2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)

Training Data Enhancements for Robust Polyp Segmentation in Colonoscopy Images

Victor de A. Thomaz, Cesar A. Sierra-Franco, and Alberto B. Raposo
Pontifical Catholic University of Rio de Janeiro,
Rio de Janeiro, Brazil.
E-mail: vthomaz@inf.puc-rio.br, {casfranco,abraposo}@tecgraf.puc-rio.br

***Abstract*—Automatic polyp detection systems are an important tools to aid in the diagnosis and prevention of colorectal cancer. Currently, methods based on deep learning approaches have presented promising results. However, the performance of these techniques is highly associated with the use of large and varied data samples for training. This is one of the main limitations of applying Deep Learning techniques in the medical field since the amount of data for training is generally limited compared to nonmedical disciplines. This work proposes a novel method to increase the quantity and variability of training images from a publicly available colonoscopy dataset. The developed approach enrich the training data adding polyps to regions of nonpolypoid samples, creating automatically new data with their appropriate labels. Performance results show that convolutional neural networks trained in these syntactically-enhanced datasets improved the accuracy on polyps segmentation task while reducing the false positive rate. These results open new possibilities for advancing the study and implementation of new methods to automatically increase the number of samples in datasets for computer-assisted medical image analysis.**

***Index Terms*—Colonoscopy, Polyp Detection, Augmentation, Training Data, Segmentation, Convolutional Neural Networks**

I. Introduction

The second leading cause of death from cancer originates from cancer of the colon [19]. The mortality rate in the United States has declined in recent decades. One of the probable reasons is the increase in the accomplishment of colonoscopy examinations in this period [21]. This demonstrates that the colonoscopy technique, which allows visualization of the elements present in the inner part of the colon, is primordial for the prevention of this type of cancer. In general, in these cases, tissue growth of the colon mucosa occurs forming a polyp. An undetected polyp may become lethal over time, which highlights the importance of early-stage polyp detection for the prevention of such disease.

Currently the colonoscopy exam represents the standard procedure for inspecting the colon for polyps and other lesions. However, colonoscopy is an operator-dependent procedure [4]. This means that the number of polyps found may vary depending on factors such as fatigue and professional experience. Some polyps are difficult to detect, even for highly trained physicians. Medical studies estimate that a rate of missed polyps by the physicians is about 25% [5]. Computer-aided polyp detection systems may reduce the rate of polyps missed during colonoscopy examination [4]. Initially, studies that presented methods for polyps detection were based on the appearance of the edges of these lesions and curvature analysis [10] [8]. Several approaches to detecting polyps can be found in the literature [2], [23], [24], [27], however, recent works [14], [17], [26] have evidenced Deep Convolutional Neural Network (CNN) as a technique for lesions detection with higher accuracy. Nevertheless, the efficiency of these techniques is directly associated with the quantity and variety of the colonoscopy images in the training data.

In the medical context, large datasets are difficult to obtain due to the high cost of skilled labor and privacy restrictions [6]. Large data are the exclusive of a few research groups, impacting the scientific production of those who do not have access to this information. An example of a publicly available dataset is the CVC-ClinicDB [2], which consists of 612 images extracted from colonoscopy videos. However, this number of images may be insufficient for training successfully a deep neural network. Another limitation founded in the CVC-ClinicDB dataset is that it presents several images for the same polyp, varying only its point of view. This shows the reduced variation of the colonoscopy images. Ideally, a suitable training data should be composed of a large number of images that present different sizes and polyps morphologies from different patients. These images must additionally contain a set of specific labels to successfully apply machine learning algorithms, such as bounding boxes for object detection or paired images with label pixels for semantic segmentation tasks.

The larger and more varied the dataset, the better the training and performance of systems that use machine learning approaches. Based on this observation, and the difficulty of obtaining datasets of suitable colonoscopy images, the present work proposes a novel method to increase the variation of polyps in colonoscopy images, improving detection accuracy in deep learning based systems. Specifically, we insert polyps from the CVC-ClinicDB database into samples with regions without polyps. Thus, the polyps variability in the images and the sample quantity were increased. To asses the effectiveness of our approach, we used a U-net [16] neural network to perform segmentation and detection tasks. We compare the performance results between training the neural network with the original data, a traditional data-augmentation approach and our synthetically enhanced dataset. The respective labels were produced according to the shape and location of each

2372-9198/19/$31.00 ©2019 IEEE
DOI 10.1109/CBMS.2019.00047

new inserted polyp. Results showed the effectiveness of our approach, improving the accuracy on polyps segmentation task and reducing the false positive detections. The remainder of this paper is organized as follows. Section II presents the literature review. Section III details the method and the steps for extracting, inserting new samples, and producing its labels. Section IV presents the experiments and results, and Section V offers some concluding remarks and future research perspectives.

