saliency maps can obtain higher values in the final saliency map. Thus, the obtained saliency map enhances common salient regions detected by these two saliency maps.

Then the final first-stage WBU saliency map $S_1$ is calculated by fusing saliency maps of different superpixel levels together,

$$S_1 = \frac{1}{L} \sum_{l=1}^{L} {S_1}^l. \tag{6}$$

The first stage WBU saliency map fuses different superpixel levels of the contrast based saliency and object-center based saliency together, therefore it can highlight the fine detail information accurately.

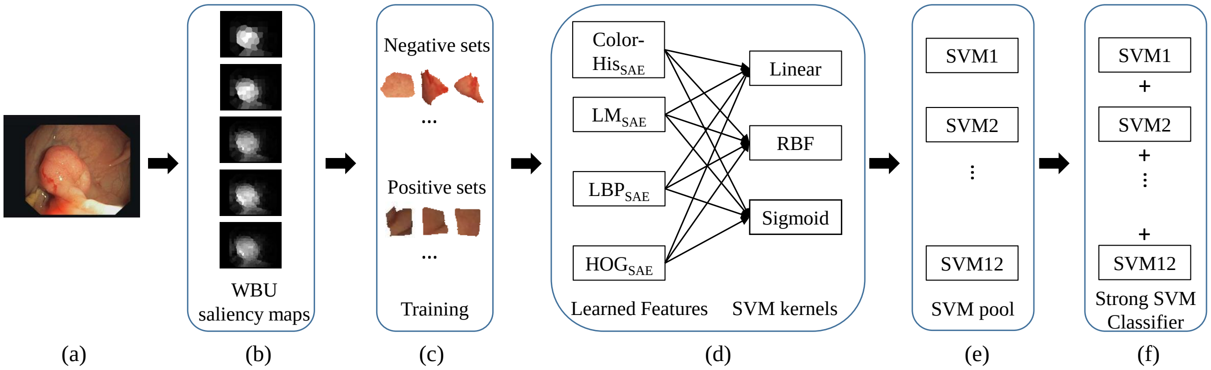


Fig. 5. Workflow of the proposed second stage STD saliency map. (a) Original polyp image, (b) Five-level WBU saliency map, (c) Selected training samples from five-level WBU saliency maps, (d) Different combinations of saliency learning, (e) SVM pooling strategy, (f) Strong SVM classifier by MKB with adaboost.

### *B. Second Stage STD Saliency Map*

Traditional top-down saliency models conduct the saliency calculation based on training samples with manual labels. However, amounts of labeled data are usually not available in medical field. Therefore, we propose to learn a robust saliency model from examples obtained through the first stage WBU saliency maps. We select the training superpixel sets from the five level WBU saliency maps. A superpixel is selected to be confident when its saliency value is larger than 80% of the mean value of the entire saliency map and its saliency score is set as 1. Instead, if a superpixel's saliency value is smaller than 25% of the mean value of the entire saliency map, it is a negative superpixel and the corresponding saliency score is -1.

With the collected training samples, the traditional SVM is able to predict the saliency value of each superpixel, and the learning based saliency map can be obtained. However, it is hard to determine the appropriate kernel for polyp datasets with various learned features from color histograms, LM filter features, LBP and HOG. To cope with this problem, we propose to use Multiple Kernel Boosting (MKB) [28] method to choose suitable kernels (linear, RBF, sigmoid) for different features. We consider SVMs with different kernels as weak classifiers and then learn a strong classifier using the boosting method. The workflow of the second stage saliency map is illustrated in Fig. 5.

Given a set of training samples $\{h_i, l_i\}^D$, where $h_i$ is the feature for $i$th sample, $l_i = \{\pm 1\}$ indicates the label of the sample and $D$ defines the total number of samples, the task of STD saliency map is to train a multi-kernel based classifier $F(h)$ to predict unlabeled samples as follows,

$$\begin{aligned} F(h) &= \sum_{m=1}^{M} \beta_m \sum_{i=1}^{D} \alpha_i y_i K_m(h, h_i) + b \\ &= \sum_{m=1}^{M} \beta_m (\alpha^T K_m(h) + b_m) = \sum_{m=1}^{M} \beta_m z_m(h) \qquad (7) \\ & with\ \beta_m \geq 0, \sum_{m=1}^{M} \beta_m = 1 \end{aligned}$$

where $K_m(h, h_i)$ defines the kernel and $\beta_m$ is the kernel weight. $\alpha = [\alpha_1 y_1, \alpha_2 y_2, ... \alpha_D y_D]$ in which $\alpha_i$ is the Lagrange multiplier and $b = \sum_m^M b_m$, where $b$ denotes the bias in the standard SVM algorithm. $M = N_f \times N_k$ denotes the number of weak classifier, where $N_f$ defines the types of features and $N_k$ indicates the number of kernels. In our paper, $M = 12$ with $N_f = 4, N_k = 3$.

