EIR - Efficient Computer Aided Diagnosis Framework for Gastrointestinal Endoscopies

Michael Riegler[*,•], Konstantin Pogorelov[*,•], Pål Halvorsen[*,•], Thomas de Lange[†,♣], Carsten Griwodz[*,•]
Peter Thelin Schmidt[‡,○], Sigrun Losada Eskeland[♣], Dag Johansen[♭]

[*]Simula Research Laboratory, Norway [†]Cancer Registry of Norway [‡]Department of Medicine, Karolinska Institute, Sweden
[•]University of Oslo, Norway [○]Center for Digestive Diseases, Solna and Karolinska University Hospital, Sweden
[♣]Bærum Hospital, Vestre Viken Health Trust, Norway [♭]The Arctic University of Norway, Norway

***Abstract*—Analysis of medical videos for detection of abnormalities like lesions and diseases requires both high precision and recall but also real-time processing for live feedback during standard colonoscopies and scalability for massive population based screening, which can be done using a capsular video endoscope. Existing related work in this field does not provide the necessary combination of detection accuracy and performance. In this paper, a multimedia system is presented where the aim is to tackle automatic analysis of videos from the human gastrointestinal (GI) tract. The system includes the whole pipeline from data collection, processing and analysis, to visualization. The system combines filters using machine learning, image recognition and extraction of global and local image features, and it is built in a modular way, so that it can easily be extended. At the same time, it is developed for efficient processing in order to provide real-time feedback to the doctor. Initial experiments show that our system has detection and localisation accuracy at least as good as existing systems, but it stands out in terms of real-time performance and low resource consumption for scalability.**

I. Introduction

During the last decades, we have witnessed a paradigm shift where computers and sensors move spatially closer and closer to the user, and we are in the process of moving devices inside the body. In this respect, our scenario is at the intersection of computer science and pathological medicine, where we target a scalable, real-time disease detection system for the gastrointestinal (GI) tract as it is depicted in figure 1. First, we study possible cancer precursors, e.g., polyps, and early cancer detection. Here, we develop both a computer-aided, live analysis system of endoscopy videos and a scalable detection system for screening systems using a wireless video capsule endoscope (VCE), i.e., a small capsule with an image sensor.

In the context of object or pattern detection and tracking in general images and videos, a lot of research has been performed, and current systems are good at detecting human faces, cars, logos, etc. However, detecting diseases in the GI tract is very different from detecting objects like cars. The GI tract can potentially be affected by a wide range of diseases with lesions visible in endoscopy, but findings may also include benign/normal or man-made lesions. The most common diseases are gastric and colorectal cancer (CRC), which are lethal when detected in a late stage (the 5-year survival rate ranges from 93% in stage I to 8% in stage IV [1]). Consequently, early detection is crucial. There are several ways of detecting pathology in the GI tract, but systematic population-wide screening is the most important tool for early detection. However, current methods have limitations regarding sensitivity, specificity, access to qualified medical staff and overall cost.

Fig. 1. The gastrointestinal (GI) tract (Image: kaulitzki/shutterstock.com).

In this scenario, both high precision and recall are of crucial importance, but so is the frequently ignored system performance that can provide feedback in real time. The most recent and most complete related work is the polyp detection system Polyp-Alert [2], which can provide near real-time feedback during colonoscopies. However, it is limited to polyp detection, and it is not fast enough for live examinations. To further aid and scale such examinations, we present EIR[1], an efficient and scalable automatic analysis and feedback system for medical data like videos and images. The system supports endoscopists in the detection and interpretation of diseases in the GI tract. EIR has initially been tested in scenarios supporting endoscopists in detection and interpretation of potential diseases in lower portions of the GI tract (large bowel). However, the main objective is to automatically detect abnormalities in the whole GI tract. Therefore, the aim is to develop both (i) a live system assisting the visual detection of, for example, polyps during colonoscopies and (ii) a future fully automated screening of the GI tract using VCEs. Both aims impose strict requirements on the accuracy of the detection to avoid false negative examinations (overlooking a disease) as well as low resource consumption. The live-assisted system also introduces a real-time processing requirement (defined as being able to process at least 30 frames or images per second). In this paper, the initial framework of our complete system is presented. To detect mucosal lesions in the colon, we built a system

[1]In Scandinavian mythology, EIR is a goddess with medical skill.