II. Related Work

In recent years many efforts have been made to develop automatic methods for polyps detection. However, with the advancement of the deep learning techniques for image processing and computer vision, there was a motivation for the adoption of these approaches in Computer-Aided Diagnosis (CAD) systems. Shin et al. [17] presented a polyp detection system based on CNN models. That approach introduced augmentation strategies such as rotation, brightening, shearing, blurring and zooming to increase the number of training samples. Shin et al. made use of transfer learning of a pre-trained Inception Resnet model, using a large image dataset named COCO (Common Objects in Context) [11]. In the case of colonoscopy images, the study of Shin et al. used the CVC-ClinicDB database for training and ETIS-LaribPolypDB [20] for the testing. They also used transfer learning and data augmentations techniques to improve detection results because of the low quantity and variability on the training data. Billah et al. [4] stated that system performance is dependent on the training data. These authors used several publicly available datasets [3] (CVC-ClinicDB, ETIS-LaribPolypDB, and ASU-Mayo [25]) and combined with their own dataset. The author's effort to form a more varied dataset confirms that there is a need for improvement on availability, quantity, and variability of colonoscopy images for machine learning approaches.

Despite advances in polyp detection systems that use deep learning techniques, little was done in relation to the datasets in the sense of contribution in the variation of the samples. It is important to remember that in the context of medical imaging, labeling the location of each lesion must be performed by clinical experts, which is a time-consuming and costly process. The potential benefit of deep learning applications is hampered by the lack of availability of image datasets properly recorded (labels) in quantities and variability to meet the requirements of detection applications for clinical use. Magnetic Resonance Imaging (MRI) initiatives were applied focusing on data augmentation and physician training [7]. In that study, a real brain MRI dataset is used as a source for the generation of new synthetic images, contributing to improve precision in diagnoses tasks.

In the context of colonoscopy images, Shin et al. [18] propose a method to add more colonoscopy images to the original CVC-ClinicDB dataset. This effort goes toward improving training processes by allowing the use of additional data. In that study, the authors used a Faster R-CNN network to test the effectiveness of a polyp detection task. The network trained on the original dataset, obtain precision metric of 59.3% and recall metric of 48% in the testing stage. In a second moment, the same network was trained on a combined dataset (original + generated), improving the performance metrics to 69.4% and 67.4%, for precision and recall respectively. The improvement in detection results demonstrates the relevance of increase and variate the samples in a dataset of colonoscopy images. These augmentation process can contribute to the better training of computer-aided systems based on deep learning techniques.

In this study, we aim to evaluate a strategy that inserts polyps in a dataset to create new samples and to compare the effect of the U-net based polyp detection over traditional data augmentations techniques.

III. Method

In this section, we present the steps of the polyps insertion process in new images. These images will be part of a new dataset composed of the original dataset complemented with the insertion of other polyps, increasing the number of samples and data variability for a deep learning solution.

A. Overview

Our proposed method is to increase the data variability by creating new samples inserting polyps on nonpolypoid images. We create an automatic execution of tasks that result in a richer dataset to be used for training a segmentation CNN. Figure 1 shows the described process. In Figure 1 (a) the polyp highlighted by the green square is copied from a source image belonging to the original dataset. Then is applied to another image (Figure 1 (d)) from the same dataset. This process form different images with richer variability.

The first step of the method is to verify (in the entire dataset) which polyps are appropriate for image extraction. We consider the $k-th$ smallest polyps in the dataset, where the size is counted by the number of pixels. Therefore, this step determines a list of polyps that meet a size criteria. This criterion could be based on the types of lesions according to the Paris Classification [1]. However, there is no information in the CVC-ClinicDB dataset corresponding to the types of polyps presented in the images. Then a bounding box around the polyp region is obtained. Here, bounding box coordinates are used to obtain a copy of the exact region in which the polyp is present on the original image. Then, this region is duplicated to latter be added on a different target image.