We treat each SVM as a weak classifier and our objective is to learn the final strong classifier $F(h)$, which is the weighted combination of all the weak classifiers $z_m(h)$. The decision function of this boosting algorithm can be rewritten as

$$F(h) = \sum_{l=1}^{L} \beta_l z_l(h), \qquad (8)$$

where $L$ indicates the total number of iterations. In this way, MKB problem will be solved by adaboost method.

We start to train each SVM with uniform weights, $w_1(i) = 1/D, i = 1, ..., D$. At the $l$-th iteration, the classification error for each of the weak classifiers is calculated by $\varepsilon_m = \frac{\sum_{i=1}^{D} w_l(i)|z_m(h_i)|U(-y_i z_m(h_i))}{\sum_{i=1}^{D} w_i|z_m(h_i)|}$, where $U(x)$ is the sign function, which equals to 1 when $x > 0$ and 0 otherwise. We select a SVM classifier that gives the smallest weighted classification error $\varepsilon_l$ on the entire dataset and set the corresponding weight of this SVM classifier as $\beta_l = \frac{1}{2} log \frac{1-\varepsilon_l}{\varepsilon_l}$. If the weight $\beta_l$ is smaller than 0, the iteration procedure terminates since even the best SVM classifier performs worse. Otherwise, the selected classifier is added to the decision function. Then we update the weights for the training samples as $w_{l+1}(i) = \frac{w_l(i)}{2\sqrt{\varepsilon_l(\varepsilon_l - 1)}} e^{-\beta_l y_i z_l(h_i)}$, where incorrectly classified samples are assigned larger weights in the next iteration.

We continue $L$ iteration utill all the $\beta_l$ and $z_l(h)$ are computed. Therefore we obtain the boosted classifier as in Eq. (8) consisting of a number of single kernel SVM classifiers. We then apply this strong classifier to the test samples (all superpixels from an input image), and five-level STD saliency maps are generated. The final second-stage saliency map $S_2$ is calculated by integrating different level saliency maps together,

$$S_2 = \frac{1}{L} \sum_{l=1}^{L} {S_2}^l. \qquad (9)$$

where $S_2^l$ defines the $l$th level of the STD saliency map. Our proposed STD saliency model could obtain good performance since it is constructed based on the MKB algorithm which learns a strong classifier by combining weak ones using the Adaboost algorithm.

### C. Final Saliency Fusion

The first stage WBU saliency map tends to detect local fine information based on the contrast-based measure. In contrast, the second stage STD saliency map highlights global shapes well. Thus the proposed saliency map for polyps is calculated by a weighted combination as following,

$$\mathbf{S}_{final}(x,y) = \gamma S_1(x,y) + (1-\gamma)S_2(x,y) \tag{10}$$

where $\gamma$ is a weight for the combination, and $\gamma$ is set as 0.4 to emphasize the STD saliency map, and $S_{final}$ is the final saliency map. The final saliency maps integrate the first stage WBU and second stage STD saliency models together, thus they are able to highlight polyps well.

## V. Results

### A. Experimental Setup

*1) Data Set Description and Implementation Details:* To evaluate the proposed saliency method for polyp detection, we utilized open database CVC-clinicDB [5]. It has been generated from 23 different colonoscopy video studies, and comprises 612 polyp images with a size of 576 ×768. These polyp images were selected from different polyp regions to reduce image similarity. Moreover, the ground truth was created by experts with manually defining the polyp mask.