978-1-4673-8695-1/16/$31.00

combining filters using machine learning, image recognition and extraction and comparison of global and local image features. Furthermore, it is easy to add new filters or other types of data, such as patient records or sensor data, to increase accuracy or enable detection of other pathologies. Moreover, we evaluate our prototype by training classifiers that are based on different image recognition approaches. It is important to point out that these classifiers can also process other input like sensor data. We also test the generated classifiers with different data and thereby evaluate the different approaches for feasibility of colonic polyp recognition and localisation. The initial results from our experimental evaluation show that, (i) the detection and localisation accuracy can reach the same performance or outperform other current state-of-the-art methods and (ii) the system performance can reach real-time in terms of video processing up to high definition resolutions. Additionally, it is extensible with more data and diseases thorough parallel detection at run time. The rest of the paper is organized as follows: Firstly, in section II, we briefly introduce our medical case study. Next, we present related work in the field and compare it to the presented system in section III. This is followed by presenting the complete system in section IV. After that we, present an evaluation of the system in section V, and in section VI we discuss two cases where our system will be used in two medical examinations by our collaborators. Finally, we conclude with section VII.

II. Gastrointestinal Endoscopy

The GI tract illustrated in figure 1 can potentially be affected by various abnormalities and diseases, e.g., CRC, a major health issue world wide. Early detection of CRC or polyps as predecessors of CRC is crucial for survival, and several studies demonstrate that a population-wide screening program improves the prognosis and can even reduce the incidences of CRC [3]. As a consequence, in current European Union guidelines, screening for colorectal cancer is recommended for the age group over 50 [4]. Colonoscopy, a common medical examination and the gold standard for visualizing the mucosa and the lumen of the entire colon, may be used either as a primary screening tool or in a second step after positive screening tests [5]. However, endoscopies are invasive procedures and may lead to great discomfort for patients. Extensive training of physicians or nurses is required to perform the examination. They are performed in real-time and therefore challenging to scale to a large population. Additionally, the procedure is expensive. In the US, for example, colonoscopy is the most expensive cancer screening process, with annual costs of 10 billion dollars (1,100$-6,000$/person) [6], and with a time consumption of about one medical-doctor-hour and two nurse-hours per examination. As a first step, we target the detection of colorectal polyps, which are known precursors of CRC (see for example figure 2). The reason for starting with this scenario is that most colon cancers arise from benign, adenomatous polyps (around 20%) containing dysplastic cells, which may progress to cancer. Detection and removal of polyps prevents the development of cancer and the risk of getting CRC in the following 60 months after a colonoscopy depend largely on the endoscopist's ability to detect polyps [7]. Nevertheless, our system will be extended to support detection of multiple abnormalities and diseases of the GI tract by training the classifiers using different datasets.

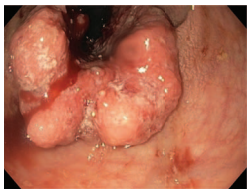

Fig. 2. Colorectal cancer that can be found using colonoscopy.

III. Related Work

Detection of diseases in the GI tract has mostly focused on polyps. This is most probably due to the lack of data in the medical field and polyps being a condition with at least some data available. However, none of the related work is able to do real-time detection or support doctors by computer-aided diagnosis during colonoscopies in real-time. Furthermore, all of them are limited to a very specific use case, which in the most cases is polyp detection for a specific type of camera. Table I gives an overview of the best working methods.

As one can see in Table I, several algorithms, methods and partial systems have been proposed and have, at first glance, achieved promising results in their respective testing environment. However, in some cases, it is unclear how well the approach would perform as a real system used in hospitals. Most of the research conducted in this field uses rather small amounts of training and testing data, making it difficult to generalize the methods beyond the specific dataset and test scenarios. Therefore, overfitting for the specific datasets can be a problem and can lead to unreliable results.

The first approach from Wang et al. [2] is the most recent and best-working one in the field of polyp detection. A list of more related work can be found in their paper. Polyp-Alert [2] is able to give near real-time feedback during colonoscopies. The system can process 10 frames per second and uses visual features and a rule-based classifier to detect the edges of polyps. Further, Polyp-Alert distinguishes between clear frames and polyp frames in its detection. The researchers report a performance of 97.7% correctly detected polyps, based on their dataset, which consists of 52 videos taken from different colonoscopes. Unfortunately, the dataset is not publicly available, and therefore, a detection performance comparison is not possible. Since neural networks (NN) are commonly used nowadays, they are also discussed in relation to the GI tract analysis. We identified two main points that make NNs less useful for our use case [17]. Firstly, (i) their training requires a lot of good training data, which is a big a problem in the medical field [18], and (ii) NNs are not easy to design for probabilistic results, which is important to support medical doctors during decision making [19].