The next step determines in which region in the target image the polyp will be added. A set of regions of interest are selected according to the output of the watershed transformation [12] on the target image. The region selected to receive the polyp should be the largest number of pixels that is not positioned on another existing polyp. If any region meets this criteria, then the rest are excluded at this stage. At this point, the region that will receive the polyp is already defined and the copy of the polyp obtained earlier is applied within this region. The output image is then obtained by composing

an image from the original dataset with the new lesion added. We explain in detail these steps on the following subsection.

B. Polyp Insertion Process

Source Image Steps: In colonoscopy images it is common to see several types and sizes of polyps. However, the perception of lesion size in the images may be affected by the camera's view. If the camera is closer to the mucosa, the polyp may appear larger than it actually is, or if its far its possible to miss the lesion detection. In this study, the polyp size is defined according to the size of the area it occupies in pixels. In the CVC-ClinicDB dataset a same polyp appears in more than one image. Thus, the proposed method will consider the size of this polyp in each image separately, even if it is the same lesion. The image of the dataset containing the position of the polyp in the image is called the ground truth image (i.e. the binary polyp mask annotated by physicians). In that picture, the number of blank pixels in determine the size of the polyp.

We perform a check size for each ground truth image, where $k - th$ minor polyps are selected. In order to insert these polyps in regions of another image, it is appropriate use the smaller polyps, since these lesion can be inserted within regions returned by the watershed process. The main idea is to insert the polyp whose size is smaller than the selected watershed region. As the watershed region tends to be the homogeneous area of the image, the polyp insertion is compatible in size with the target region.

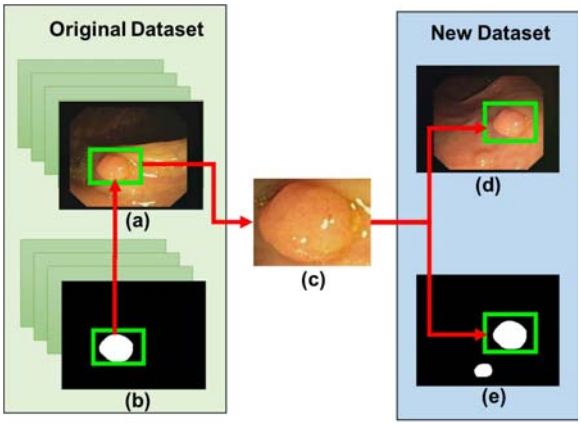

Fig. 1. An illustration of the polyp selection and insertion process. (a) Respective source image polyp bounding box. (b) Polyp selection bounding box in ground truth image. (c) Duplicated polyp sub-image. (d) Polyp applied to the selected watershed region. (e) The ground truth created for the added polyp.

After select the smallest polyps, the smallest bounding box capable of accommodate the polyp area is defined by means of the ground truth image, as seen in Figure 1 (b). With the size and position information of the bounding box defined it is possible to extract the polyp in the corresponding colonoscopy image. Figure 1 (c) show the polyp duplication area in the source image (Figure 1 (a)) that belongs to the original dataset.

Target Image Steps: once the polyp is ready to be added into the target image we must decide on which region of the image it will be positioned. This choice of the best region is important because it helps to maintain consistency, preventing impairing appearance and keeping similar as possible to the real colonoscopy images. We based the region choice on the watershed strategy for target image segmentation. Figure 2 represents the steps of this process, where the target image in Figure 2 (A) is used as the basis for finding a set of regions illustrated in Figure 2 (J). The polyp will be added in one of these regions, if any is compatible with the polyp to be inserted. In case of incompatibility, this target image is discarded and a new one is selected within the original dataset.