Experiments are conducted using MATLAB on a desktop computer with an Intel i7-3770 CPU (3.4 GHz) and 32GB RAM. In the feature extraction step by SAE, there are four free parameters: the weight decay parameter ($\alpha$), the sparsity penalty parameter ($\beta$), the sparsity parameter ($\rho$) and the learning rate ($\mu$). We experimentally chose $\alpha = 0.002,\ \beta = 5,\ \rho = 0.5,\ \mu = 0.02$ with the practical tricks introduced in [37]. In our SAE, we just used one hidden layer to learn features and we defined feature dimension as 10 in our experiment. A negligible (1%) difference in total precision of saliency detection was obtained by using different values (8-15) of feature dimension; thus, high robustness of our saliency model is shown.

*2) Evaluation Criteria:* To quantitatively evaluate different saliency methods, we calculated common criteria: Precision and Recall (PR) curves, to evaluation of saliency models. The PR curves are generated by binarizing the calculated saliency maps via varying the threshold from 0 to 255. We computed the precision/recall pairs of all images to plot the precision-recall curves.

We calculated mean precision, recall and F-measure by an adaptive thresholding to evaluate the saliency performance. The threshold is chosen as twice of the mean saliency value within an image. The F-measure, which is a harmonic mean of precision and recall, is also calculated as follows to evaluate the saliency performance,

$$F_{measure} = \frac{(1+\gamma^2) \times Precision \times Recall}{\gamma^2 \times Precision + Recall}. \tag{11}$$

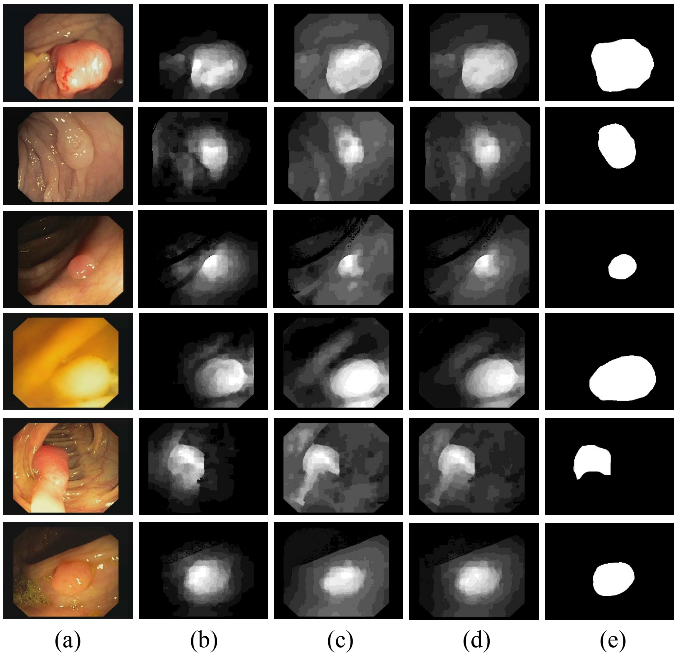


Fig. 6. Illustration of our proposed saliency results. (a) Polyp images, (b) First stage WBU saliency maps, (c) Second stage STD saliency maps, (d) The final saliency maps, (e) Ground truth.

where $\gamma^2$ is set to 0.3 as suggested in [26].

In addition, we also computed the Area Under Curve (AUC), linear Correlation coefficient (CC) and Normalized Scanpath Saliency (NSS) to evaluate saliency maps.

### B. Saliency Results

The first experiment evaluates the performance of the proposed saliency method for polyp images. As shown in Fig. 6, we illustrated our methods with six example polyp images. Fig. 6(b) shows the first stage WBU saliency maps while Fig. 6(c) shows second stage STD saliency maps through learning a robust saliency estimator from a set of training examples. We found that these two saliency maps both provide good performance to detect polyp regions. Moreover, we found that the STD saliency maps could further refine the polyp regions as shown in third and fifth example images. With the fusion strategy, the final saliency maps are shown in Fig. 6(d), where the polyp mucosa preserve high saliency values than normal mucosa region. The WBU and STD saliency maps emphasize polyp region in different aspects and complement each other to generate the final saliency result. The good performance in Fig. 6 demonstrates the effectiveness of the saliency fusion methods.

