EXPLORING THE CLINICAL POTENTIAL OF AN AUTOMATIC COLONIC POLYP DETECTION METHOD BASED ON ENERGY MAPS CREATION

Fernández-Esparrach G[1], Bernal J[2], López-Cerón M[1], Córdova H[1], Sánchez-Montes C[1], Rodríguez de Miguel C[1], Sánchez FJ[2].

[1] Endoscopy Unit, Gastroenterology Department, Hospital Clínic, IDIBAPS, CIBEREHD, University of Barcelona, Barcelona, Spain

[2] Computer Science Department, Universitat Autònoma de Barcelona and Computer Vision Center, Barcelona, Spain

Author for correspondence: G. Fernández-Esparrach
Hospital Clínic
Villarroel 170
08036 Barcelona
Phone:+34932275400
mgfernan@clinic.ub.es

ABSTRACT

Background and aims: Polyp miss-rate is a drawback of colonoscopy that increases significantly in small polyps. We explored the efficacy of an automatic computer vision method for polyp detection.

Methods: Our method relies on a model that defines polyp boundaries as valleys of image intensity. Valley information is integrated into energy maps which represent the likelihood of polyp presence.

Results: In 24 videos containing polyps from routine colonoscopies, all polyps were detected in at least one frame. Mean values of the maximum of energy map were higher in frames with polyps than without ($p<0.001$). Performance improved in high quality frames (AUC= 0.79, 95%CI: 0.70-0.87 vs 0.75, 95%CI: 0.66-0.83). Using 3.75 as maximum threshold value, sensitivity and specificity for detection of polyps were 70.4% (95%CI: 60.3-80.8) and 72.4% (95%CI: 61.6-84.6), respectively.

Conclusion: Energy maps showed a good performance for colonic polyp detection. This indicates a potential applicability in clinical practice.

INTRODUCTION

Colonoscopy is considered the gold standard test for colorectal cancer (CRC) screening as it is able to detect and resect polyps [1, 2]. However, a significant number of polyps are missed during colonoscopy with a reported miss-rate of up to 22% [3]. Detection rates depend on the proportion of the mucosal surface inspected and correlate with the cleanliness and time dedicated to mucosal inspection during withdrawal [4]. Depending on the number, size and histologic features of polyps, a surveillance colonoscopy is recommended at different intervals [2]. For these reasons, several efforts have been made to improve polyp detection, including better tolerated cleaning solutions, mucosal enhancement techniques and improvements in endoscopes and devices in order to increase mucosal visualization [5,6].

Intelligent systems for colonoscopy couple clinicians and computer scientists to build systems that aid gastroenterologists in the different stages on the intervention. Existing works are mainly focused on polyp detection, commonly guided by shape [7], texture and color [8] cues. The recently described WM-DOVA (Window Median Depth of Valleys Accumulation) energy mapssystem [9] accurately highlights the region of an image containing a polyp. These maps are based on a model of appearance for polyps defining polyps as protrusions in the mucosa defining their boundaries in terms of intensity valleys. Polyp localization in the image is a required step for automatic characterization and further optical histology presumption.

The objective of this study was to assess the potential of WM-DOVA maps for polyp detection in two settings, a subset of frames all containing a polyp and another with full sequences with frames with and without polyps, and to evaluate if polyp morphology influences the model outcome.

MATERIAL AND METHODS

24 videos containing 31 different polyps with as many different appearances as possible were recorded from routine colonoscopies performed with a standard-resolution, white light videocolonoscope (Olympus Q160AL and Q165L, Olympus Europe, Hamburg, Germany). The study was approved by the Hospital Clinic of Barcelona Ethics Committee.

For the first experiment with polyp frames (experiment A), we used the CVC-ClinicDB database [9] which contains 612 polyp images (size 576×768 pixels) from all 24 videos (table 1). Polyps were classified as protruded (0-Is, 0-Ip) or flat (0-II) following the Paris classification [10]. An average number of 20 frames (range 2-25) were extracted from each video, rejecting frames with excessive stool due to extremely poor patient preparation. For validation purposes, a group of expert endoscopists (previously trained for reaching an agreement) manually defined a mask (ground truth) on the region covered by the polyp (figure 1). Bowel preparation at the segment containing the polyp was evaluated according to Boston classification and scored from 0 to 3 [11].