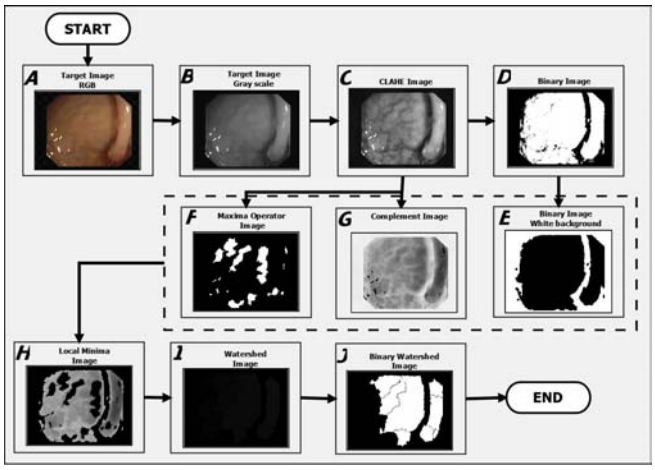

Fig. 2. Flow chart of the generation procedure of target image watershed region.

The generation process is initiated by the creation of a gray scale version of the target image. After this, the equalization process is applied to increase contrast and limit noise in homogeneous areas [15], resulting in the CLAHE image as seen in Figure 2 (C). The CLAHE image is used as the basis for the creation of three versions: a binary image, a maxima operator and complement (Figure 2 (D), (F) e (G) respectively). The binary image is used as the basis for the creation of a modified version of the binary image described in Figure 2 (E) as a binary image white background. This latter one is an image where corresponding black background areas in the Figure 2 (D), are surrounded by blank areas and then filled to form a homogeneous area.

In Figure 2 (E), we inverted the background from black to white and the foreground from white to black. Figure 2 (F) was also obtained from Figure 2 (C) where the groups of pixels identified with high intensity in relation to surrounding pixels [22]. The Figure 2 (G) is the complement of the Figure 2 (C) (dark areas become lighter and light areas become darker). This is useful for applying the watershed transform, which identifies clear points in further processing. Figure 2 (H) is a combination composed by the Figure 2 (E), (F) and (G). The image in Figure 2 (H) is modified from a logic operation OR

from the pixel values of the images Figure 2 (E) and Figure 2 (F). The intensity values in Figure 2 (G) are forced to become local minimums wherever the corresponding values resulting from logic operation OR with Figure 2 (E), Figure 2 (F) are non-zero. Thus, we obtain image in Figure 2 (H) as the input to the watershed algorithm, which is able to recognize regions according to the level of intensity. The watershed process returns regions found in the Figure 2 (I), then converted to a binary version where the regions are light areas from the Figure 2 (J). Finally, the polyp is positioned in one of regions on the Figure 2 (J).

Once the regions are returned by the watershed process in the target image, now is necessary choose in which region the polyp will be added. For this, a region size check is performed in terms of the number of pixels. Then, the largest region is selected as a candidate. This region will only receive the polyp if its area is not on a polyp that is already present in the target image. The watershed selection process finishes when the new polyp is positioned over the largest watershed region and no area is over the polyp that was already present in the target image. In this process, there may be cases where there are no compatible areas in the target image for polyp insertion. In that case, the image is discarded and a new sample from the original dataset is chosen as target image.

Finally, after the selection of a compatible watershed region, the area containing the polyp (Figure 1 (c)) is inserted into the target image (Figure 1 (d)). However, there are texture and lighting differences between the source image and the target image. This problem is minimized using the Poisson technique [13]. This approach allows polyp insertion despite the presence of complex structures or contours in new backgrounds (target image). The Poisson approach eliminates the need to delineate the polyp borders with high precision. It is interesting to use a bounding box so that regions close to the polyp in the source image are also obtained with the aim of favoring the smoothing of texture, lighting and color in the target image. This smoothing strategy aims to make the target image visually coherent, so it is not noticeable that a polyp belonging to a source image has been inserted. In addition, this method create automatically the ground truth image required by the deep learning algorithm.

IV. Experiments And Results

In this section, we describe the training and evaluation procedures. We compare the performance results between training the neural network with the original data, a traditional data-augmentation approach and our synthetically enhanced dataset.

A. Training and Testing Datasets

We use the 612 images from CVC-ClinicDB dataset to constitute training sets A and B, and also testing set A as shown in Table I. The ETIS-LaribPolypDB dataset contains 196 images and was used only for evaluation (testing set B in Table I). We propose two training datasets (i.e. training set C and D in Table I) in which polyps were inserted according to the process described in Section III. Therefore two polyps were inserted in the original CVC-ClinicDB dataset to form the training set C with 1071 images. The training set D was built following the same methodology, but seven polyps were added, resulting in 3823 images. Among the images that constitute the CVC-ClinicDB dataset, we selected one hundred images such that three or four images represent each polyp in the CVC-ClinicDB dataset. Thus, the 612 images were split into the training set A and testing set A (Table I). We perform the data augmentation strategies in the training set A and B to increase the number of training samples. For preprocessing, the resolution of all the images was reduced to 256x256 pixels.