### C. Component Analysis of the Proposed Saliency Method

We then constructed four baseline experiments on polyp images to demonstrate benefits of the proposed components (first stage WBU saliency map, second stage STD saliency map, automatic feature learning step and object-center based saliency step) in our saliency models. We first analyzed the effectiveness of two complementary saliency steps, i.e. the

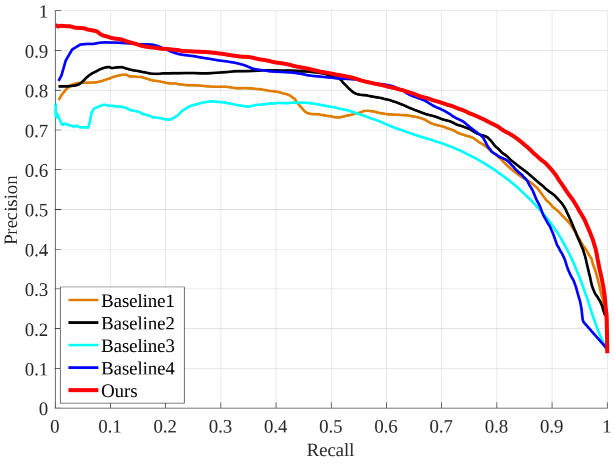


Fig. 7. PR curves of our proposed method and four baseline experiments.

first stage saliency and second stage saliency. We separately calculated the first stage WBU saliency (Baseline 1) and the second stage STD saliency (Baseline 2) on our datasets, and the corresponding PR curves are showed in Fig. 7. The quantitative results show that the integrated saliency maps perform consistently better than either of two results respectively, which also verified by visual examples in Fig. 6. This result validates that the integration of the first and second saliency maps is very simple but effectively incorporates the original WBU saliency maps and the STD saliency maps, which makes the final results accurate.

In our proposed saliency method, SAE is designed to learn feature representation for superpixels. We then constructed the third baseline method and evaluated the contribution of the SAE feature learning step in the polyp saliency calculation by comparing it with those which only using low-level hand-crafted features. The corresponding PR curve is shown in Fig. 7 with Light Blue color (Baseline 3). This result shows that the performance of our results dramatically surpasses the result of saliency maps without SAE feature learning step. This comparison well demonstrates the effectiveness of automatic feature learning step.

Then we analyzed the contribution of proposed object-center based saliency map. In the fourth baseline method, we excluded the object-center based step while keeping the contrast based saliency map unchanged to calculate polyp saliency map. The PR curve for the fourth baseline method is lower than our proposed method. This indicates more accurate saliency maps for polyps are calculated by our proposed method. One possible reason for this result is that the boundary exclusion problem is alleviated since our method highlights the object coarsely.

These results demonstrate that the proposed components (first stage WBU saliency map, second stage STD saliency map, automatic feature learning step and object-center based saliency step) contribute to the final results and they complement each other perfectly in detecting the polyp saliency precisely.

TABLE I
COMPARISON RESULTS OF AUC, CC AND NSS VALUES BY DIFFERENT SALIENCY METHODS.

| Method | AUC | CC | NSS |
|---|---|---|---|
| FT [15] | 0.7536 | 0.3018 | 1.0051 |
| MSSS [16] | 0.7646 | 0.3127 | 1.1232 |
| SDSP [17] | 0.8269 | 0.3212 | 1.1874 |
| HDCT [21] | 0.8790 | 0.3385 | 1.2815 |
| LPS [22] | 0.8237 | 0.3153 | 1.2275 |
| **Ours** | **0.8986** | **0.3547** | **1.3798** |

### D. Performance Comparison with Existing Saliency Models

We further compared our saliency method with five state-of-the-art saliency models: FT [15], MSSS [16], SDSP[17], HDCT [21] and LPS [22].

*1) Qualitative Results:* The visual comparison results of our methods and five state-of-the-art saliency models are shown in Fig. 8. Our approach (Fig. 8 (g)) deals well with challenging cases when the texture of background is similar to that of the polyp. As shown in the second and sixth rows, our method almost successfully highlights the whole polyp region while other approaches are usually distracted by the similar textures on background and the corresponding results are not satisfactory. Moreover, we found that our approach performs well when polyps locate at the image border as in the first and fourth example images in Fig. 8.