In the second experiment with full videos (experiment B), the same experts categorized all 47,886 frames from the 24 videos into two groups: those with and without a polyp. Frames with a polyp were further sub-classified according to image quality, labelling as low quality those frames with poor patient preparation, blurring, presence of endoscopic tools or excessive camera movement.

Creation of energy maps

Energy maps creation is based on a model of polyp appearance which considers polyps as protrusions in the mucosa. When protruding surfaces are illuminated

perpendicularly, boundaries appear as shadows in the image that can be associated to valleys in the intensity image (figure 2). Following this model, which fits better for zenithal views of the polyp, a pixel inside a polyp should be surrounded by high-intensity valleys in the majority of directions. Considering this, we use a series of radial sectors centered on each pixel to search for these valley contributions in each of the possible directions (figure 3). The final value for a given pixel integrates contributions from all directions and also imposes conditions (continuity, concavity, completeness, robustness) [9] to the final output in order to mitigate the impact of other elements of the scene with valley information (folds, blood vessels).

Our method takes as input a colonoscopy frame, and produces an energy map in which brighter regions correspond to pixels with high likelihood of polyp presence (figure 4). In experiment A, detection was considered as correct when the position of the brightest area fell within the polyp mask. In experiment B, we calculated the maximum energy level for each frame which corresponds to the maximum likelihood value for all the pixels in the particular frame and we explored if the value of this maximum was related to presence of a polyp. We defined MeanMaxDOVA for each video as the mean of the maximum of WM-DOVA maps for all the frames in three different scenarios: frames without polyp, frames with polyps and frames with only high quality (HQ) polyp frames.

Statistical analysis

Chi Square test was used to compare the proportion of frames with correct polyp detection between small type 0-II polyps and all the other types and to evaluate the association between bowel cleanliness and the correct location of polyps. MeanMaxDOVA in frames with and without a polyp were compared with the Mann-Whitney test for independent data. These calculations were done with the SPSS 23

statistical software (SPSS Inc, Chicago III).

Receiver operator characteristic (ROC) curves for the use of maximum energy values to determine polyp presence were constructed with Matlab. Area Under Curve (AUC) and optimal operating point (OP) with its sensitivity and specificity and 95% confidence interval (CI) were calculated.

All tests were 2-sided and p values less than 0.05 were considered statistically significant.

RESULTS

In both experiments, all different polyps were correctly detected in at least one frame regardless of the size and appearance (table 1).

Experiment A: where is the polyp in the image?

The proportion of correctly identified polyps was higher for small type 0-II polyps compared with all the other types: 169/218 (77.5%; 95%CI: 71.5-82.6) vs 261/394 (66.2%; 95%CI: 61.4-70.7); p<0.01. The methods' performance did not depend on patient preparation (table 2). In the 182 frames without a correct detection, the causes of failure were: strong deviation from polyp appearance model (mainly non-zenithal views) in 94 (51.6%), low image quality in 71 (39%), solid fecal particles in 6 (3.3%), poor preparation in 5 (2.7%), and others in 6 (3.3%) (Figure 5).

Experiment B: is there a polyp in the image?

Figure 6 shows that MeanMaxDOVA was higher in frames containing a polyp than in frames without a polyp (p<0.001) and this difference increased when only HQ polyp frames were considered (p<0.001). ROC curves (figure 7) show that system performance again depends on image quality leading to higher AUC scores in HQ

6

frames compared with all frames (0.79, 95%CI: 0.70-0.87 vs 0.75, 95%CI: 0.66-0.83). Optimal OP was achieved by using 3.75 as threshold value for energy map maximum with a sensitivity of 70.4% (95%CI: 60.3-80.8) and specificity of 72.4% (95%CI: 61.6-84.6). In other words, we missed around 30% of the frames with a polyp but provided only a false alarm for each 3 correct detections.

DISCUSSION

WM-DOVA maps are able to detect polyps in at least one frame in all sequences by using the maximum value of the energy map and regardless of their size and appearance. Almost all types of polyps were represented in the database with frequencies similar to their prevalence in the general population, supporting the robustness of our method. Our method appears especially useful for small and flat (type 0-II) lesions which are the most difficult to detect in real conditions. Although WM-DOVA maps were initially developed to detect the position of polyps in frames containing polyps, they are also able to discern between polyp presence or absence in the images when full videos are analyzed by using the value of the maximum of the map as threshold.