TABLE I
SUMMARY OF THE TRAINING AND TESTING DATASETS.

Set Name	Dataset Used	Num. of Images
Train. Set A	CVC-ClinicDB	512
Train. Set B	CVC-ClinicDB	612
Train. Set C	CVC-ClinicDB + Two Polyps Added	1071
Train. Set D	CVC-ClinicDB + Seven Polyps Added	3823
Test. Set A	CVC-ClinicDB	100
Test. Set B	ETIS-LaribPolypDB	196

B. Evaluation Metrics and Experiments Setup

Table I list the evaluation metrics used to asses the training performance. We measure for each experiment the accuracy, precision, recall (sensitivity), specificity and F1 score. A description of these metrics can be found in [3]. Additionally, we include the false positive rate (FPR) metric to quantify how much the detected polyp location falls outside the ground truth.

Our polyp segmentation approach is based on the U-net architecture. Their output is a a binary segmentation map that provides the predicted polyp location for each input image. In the training stage, we used an Adam optimizer [9] with learning rate of 0.0001 and batch size equal to 4. We apply augmentation that comprises the transformations of rotate (40 degrees), zoom, shift horizontally and vertically (20%), and horizontal flip. We also use early stopping to stops the execution of the training process when there is no improvement in ten consecutive epochs. The experiments were carried out on Windows 10 operating system using a standard PC with a 2.80GHz Intel (R) Core (TM) i7-7700HQ CPU, 16GB of RAM and a NVidia GeForce GTX 1060 GPU 6GB.

C. Detection Results

We investigate the effect of our training sets C and D (i.e. with polyps inserted) in comparison with augmented versions of the training set A and B. The Table II depicts the performance evaluation results for the test performed in the CVC-ClinicDB dataset (testing set A). We can observe that the performance against almost all metrics is similar between the augmented training set A and the proposed training set C, except for FPR which was reduced to 0.005 with the use of this proposed dataset.

The results listed in Table III shows the comparison of the augmented training set and our proposed training set in the context of testing set B (ETIS-LaribPolypDB). In the comparison between training set B (augmented) and training set C (proposed), we note that an improvement of 12.5% for the accuracy and recall, 13.1% for specificity, 9.5% for F1-Score and FPR was reduced to 0.0292. We also observed the evaluation among the training set B (augmented) with 3825 images and training set D (proposed) with 3823 images (See Table III). Note that there are improvements in terms of accuracy and recall of 21.7%, specificity of 23.9% and F1-Score 14%, with reduction of FPR to 0.079 by considering our proposed training set D. The values obtained using the training set D (proposed) were higher in all metrics except precision that maintained a value close to that presented by training set B (augmented). The results show that our method is promising and superior in this evaluation than the ones that use traditional data augmentations techniques.

TABLE II
PERFORMANCE OF THE TRAINING DATASETS ON THE TESTING DATASET A (CVC-CLINICDB).

Training Set	Acc	Prec	Rec	Spec	F1	FPR
Train. Set A (Original) - 512 images	0.959	0.958	0.959	0.979	0.958	0.020
Train. Set A (Augmented) - 1152 images	0.962	0.961	0.962	0.981	0.961	0.018
Train. Set C (Proposed) - 1071 images	**0.985**	**0.985**	**0.985**	**0.994**	**0.985**	**0.005**

TABLE III
PERFORMANCE OF THE TRAINING DATASETS ON THE TESTING DATASET B (ETIS-LARIBPOLYPDB).