*2) Quantitative Results:* To evaluate the saliency detection performance of different saliency models quantitatively, we calculated corresponding PR curves. Fig. 9(a) shows PR curves of different saliency models, which effectively demonstrates that our proposed saliency model can achieve better performance compared with the other models. Moreover, the mean precision, recall and F-measure of each segmentation based on a saliency map are averaged over all polyp images, and the results are shown in Fig. 9(b). Our approach reduced 20.94%, 28.67%, 2.05%, 4.41%, and 16.52% overall error rates on precision, compared with FT [15], MSSS [16], SDSP[17], HD-CT [21] and LPS [22], respectively when evaluated using the polyp dataset. Among these methods, our algorithm achieves the best performance with highest recall (0.8180) and best F-measure values (0.7431) and relative high precision (0.7232). Table I provides the comparison results of detailed AUC, CC and NSS values for different saliency methods. These comparisons demonstrate that our proposed saliency method outperforms state-of-the-art algorithms significantly. The main reason for the superior performance lies in our two stage saliency architecture, where WBU saliency map highlights the polyp region in detail while STD saliency map depicts accurate global contours. Furthermore, compared with the traditional saliency models that utilize hand-crafted features, our method introduces the SAE model to enable powerful discriminative features.

### E. Performance Comparison with Existing Polyp Detection Methods

We finally applied an automatic Otsu's threshold on the obtained saliency map to get the final segmentation results.

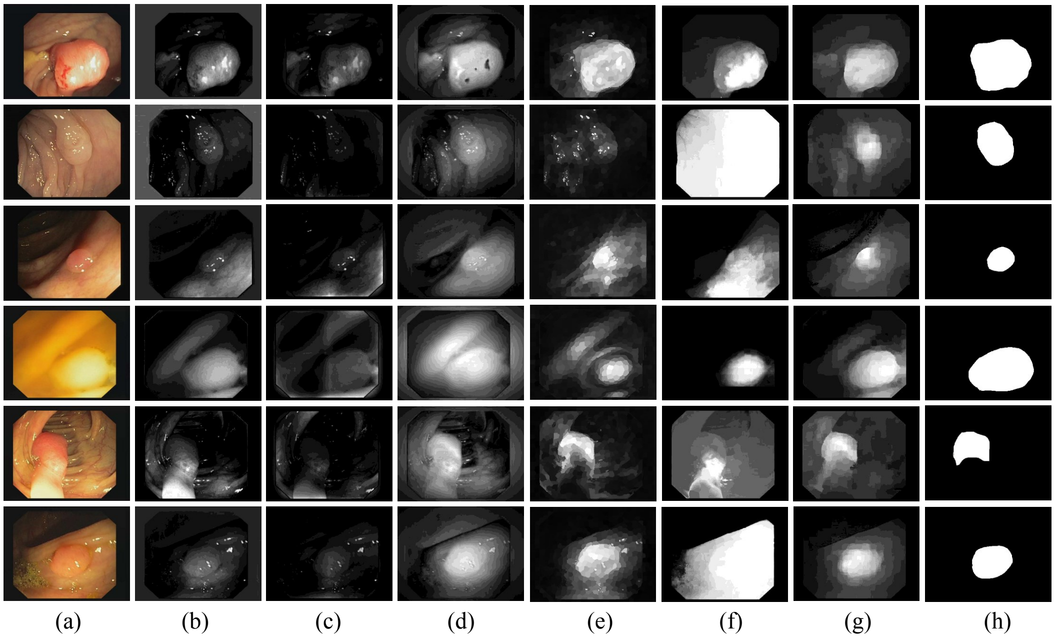


Fig. 8. Comparison of saliency estimation results between the state-of-the-art methods and the proposed one by using six polyp images as examples. (a) Original images, (b) FT, (c) MSSS, (d) SDSP, (e) HDCT, (f) LPS, (g) The proposed method, (h) The binary-labeled ground truth.