In summary, a lot of good related work with interesting approaches for polyp detection exists. However, existing systems are either (i) too narrow for a flexible, multi-disease detection system; (ii) have been tested on limited datasets too small to show whether the method would work in a real

TABLE I
A PERFORMANCE COMPARISON OF POLYP DETECTION APPROACHES. NOT ALL PERFORMANCE MEASUREMENTS ARE AVAILABLE FOR ALL METHODS, BUT INCLUDING ALL AVAILABLE INFORMATION GIVES AN IDEA ABOUT EACH METHOD'S PERFORMANCE.

Publ./System	Detection Type	Recall / Sensitivity	Precision	Specificity	Accuracy	FPS	Dataset Size
Wang et al. [2]	polyp / edge, texture	97.70%	–	–	95.70%	10	1.8m frames
Wang et al. [8]	polyp / shape, color, texture	81.4%	–	–	–	0.14	1,513 images
Mamonov et al. [9]	polyp / shape	47%	–	90%	–	–	18,738 frames
Hwang et al. [10]	polyp / shape	96%	83%	–	–	15	8,621 frames
Li and Meng [11]	tumor / textural pattern	88.6%	–	96.2%	92.4%	–	–
Zhou et al. [12]	polyp / intensity	75%	–	95.92%	90.77%	–	–
Alexandre et al. [13]	polyp / color pattern	93.69%	–	76.89%	–	–	35 images
Kang et al. [14]	polyp / shape, color	–	–	–	–	1	–
Cheng et al. [15]	polyp / texture, color	86.2%	–	–	–	0.076	74 images
Ameling et al. [16]	polyp / texture	AUC=95%	–	–	–	–	1,736 images
EIR-system	abnormalities/30 features	98.50%	93.88%	72.49%	87.70%	30-65	18,781 frames

scenario and; (iii) provide a performance too low for a real-time system or ignore the system performance entirely. Last, but not least, we are targeting a holistic end-to-end system where a VCE that traverses the entire tract with its video signals is algorithmically analyzed.

IV. EIR BASIC IDEA

Our objective is to develop a system that supports doctors in disease detection in the GI tract. The system must (i) be easy to use and less invasive for the patient that existing methods, (ii) be easy to extend to different diseases, (iii) handle of multimedia content in real time, (iv) be usable for real-time computer-aided diagnosis, (v) achieve high classification performance with minimal false-negative classification results and (vi) have a low resource consumption. These properties potentially provide a scalable system with regard to cost, medical specialists required for a larger population, and number of users potentially willing to be screened. Therefore, EIR consists of three parts: The annotation subsystem, the detection and automatic analysis subsystem and the visualization and computer-aided diagnosis subsystem.

A. Annotation Subsystem

The purpose of the annotation subsystem is the efficient collection of training data for the detection and automatic analysis subsystem. It is well known that training data is very important for a good classification system. Nevertheless, in the medical field, the time of the experts and access to multimedia data are two resources that are quite limited. This is primarily because of high everyday workload for physicians, but also due to legal issues. For each image or video, patient consent has to be collected before research can be done, making it a very cumbersome task. Moreover, the annotation of videos itself is very time-consuming, and the quality of annotations depends on the experience and concentration of the physicians [20]. For example, in a VCE procedure, there are about $216,000$ images per examination, and a very experienced endoscopist needs at least 60 minutes to view and analyse all the video data [21]. Due to this limitation, it is important to develop automatic methods that can reduce the burden on physicians and speed up the screening process. We therefore developed an efficient semi-automatic annotation subsystem [22]. This annotation system is the entry point into our whole system. Since the medical doctor is usually located in a hospital with restrictions to data security, the implementation of the software is done with standard web technologies, which do not require any installation on the hospital's systems. This includes the storing of all information on the system-side and moves the responsibility of maintaining the system and the data integrity from the user to the system. Besides getting data for the EIR system to enable automatic screening, the annotation subsystem makes it possible to use the annotated videos in a medical video archive for documentation or teaching purposes.