Training Set	Acc	Prec	Rec	Spec	F1	FPR
Train. Set B (Original) - 612 images	0.569	0.929	0.569	0.566	0.688	0.433
Train. Set B (Augmented) - 1071 images	0.581	0.934	0.581	0.576	0.697	0.423
Train. Set C (Proposed) - 1071 images	0.706	**0.939**	0.706	0.707	0.792	0.292
Train. Set B (Augmented) - 3825 images	0.681	**0.939**	0.681	0.681	0.774	0.318
Train. Set D (Proposed) - 3823 images	**0.898**	0.936	**0.898**	**0.920**	**0.914**	**0.079**

To see the difference clearly between the training sets proposed and augmented datasets Figure 3 depicts a comparison among the precision and recall found after testing using ETIS-LaribDB dataset. The number of images was adjusted to allow comparison of the training datasets with a similar amount of images. The obtained results compared to use of original training set B only, show improved performances in recall considering to training set C(proposed) and training set D(proposed). Specifically, the use of 3823 inserted polyp images (Training set D proposed) shows 21.7% of recall improvements compared to the use of augmented training set B with 3825 images. Overall, on the training set D, we can see considerable improvement for recall while maintaining compatible values of precision. The same comparison for the tests on the CVC-ClinicDB dataset with 100 images (testing set A, Table I) presented close values for both precision and recall for different training datasets as seen in Table II.

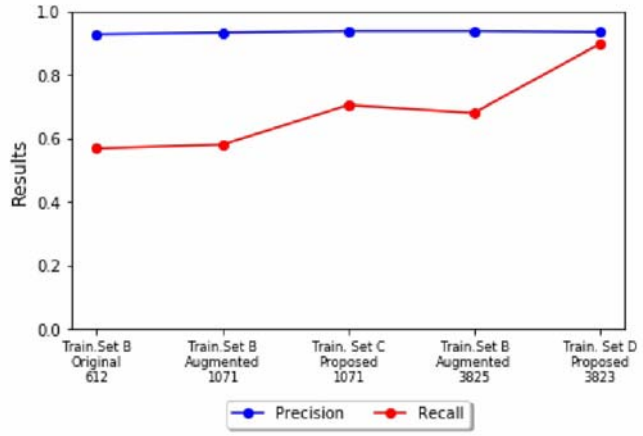

Fig. 3. The comparison from precision and recall values from testing set B results (ETIS-LaribPolypDB dataset).

The tests results (Table II and III) have shown to be effective in reducing false alarms. More specifically, in Table II, the lower values for FPR results were obtained because the original training set was modified by the polyp insertion process. We noticed that the FPR results (Table III) tend to be more consistent because we use the images from CVC-ClinicDB exclusively for the training and the ETIS-LaribPolypDB for testing. In the case of the tests on the ETIS-LaribPolypDB dataset (testing set B), FPR values also had a important reduction as presented in Figure 4. Results show how the use of the enhanced training set D (proposed) lead to a significant performance improvements.

The results showed that there is a significant reduction in the FPR. This means that the polyp insertion process contribute to the network learning regards the hard mimics, such as specular highlights, fecal content, and bubbles.

V. CONCLUSION

In this paper, we propose a method for automatic polyp insertions in colonoscopy images for enriching the training dataset and to improve the performance of convolutional neural network approaches in the polyp detection tasks. We evaluated our method with a training set of CVC-ClinicDB database and ETIS-LaribPolypDB testing set. The proposed method results

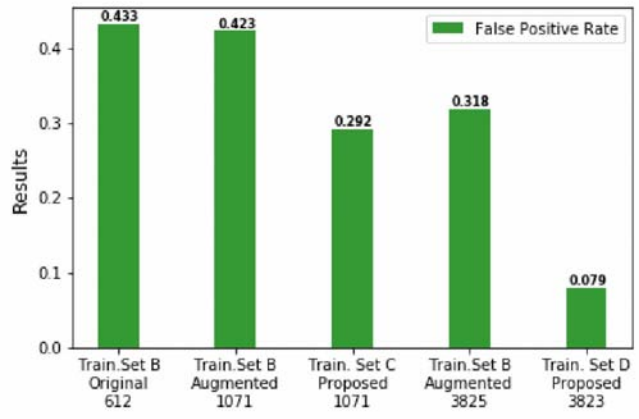

Fig. 4. The False Positive Rate values from Testing set B (ETIS-LaribPolypDB).

show that our polyp insertion process is effective to reduce false positives and can be a useful alternative to traditional data augmentation. The evaluation performed show a low FPR while maintaining a substantial recall/sensitivity. We achieve 91.4% of F1-Score and 0.079 of False Positive Rate with our modified training set over the ETIS-LaribPolypDB testing set. Additionally, it has the potential to increase the variability and number of samples in a reduced polyps dataset like CVC-ClinicDB. For future work, we plan create new polyps of varied shapes, sizes, and texture by means of a learning process considering all the polyps of the dataset.