WM-DOVA enery maps appear especially useful for small type 0-II lesions which are the most difficult to detect. This is due to the fact that these lesions tend to be observed zenithally and the complete polyp boundary can be identified, meaning that we have complete valley information representing the polyp. In this case we took advantage of the different constrains applied in our method to discriminate polyp

boundaries (completeness, continuity and local circularity) which, in the case of zenithal polyp views, are completely fulfilled.

Our method seems to be robust in non-ideal conditions as poor preparation does nott negatively affect the results, whereas colonoscopy studies show a clear dependence on patient preparation[12]. One possible explanation is that, as the current version of the method does not use color and texture cues, solid fecal particles mimick the apparence of polyps fitting the model conditions (see Figure 5 c). However, other type of fecal content not fitting the appearance model, does not represent a problem.

Although this study shows good performance of WM-DOVA maps, there are some cases in which lateral observations of a polyp lead to detection errors due to the presence of other elements in the scene with valley information (blood vessels, folds). To improve our results, our model should be extended to both cope better with lateral views and to better discriminate polyps from these other elements of the scene.

Finally, our results also show the positive impact of high image quality on the methods' performance. It is worth to mention that our method was tested over standard-resolution images and one could expect better results with high resolution endoscopes where a better definition of polyp boundaries would definitely help WM-DOVA maps performance. This fact reinforces the need of an ongoing collaboration between clinicians and computer scientists to create computational systems that provide high quality data.

Computer- aided detection (CAD) utility has already been demonstrated in CT-colonography, with a reported sensitivity for polyps <10 mm close to 80% [13]. CAD has also shown to increase the radiologists'performance, especially in non-expert and

moderately experienced practitioners [14]. Although WM-DOVA performance in its current version is slightly lower than these results, it is impacted by the presence of other elements in the endoscopic luminal scene (vessels, specularities…) that are not present in the radiological image and, once mitigated, we could expect an improvement in the results. Similarly to CAD in colonography, the use of WM-DOVA in a concurrent-reader paradigm could be more time-efficient than colonoscopy alone and could increase its sensitivity detecting difficult polyps without significantly reducing specificity [15]. These potential benefits could even be more evident among colonoscopists with a moderate level of experience which represent a substantial fraction of endoscopists performing regular colonoscopies.

We foresee polyp characterization –automatic optical biopsy- as the most clinically useful computational tool to be developed. Although WM-DOVA maps do not allow to distinguish between adenomatous and non-adenomatous polyps, they can be used for detecting the polyp,which is the first stage in the whole characterization pipeline. Once lesion is detected, following stages would focus on the computational analysis of the content of the detected polyp region aiming to obtain a first indication of polyp histology.

In conclusion, our results definitely indicate a potential of WM-DOVA maps as an accurate polyp detection tool, warranting further clinical studies.

REFERENCES

1. Levin B, Lieberman DA, McFarland B et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the

American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 2008;134:1570-95.

2. Hassan C, Quintero E, Dumonceau JM, et al. Post-polypectomy colonoscopy surveillance: ESGE Guideline. Endoscopy 2013;45:842-51.

3. van Rijn JC, Reitsma JB, Stoker J et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006;101:343-50.

4. Lee TJ, Rees CJ, Blanks RG et al. Colonoscopic factors associated with adenoma detection in a national colorectal cancer screening program. Endoscopy 2014;46:203-11.

5. Belsey J, Crosta C, Epstein O, et al. Meta-analysis: the relative efficacy of oral bowel preparations for colonoscopy 1985-2010. Aliment Pharmacol Ther 2012;35:222-37.

6. ASGE Technology Committee, Konda VI, Chauhan SS, Dayyeh BKA et al. Endoscopes and devices to improve colon polyp detection. Gastrointest Endosc 2015;81:1122-9.

7. Tajbakhsh N, Gurudu SR, Liang J. Automatic polyp detection using global geometric constraints and local intensity variation patterns. Med Image Comput Assist Interv 2014;17:179-87.

8. Bernal J, Sánchez FJ, Vilariño F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 2011; 45:3166-82.

9. Bernal J, Sánchez FJ, Fernández-Esparrach G et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput Med Imaging Graph 2015;43:99–111. 10. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract . Endoscopy 2005; 37: 570 – 8.

11. Calderwood AH, Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. Gastrointest Endosc 2010;72:686-92.

12. Lebwohl B, Kastrinos F, Glick M et al. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. Gastrointest Endosc 2011;73:1207-14.