B. Detection and Automatic Analysis Subsystem

These subsystems for algorithmic analysis are designed in a modular way, so that they can be extended to different diseases or subcategories of disease, as well as other tasks like size determination, etc. At the moment, this subsystem consists of two parts, the detection subsystem that detects irregularities in video frames and images and the localisation subsystem that localizes the exact position of the disease. The detection can not determine the location of the found irregularity. The location determination is done by the localisation subsystem. The localisation subsystem uses the output of the detection system as input.

1) Detection Subsystem: This part of the system is not designed to detect the precise position of an abnormality like a polyp or bleeding, but rather to detect whether there is something in the current frame of the video or not. All the frames that we process can be separated into two disjoint sets which can also be seen as the model for the detector. These two sets contain example images for abnormalities and images without any abnormality. Each of these sets can be seen as the model for a specific disease. The detection system is built in a modular way and can easily be extended with new models. This would for example allow to first detect a polyp and then to distinguish between a polyp with low or high risk to developing into CRC by using the *NICE* classification[2]. To compare and determine the abnormalities in a given video frame, we use global image features, i.e., because they are easy and fast to calculate, and because the exact position is at this point of the system not needed. In previous work, we showed that global features can indeed outperform or at least reach the same results as local features [23]. The basic idea is based

[2]http://www.wipo.int/classifications/nice/en/

on an improved version of a search based method for image classification presented in [23]. We create the indexes from visual features extracted from video frames or images. However, the number of needed examples is rather low compared to other methods. The index also contains information about the presence and type of any disease in the frame or image. A classifier can then search the index for the frames that are most similar to a given input frame. Based on the classification of the results, the detection subsystem then decides which abnormality the input frame belongs to. The whole detector is realised with two separate tools, an indexer and a classifier. We have released the indexer and the classifier as a separate project called *OpenSea*[3]. The computational nature of the indexing part is similar to what we know as batch processing. Therefore, creating the models for the classifier could be done off-line, and it is not influencing the real-time capability of the system, because it is only done once at the very first time when the training data is inserted into the system. Visual features to calculate and store in the indexes can be chosen based on the abnormality because, for different types of disease different set of features or combinations are better. For example, bleeding is easier to detect using color features, whereas polyps require shape and texture information.

The classifier can be used to classify video frames from an input video into as many classes as the detection subsystems model consists of. The classifier uses indexes generated by the indexer described before. In contrast to other classifiers that are commonly used, this classifier is not trained in a separate learning step. Instead, the classifier searches previously generated indexes, which can be seen as the model, for similar visual features. The output is weighted based on the ranked list of the search results. Based on this, a decision is made. The classifier is parallelized and allows to choose how many CPU cores are used. Ongoing work includes to port parts of the system to GPUs to further increase the performance.

2) Localisation Subsystem: The localisation subsystem is intended for exact positioning of a lesion, which is used to show markers on the frame or image containing the disease. This information is then used in the visualization subsystem. All images that we process during the localisation step come from the positive frames list generated by the detection subsystem. Processing of the images is implemented as a sequence of intraframe pre- and main-filters. Pre-filtering is needed because we use local image features to find the exact position of objects in the frames. Lesion objects or areas itself can have different shapes, textures, colors and orientations. They can be located anywhere in the frame and also partially be hidden and covered by biological substances, like seeds or stool, and lighted by direct and ambient light. Moreover, the image itself can be interleaved, noisy, blurry and over or under exposed, and it can contain borders and subimages. Apart from that, it can have various resolutions depending on the type of endoscopy equipment used. Endoscopic images usually have a lot of flares and flashes caused by high power light source located close to the camera. All these nuances affect the local features detection methods negatively and have to be specially treated to reduce localisation precision impact. In our case, several, sequentially applied filters are used to prepare raw input images for the following analysis. These analyses are RGB to YCbCr color space conversion, borders and subimages removing, flares masking and low-pass filtering. After the pre-filtering, the images are used for further analysis.