REFERENCES

[1] A Axon, MD Diebold, M Fujino, R Fujita, RM Genta, JJ Gonvers, M Guelrud, H Inoue, M Jung, H Kashida, et al. Update on the paris classification of superficial neoplastic lesions in the digestive tract. *Endoscopy*, 37(6):570–578, 2005.

[2] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015.

[3] Jorge Bernal, Nima Tajkbaksh, Francisco Javier Sánchez, Bogdan J Matuszewski, Hao Chen, Lequan Yu, Quentin Angermann, Olivier Romain, Bjørn Rustad, Ilangko Balasingham, et al. Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge. *IEEE transactions on medical imaging*, 36(6):1231–1249, 2017.

[4] Mustain Billah, Sajjad Waheed, and Mohammad Motiur Rahman. An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. *International journal of biomedical imaging*, 2017, 2017.

[5] Michael Gschwantler, Stephan Kriwanek, Erich Langner, Bernhard Göritzer, Christiane Schrutka-Kölbl, Eva Brownstone, Hans Feichtinger, and Werner Weiss. High-grade dysplasia and invasive carcinoma in colorectal adenomas: a multivariate analysis of the impact of adenoma and patient characteristics. *European journal of gastroenterology & hepatology*, 14(2):183–188, 2002.

[6] John T Guibas, Tejpal S Virdi, and Peter S Li. Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872*, 2017.

[7] Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama. Gan-based synthetic brain mr image generation. In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 734–738. IEEE, 2018.

[8] Sae Hwang, JungHwan Oh, Wallapak Tavanapong, Johnny Wong, and Piet C De Groen. Polyp detection in colonoscopy video using elliptical shape feature. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, pages II–465. IEEE, 2007.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[10] SM Krishnan, X Yang, KL Chan, S Kumar, and PMY Goh. Intestinal abnormality detection from endoscopic images. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pages 895–898. IEEE, 1998.

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[12] Fernand Meyer. Topographic distance and watershed lines. *Signal processing*, 38(1):113–125, 1994.

[13] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. *ACM Transactions on graphics (TOG)*, 22(3):313–318, 2003.

[14] K. Pogorelov, O. Ostroukhova, M. Jeppsson, H. Espeland, C. Griwodz, T. de Lange, D. Johansen, M. Riegler, and P. Halvorsen. Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 381–386, June 2018.

[15] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015.

[17] Y. Shin, H. A. Qadir, L. Aabakken, J. Bergsland, and I. Balasingham. Automatic colon polyp detection using region based deep cnn and post learning approaches. *IEEE Access*, 6:40950–40962, 2018.

[18] Younghak Shin, Hemin Ali Qadir, and Ilangko Balasingham. Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access*, 6:56007–56017, 2018.

[19] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2017. *CA: a cancer journal for clinicians*, 67(1):7–30, 2017.

[20] Juan Silva, Aymeric Histace, Olivier Romain, Xavier Dray, and Bertrand Granado. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293, 2014.

[21] American Cancer Society. Key statistics for colorectal cancer, 2017. [Online]. Available: http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-key-statistics/.

[22] Pierre Soille. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[23] Bilal Taha, Naoufel Werghi, and Jorge Dias. Automatic polyp detection in endoscopy videos: A survey. In *Biomedical Engineering (BioMed), 2017 13th IASTED International Conference on*, pages 233–240. IEEE, 2017.

[24] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 79–83. IEEE, 2015.

[25] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE transactions on medical imaging*, 35(2):630–644, 2016.

[26] Gregor Urban, Priyam Tripathi, Talal Alkayali, Mohit Mittal, Farid Jalali, William Karnes, and Pierre Baldi. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*, 2018.

[27] Yi Wang, Wallapak Tavanapong, Johnny Wong, Jung Hwan Oh, and Piet C De Groen. Polyp-alert: Near real-time feedback during colonoscopy. *Computer methods and programs in biomedicine*, 120(3):164–179, 2015.