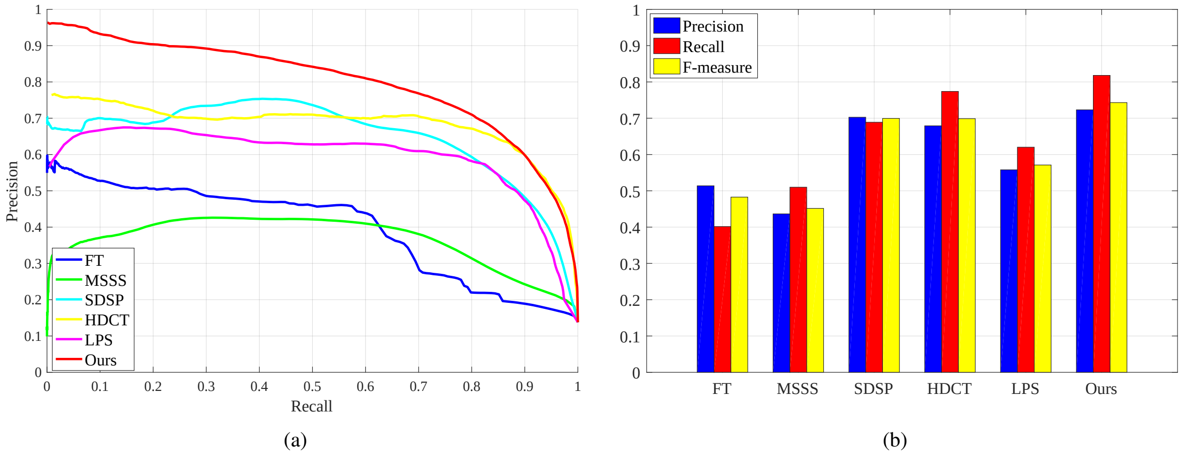


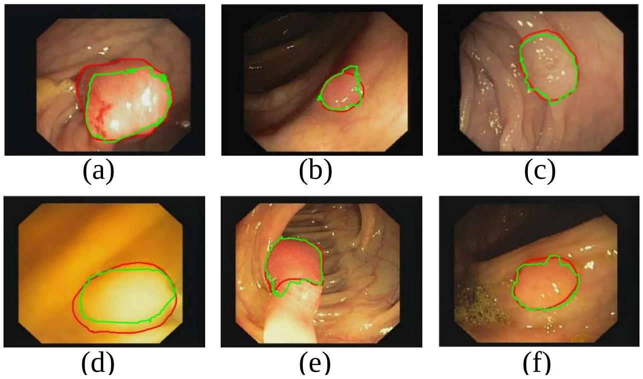
Fig. 9. Experimental results on polyp images for the proposed method and other state-of-the-art algorithms. (a) PR curves. (b) PR bar with F-measure.

The segmentation results are shown in Fig 10. We found that our proposed method achieves good segmentation result for the polyps.

Although lots of methods have already been proposed to deal with polyp detection problem, most of them only focus on the localization with given curves (circles) or calculation of the number of polyps within a certain area, which makes comparison challenging. To analyze our results accurately, we compared our proposed method with the state-of-the-art polyp segmentation methods [4], [9], which provide pixel-based measurements of the segmentation results. The dice similarity coefficient (DSC) was used to quantitatively evaluate the similarity of the segmentation results to their corresponding ground truth. It was defined as: $DSC(A,B)=\frac{2N(A\cap B)}{N(A)+N(B)}$, where $A$ denotes the segmentation result and $B$ is the ground truth, $N(.)$ represents the number of pixels in the corresponding set. The direct comparison validates the effectiveness of our method for polyp detection since we improved DSC from 65.73% and 71.45% in [4] and [9] to 76.26% for the polyp dataset.

## VI. Conclusion

In this paper, we proposed a novel unified bottom-up and top-down saliency model to detect polyps. The first stage WBU saliency model fuses the contrast based saliency and object-center based saliency together. The contrast-based saliency map highlights image parts that show different appearances compared with surrounding areas while the object-center based saliency map is estimated by key points to emphasize the center of the salient object. In the second stage, we propose to calculate the STD saliency map by learning a solid classifier with MKB through samples directly from multi-level WBU saliency maps. Our method achieves promising results with 0.7232 precision, 0.8180 recall and 0.7431 F-measure,



(a) (b) (c)

(d) (e) (f)

Fig. 10. Segmentation results of our proposed saliency methods on six polyp images. Red lines represent the manual annotation by clinicians while green lines show our proposed segmentation results.

demonstrating the effectiveness of our method. Extensive comparison results on polyp datasets show better performance of the proposed method compared with state-of-the-art saliency models and polyp segmentation methods.

## References

[1] A. C. Society, "Key statistics for colorectal cancer," http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-key-statistics/, accessed Jan 11, 2017.