At the moment, we have implemented the detection of colon polyps using our local features approach. The main idea of this localisation algorithm is to use the polyps' physical shape to find the exact position in the frame. In most cases, the polyps have the shape of a hill located on relatively flat underlying surface or the shape of a more or less round rock connected to an underlying surface with stalks of varying thickness. These polyps can be approximated with an elliptically shaped region consisting of local features that differ from the surrounding tissue with high probability. To detect those types of objects, we use the following sequence of filters: binary noise reduction filter, 2D-gradient filter, threshold borders detection filter and binary noise removing filter. The next step creates a filtered binary image approximated by a set of ellipses from which we build energy maps based on the ellipse's size and border points precision approximation and matching. The final coordinates of one or more polyps in the frame are chosen by looking for the maximum in the energy map.

C. Visualization and Computer Aided Diagnosis Subsystem

This subsystem has two main purposes. First, it should help in evaluating the performance of the system and get insights into why things work well or not. Second, it can be used as a computer-aided diagnostic system for medical experts. Therefore, we have the TagAndTrack subsystem [22] that can be used for visualization and computer-aided diagnosis. We developed a web technology-based visualization that can be used to support medical experts while being very easy to use and distribute. This tool simply takes the output of the systems detection and localisation part and creates a web-based visualization, which can then be combined with a video sharing platform where doctors are able to watch, archive, annotate and share information.

V. Evaluation

For all of the subsequent measurements, we used the same computer (32 AMD CPU cores Linux server, 128GB ram). It is important to point out that the used hardware is quite old (ca. 4 years). Newer hardware would most probably lead to better performance for all the tests, but the evaluation shows that even on old hardware the system performs as intended. For all experiments, we used the ASU-Mayo Clinic polyp database[4]. This is currently the biggest publicly available dataset consisting of 20 videos from standard colonoscopies (converted from WMV to MPEG-4 for the experiments) with a total of 18,781 frames and different resolution up to full

[3]https://bitbucket.org/mpg_projects/opensea

[4]http://polyp.grand-challenge.org/

HD [24]. For the detection and localisation accuracy, we used the common standard metrics precision, recall/sensitivity and F1 score. We conducted a leave-one-out cross-validation to evaluate this part of the system, which is a method that assesses the generalization of a predictive model. In our case, it describes the process where the training and testing datasets are rotated, leaving out a single different non-overlapping item or portion for testing, and using the remaining items for training. This process is repeated until every item or portion has been used for testing exactly once [25].

EIR allows us to use several different global image features for the classification. The more image features we use, the more computationally expensive the classification becomes. Further, not all image features are equally important or provide equally good results for our purpose. As a first step, we therefore need to find out which image features we want to use for classification. In order to understand which image features provide the best results, we generated indexes containing all possible image features for all frames of all video sequences from the ASU-Mayo Clinic database. These indexes can be used for several different measurements and also for leave-one-out cross-validation. Using our detection system, the built-in metrics functionality can provide information on the performance of different image features for benchmarking. Further, it provides us with separate information for every single image feature, as well as the late fusion of all the selected image features. All used features are described in detail in [26].

Accuracy. Based on the evaluation of different combinations of image features using 30 different features and information gain analysis, the image features JCD and Tamura were identified to be the best ones for polyp detection. The last row of table I shows our approaches' performance to give a comparison. We achieve an average precision of 0.9388, an average recall of 0.9850, and an average F1 score value of 0.9613. In other words, the results mean that we can detect polyps with a precision of almost 94%, and we detect almost 99% of all polyp containing frames. If we compare this to the best performing system in table I, it seems that Polyp-Alert reaches slightly higher detection accuracy. But, our system is faster and can detect polyps in real-time. Furthermore, our system is not designed and restricted to detect only polyps, and can be expanded to any possible disease if we have the correct training data.To evaluate the performance of the localisation subsystem we used the exact positions of the polyps as provided in the ASU-Mayo clinic polyp database as ground truth. Overall, we reached for the localisation an average precision of 0.3207, a recall of 0.3183 and a F1 score of 0.3195.

Speed. We also performed some initial system performance tests. For all these tests, we used 3 videos from 3 different endoscopic devices and different resolutions. The three videos have the resolutions $1,920x1,080$, $856x480$ and $712x480$. We chose these videos to show the performance under different requirements that the system will have to face when it is used. As figure 3 shows, EIR reaches the required 30 frames per second with 16-26 CPUs. This is true for all three videos that we used. For the future an implementation using GPUs will be important to cope with the high number of needed cores.