13. Regge D, Della Monica P, Galatola G et al. Efficacy of computer-aided detection as a second reader for 6-9 mm lesions at CT colonography. Radiology 2013;266:168-76.

14. Mang T, Bogoni L, Anand VX et al. CT colonography: effect of computer-aided detection of colonic polyps as a second and concurrent reader for general radiologists with moderate experience in CT colonography. Eur Radiol 2014;24:1466-76.

15. Halligan S, Mallett S, Altman DG et al. Incremental benefit of computer-aided detection when used as a second and concurrent reader of CT colonographic data: multiobserver study. Radiology 2011;258:469–476.

ACKNOWLEDGEMENTS

TABLES

Table 1. Distribution of polyps in the CVC-Clinic database according to size and Paris classification and number of frames with correct detection of polyps.

Polyp size	Paris classification	# of polyps	# of frames	# of frames with correct localization	# of polyps with at least one frame with correct localization
< 10 mm (small)	All	22	430	308 (71.6%)	22 (100%)
	0-II	11	218	169 (77.5%)	11 (100%)
	0-Is	9	162	105 (65%)	9 (100%)
	0-Ip	2	50	34 (68%)	2 (100%)
≥ 10 mm (normal)	All	9	182	122 (67%)	9 (100%)
	0-II	1	12	6 (50%)	1 (100%)
	0-Is	2	46	31 (67.4%)	2 (100%)
	0-Ip	6	124	85 (68.5%)	6 (100%)

Table 2. Impact of patient preparation in localization results. Boston 2 and 3 have been grouped because they represent good preparation (p=NS).

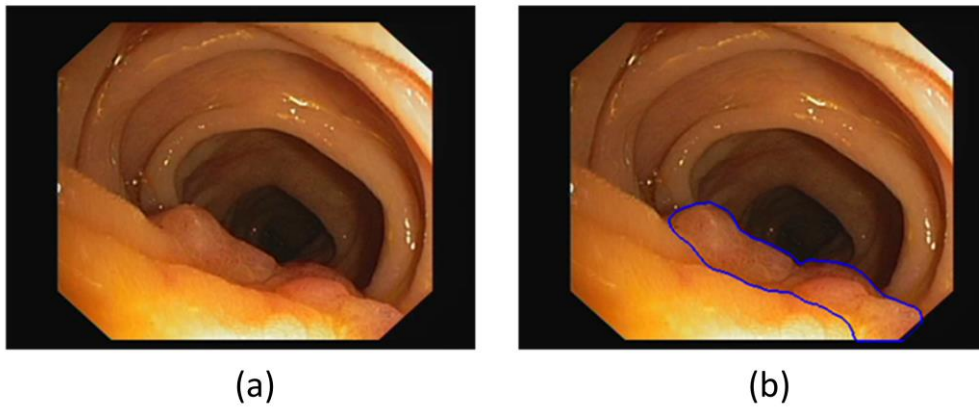| Boston score	# of frames with correct localization	# of frames with failed localization	# Total frames
1	39 (78%)	11	50
2+3	391 (70%)	171	562