[2] Y. Iwahori, T. Shinohara, A. Hattori, R. J. Woodham, S. Fukui, M. K. Bhuyan, and K. Kasugai, "Automatic polyp detection in endoscope images using a hessian filter." in *MVA*, 2013, pp. 21–24.

[3] J. Bernal, J. Sánchez, and F. Vilarino, "Integration of valley orientation distribution for polyp region identification in colonoscopy," in *Abdominal Imaging. Computational and Clinical Applications*. Springer, 2011, pp. 76–83.

[4] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2144–2151, 2012.

[5] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imaging Graph.*, vol. 43, pp. 99–111, 2015.

[6] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 630–644, 2016.

[7] M. Breier, S. Gross, and A. Behrens, "Chan-vese-segmentation of polyps in colonoscopic image data," in *Proc. Int. Conf. Electrical Engineering*, vol. 2011, 2011.

[8] H. Xu, H. D. Gage, P. Santago, and Y. Ge, "Colorectal polyp segmentation based on geodesic active contours with a shape-prior model," in *Virtual Colonoscopy and Abdominal Imaging. Computational Challenges and Clinical Opportunities*. Springer, 2010, pp. 134–140.

[9] N. Du, X. Wang, J. Guo, and M. Xu, "Attraction propagation: A user-friendly interactive approach for polyp segmentation in colonoscopy images," *PloS one*, vol. 11, no. 5, p. e0155371, 2016.

[10] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin 2009*. Springer, 2009, pp. 346–350.

[11] M. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in hd colonoscopy," in *International MICCAI Workshop on Medical Computer Vision*. Springer, 2013, pp. 205–213.

[12] S.-H. Bae and K.-J. Yoon, "Polyp detection via imbalanced learning and discriminative feature learning," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2379–2393, 2015.

[13] Q. Angermann, A. Histace, and O. Romain, "Active learning for real time detection of polyps in videocolonoscopy," *Procedia Computer Science*, vol. 90, pp. 182–187, 2016.

[14] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1597–1604.

[16] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," in *Proc. IEEE Conf. Image Process.*, 2010, pp. 2653–2656.

[17] L. Zhang, Z. Gu, and H. Li, "Sdsp: A novel saliency detection method by combining simple priors." in *Proc. IEEE Conf. Image Process.*, 2013, pp. 171–175.

[18] N. Tong, H. Lu, L. Zhang, and X. Ruan, "Saliency detection with multi-scale superpixels," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1035–1039, 2014.

[19] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2083–2090.

[20] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "Sun: Top-down saliency using natural statistics," *Vis? cogn?*, vol. 17, no. 6-7, pp. 979–1003, 2009.

[21] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 9–23, 2016.

[22] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: salient object detection in the wild," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, 2015.

[23] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *J. Vis.*, vol. 9, no. 12, pp. 15–15, 2009.

[24] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, 2012.

[25] R.-J. Lin and W.-S. Lin, "A computational visual saliency model based on statistics and machine learning," *J. Vis.*, vol. 14, no. 9, pp. 1–1, 2014.

[26] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, 2015.

[27] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[28] F. Yang, H. Lu, and M.-H. Yang, "Robust visual tracking via multiple kernel boosting with affinity constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 2, pp. 242–254, 2014.

[29] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[30] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 886–893.

[31] Y. Yuan and M. Q.-H. Meng, "A novel feature for polyp detection in wireless capsule endoscopy images," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2014, pp. 5010–5015.

[32] Y. Yuan and M. Meng, "Deep learning for polyp recognition in wireless capsule endoscopy images," *Medical Physics*, vol. 44, no. 4, pp. 1379–1389, 2017.

[33] S. M. Alsaleh, A. I. Aviles, P. Sobrevilla, A. Casals, and J. K. Hahn, "Automatic and robust single-camera specular highlight removal in cardiac images," in *Proc. 37th IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2015, pp. 675–678.

[34] Y. Yuan and M. Q.-H. Meng, "Polyp classification based on bag of features and saliency in wireless capsule endoscopy," in *Proc. IEEE Int. Conf. Robotics Automation*, 2014, pp. 3930–3935.

[35] Y. Yuan, J. Wang, B. Li, and M. Meng, "Saliency based ulcer detection for wireless capsule endoscopy diagnosis," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, 2015.

[36] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, pp. 79–86, 1951.

[37] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, 2015.