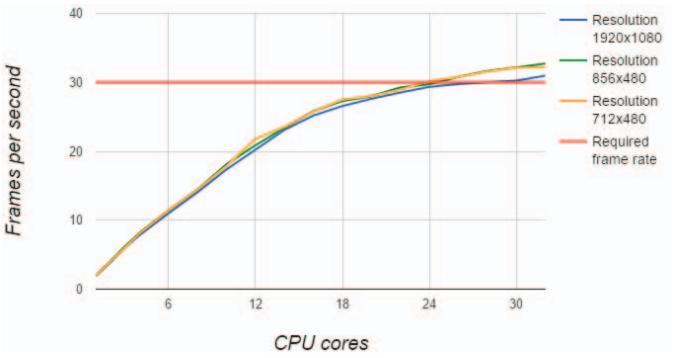

Fig. 3. Processing performance in frames per second.

VI. Real World Use Cases

In this section, we will describe two real world use cases where the presented system can be used. The first one is a live system that will support medical doctors during endoscopies. Currently, we are working on setting it up in one of our partner hospitals. The second one is a system that will automatically analyse videos captured by VCEs. Several hospitals all over Europe and US are involved in this part, and currently, we are collecting data. The first use case requires fast and reliable processing, and the second requires a system that is able to process a large amount of data in a reliable and scalable way.

Live System. Endoscopy is a common gastrointestinal examination and is essential for the diagnosis of most mucosal diseases in the gastrointestinal tract, particularly diagnosis of CRC and its precursors. Previous studies have demonstrated that a major challenge is the detection rate of lesions [27]. The aim of the live system is to provide live feedback to the doctors, i.e., a computer aided diagnosis in real-time. While the endoscopist performs the colonoscopy, the system analyses the video frames that are recorded by the colonoscope. At the beginning, we plan to optically show the physician (for example with a red or green frame around the video) when the system detects something abnormal in the actual frame or not. This can also be extended to the determination of what disease the system most probably detected and provide this information to the doctor. Apart from supporting the medical expert during the colonoscopy, the system can also be used to document the procedure. After the colonoscopy, an overview can be given to the doctors where they can make changes or corrections, and add additional information. This can then be stored for later purposes or used as a written endoscopy report. A demo of the live system is presented and described in [28]

Wireless Video Capsule Endoscope. The present VCEs have a resolution of around $256x256$ with 3-35 frames per second (adaptive frame rate with a feedback loop from the receiver to the transmitter). They do not have optimum lighting, making it difficult use the images. Nevertheless ongoing work tries to improve the state-of-the-art technology which will make it possible to use the methods and algorithms developed for colonoscopies also for VCEs [29]. The multi-sensor VCE

is swallowed in order to visualize the GI tract for subsequent diagnosis and detection of GI diseases. Thus, people may be able to buy VCEs at the pharmacy, and connect and deliver the video stream from the GI tract to the phone over a wireless network. The video footage can be processed in the phone or delivered to our system, which finally analyses the video automatically. In the best case, the first screening results are available within eight hours after swallowing the VCE, which is the time the camera typically spends traversing the GI tract.

VII. Conclusion

In this paper, a multimedia system for disease detection and classification in the GI tract has been presented. We briefly described the whole pipeline of the system from annotation (data collection for system learning) to visualization (doctor feedback). A detailed evaluation in terms of detection and localisation accuracy and system performance has been performed. These experiments showed that the proposed system can achieve equal results to state-of-the-art methods in terms of detection accuracy for state-of-the-art endoscopic data. Further, we showed that the system outperforms state-of-the-art systems in terms of system performance, that it scales in terms of data throughput and that it can be used in a real-time scenario. We also presented automatic analysis of VCE videos and live support of colonoscopies as two real-world use cases that will benefit from the proposed system and will actually be tested and used in our partner hospitals. For future work, we plan to improve the detection and localisation accuracy of the system and include more different abnormalities to detect. Presently, we are working with medical experts to collect more training data. Additionally, we currently work on the set-up of the real-world use cases in the hospitals. Finally, to further improve the performance of the system, we work on an extension that allows the system to use GPUs to further utilize the parallelization potential of the workload [30].

Acknowledgment

This work is founded by the FRINATEK project "EONS" #231687.