LEGEND OF FIGURES

Figure 1. Creation of a polyp mask (ground truth) to validate WM-DOVA maps: (a) frame containing a polyp 0-Is; (b) the same polyp with a blue line delineating the boundary.
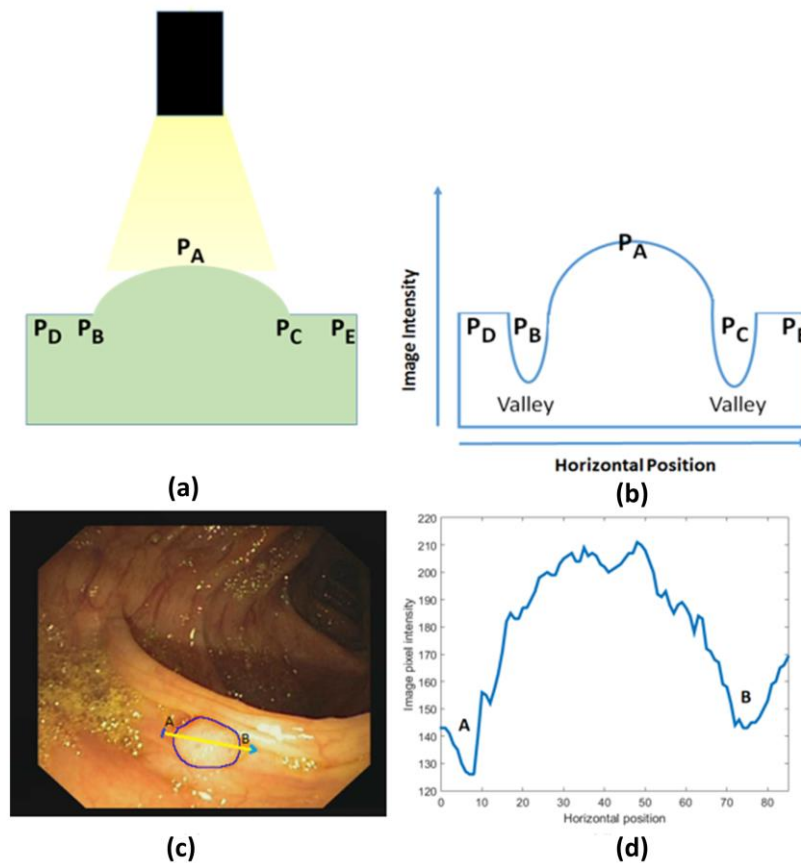
Figure 2. Model of appearance of polyps: (a) synthetic image in which a polyp is illuminated frontally by the light of the endoscope; (b) synthetic image intensity plot; (c) real colonoscopy image and (d) colonoscopy image intensity plot. To ease the understanding of the images and make a better correspondence between images and corresponding intensity plots, key points have been marked in both synthetic and real images.

Figure 3. Polyp boundaries as intensity valleys: (a) original image; (b) preprocessed image for mitigation of other structures that generate valleys; (c) output of valley detection and (d) output of valley detection with series of radial sectors used to calculate WM-DOVA maps superimposed (yellow lines mark the limits of each sector, blue circles represent maximum of valley image under sector).

Figure 4. WM-DOVA energy maps generation: (a) original image; (b) preprocessed image for the mitigation of other structures that generate valleys (c) WM-DOVA energy map and (d) output image with a correct polyp detection: the maximum of WM-DOVA energy map (green square) falls within the polyp mask.

Figure 5. Causes of localization errors: (a) deviation from the model –lateral views-; (b) presence of vessels with strong valley information contribution; (c) presence of fecal content and (d) low image quality. Polyp boundary is painted as a blue line and the position of the maximum of WM-DOVA map is marked as a red square.

Figure 6. Distribution of MeanMaxDOVA in frames with and without a polyp and in high-quality (HQ) frames with a polyp. * p<0.001 between frames without polyp vs frames with polyp in all images and ** p<0.001 between frames without polyp vs frames with polyp only in HQ images.

Figure 7. Receiver Operating Curves (ROC) for the maximum of WM-DOVA map. Red line represents ROC curve calculated using only HQ frames whereas blue line is calculated using all the frames.

Figure 1. Creation of a polyp mask (ground truth) to validate WM-DOVA maps: (a) frame containing a polyp 0-Is; (b) the same polyp with a blue line delineating the boundary.

Figure 2. Model of appearance of polyps: (a) synthetic image in which a polyp is illuminated frontally by the light of the endoscope; (b) synthetic image intensity plot; (c) real colonoscopy image and (d) colonoscopy image intensity plot. To ease the understanding of the images and make a better correspondence between images and corresponding intensity plots, key points have been marked in both synthetic and real images.

Figure 3. Polyp boundaries as intensity valleys: (a) original image; (b) preprocessed image for mitigation of other structures that generate valleys; (c) output of valley detection and (d) output of valley detection with series of radial sectors used to calculate WM-DOVA maps superimposed (yellow lines mark the limits of each sector, blue circles represent maximum of valley image under sector).
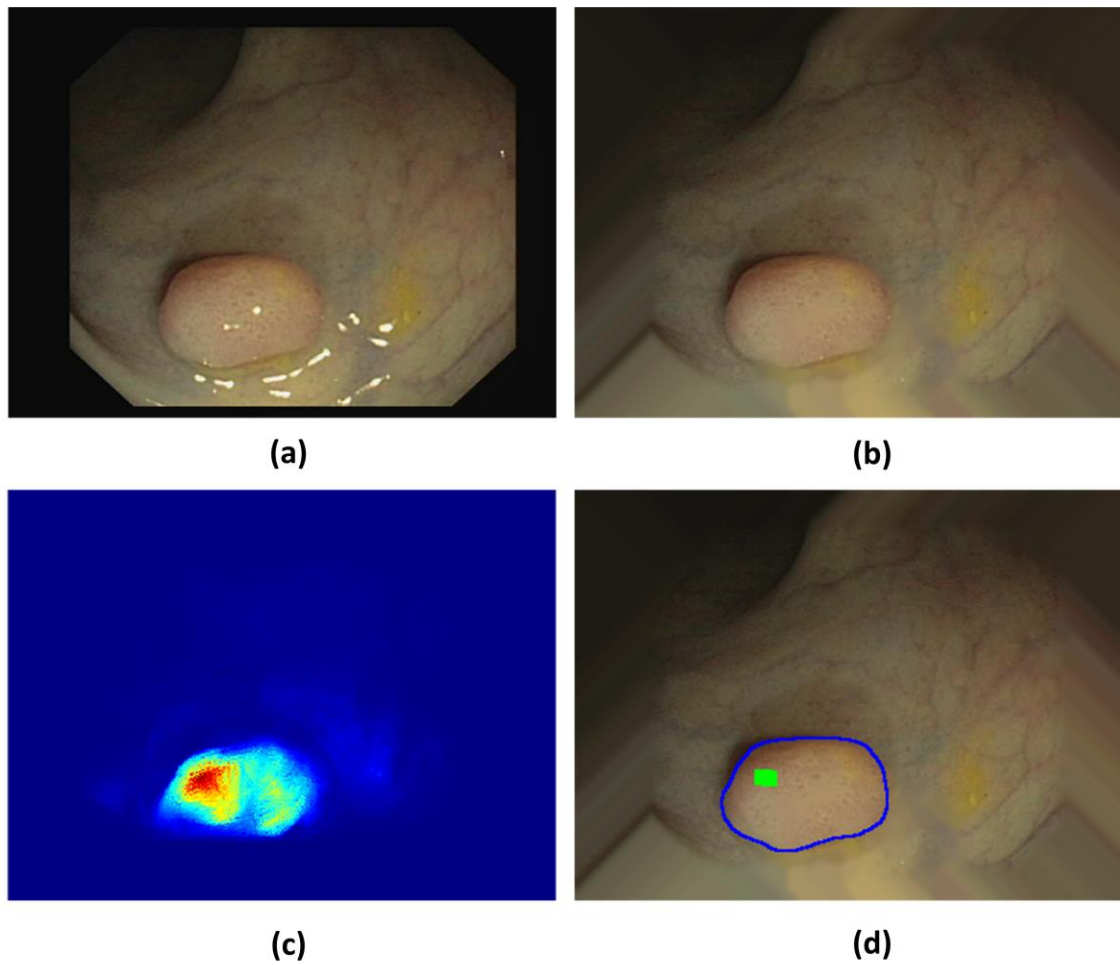
Figure 4. WM-DOVA energy maps generation: (a) original image; (b) preprocessed image for the mitigation of other structures that generate valleys (c) WM-DOVA energy map and (d) output image with a correct polyp detection: the maximum of WM-DOVA energy map (green square) falls within the polyp mask.
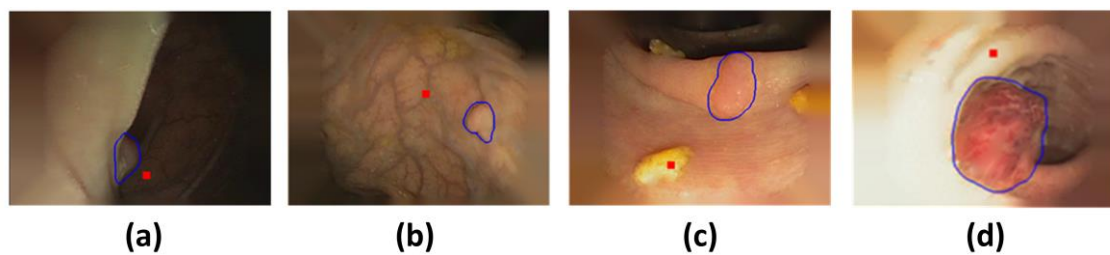

Figure 5. Causes of localization errors: (a) deviation from the model –lateral views-; (b) presence of vessels with strong valley information contribution; (c) presence of fecal content and (d) low image quality. Polyp boundary is painted as a blue line and the position of the maximum of WM-DOVA map is marked as a red square.
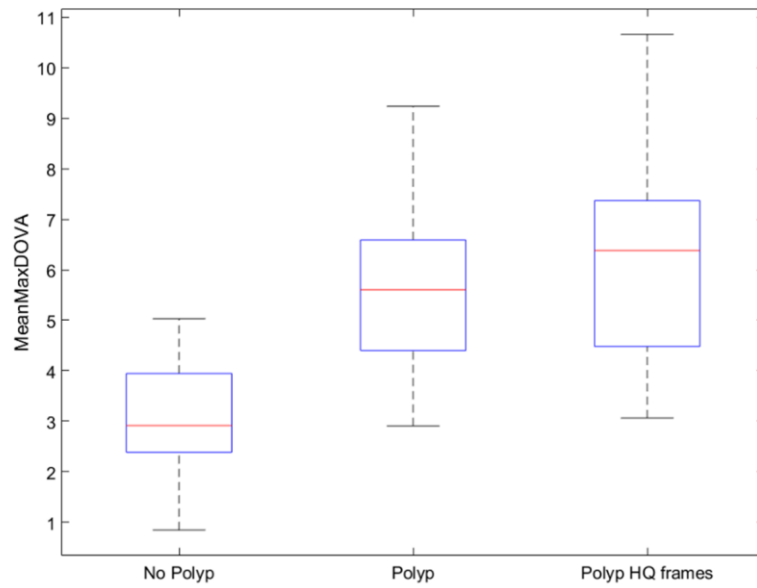
Figure 6. Distribution of MeanMaxDOVA in frames with and without a polyp and in high-quality (HQ) frames with a polyp. * p<0.001 between frames without polyp vs frames with polyp in all images and ** p<0.001 between frames without polyp vs frames with polyp only in HQ images.
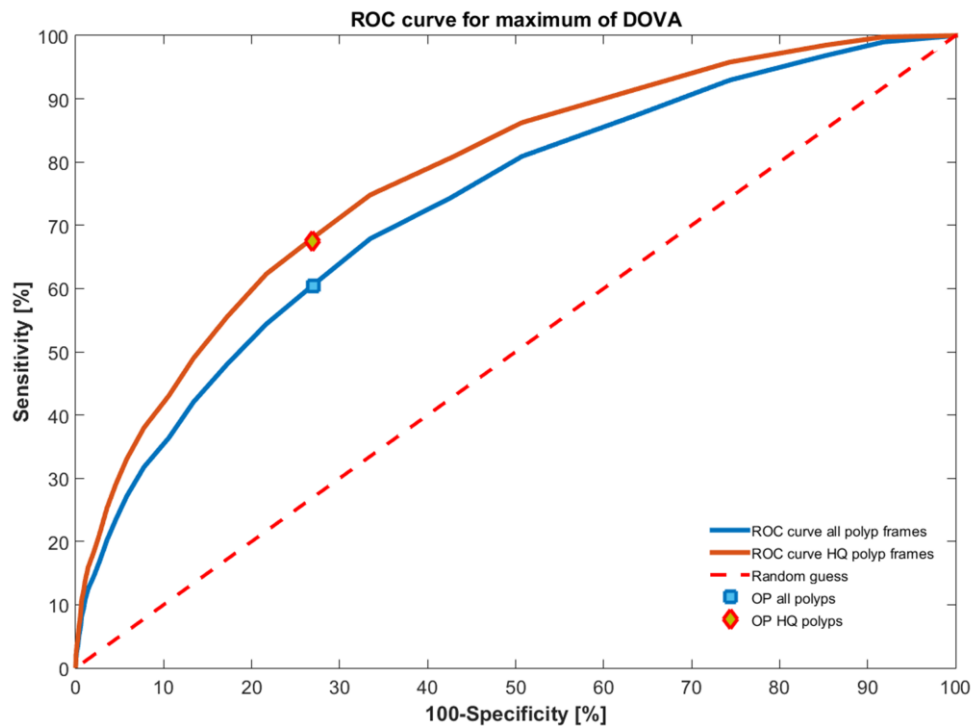

Figure 7. Receiver Operating Curves (ROC) for the maximum of WM-DOVA map. Red line represents ROC curve calculated using only HQ frames whereas blue line is calculated using all the frames.