References

[1] J. B. O'Connell, M. A. Maggard, and C. Y. Ko, "Colon cancer survival rates with the new american joint committee on cancer sixth edition staging," *NCI*, vol. 96, 2004.

[2] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: Near real-time feedback during colonoscopy," *CMBP*, no. 3, 2015.

[3] O. Holme, M. Bretthauer, A. Fretheim, J. Odgaard-Jensen, and G. Hoff, "Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals," *CDSR*, 2013.

[4] L. von Karsa, J. Patnick, and N. Segnan, "European guidelines for quality assurance in colorectal cancer screening and diagnosis. first edition–executive summary," *Endoscopy*, 2012.

[5] S. Mallery and J. Van Dam, "Advances in diagnostic and therapeutic endoscopy," *Med Clin North Am*, vol. 84, no. 5, pp. 1059–83, 2000.

[6] The New York Times, "The $2.7 Trillion Medical Bill," http://goo.gl/CuFyFJ, [last visited, Nov. 29, 2015].

[7] M. F. Kaminski, J. Regula, E. Kraszewska, M. Polkowski, U. Wojciechowska, J. Didkowska, M. Zwierko, M. Rupinski, M. P. Nowacki, and E. Butruk, "Quality indicators for colonoscopy and the risk of interval cancer," *JM*, vol. 362, no. 19, pp. 1795–1803, 2010.

[8] Y. Wang, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen, "Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy," *In proc. of BHI*, vol. 18, no. 4, pp. 1379–1389, 2014.

[9] A. Mamonov, I. Figueiredo, P. Figueiredo, and Y.-H. Tsai, "Automated polyp detection in colon capsule endoscopy," *IEEE Transactions on Medical Imaging*, vol. 33, no. 7, pp. 1488–1502, July 2014.

[10] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. de Groen, "Polyp detection in colonoscopy video using elliptical shape feature," in *Proc. of ICIP*, Sept 2007, pp. 465–468.

[11] B. Li and M.-H. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and svm-based feature selection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 323–329, May 2012.

[12] M. Zhou, G. Bao, Y. Geng, B. Alkandari, and X. Li, "Polyp detection and radius measurement in small intestine using video capsule endoscopy," in *Proc. of BMEI*, Oct 2014, pp. 237–241.

[13] L. A. Alexandre, J. Casteleiro, and N. Nobreinst, "Polyp detection in endoscopic video using svms," in *Proc. of PKDD*, 2007, pp. 358–365.

[14] J. Kang and R. Doraiswami, "Real-time image processing system for endoscopic applications," in *Proc. of CCECE*, vol. 3, 2003, pp. 1469–1472.

[15] D.-C. Cheng, W.-C. Ting, Y.-F. Chen, Q. Pu, and X. Jiang, "Colorectal polyps detection using texture features and support vector machine," in *In proc. of MDAISM*. Springer, 2008, pp. 62–72.

[16] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Proc. of BfM*. Springer, 2009, pp. 346–350.

[17] C. Chin and D. E. Brown, "Learning in science: A comparison of deep and surface approaches," *Journal of Research in Science Teaching*, vol. 37, no. 2, pp. 109–138, 2000.

[18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[19] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[20] B. Giritharan, X. Yuan, J. Liu, B. Buckles, J. Oh, and S. J. Tang, "Bleeding detection from capsule endoscopy videos," in *Proc. of EMBS*, 2008.

[21] B. Li and M. Q. H. Meng, "Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments," *CBM*, vol. 39, no. 2, pp. 141–147, 2009.

[22] Z. Albisser, M. Riegler, P. Halvorsen, J. Zhou, C. Griwodz, I. Balasingham, and C. Gurrin, "Expert driven semi-supervised elucidation tool for medical endoscopic videos," in *Proceedings of the 6th ACM Multimedia Systems Conference*, ser. MMSys '15. New York, NY, USA: ACM, 2015, pp. 73–76. [Online]. Available: http://doi.acm.org/10.1145/2713168.2713184

[23] M. Riegler, M. Larson, M. Lux, and C. Kofler, "How 'how' reflects what's what: Content-based exploitation of how users frame social images," in *In proc. of MM*, ser. MM '14. New York, NY, USA: ACM, 2014, pp. 397–406. [Online]. Available: http://doi.acm.org/10.1145/2647868.2654894

[24] N. Tajbakhsh, S. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," 2015.

[25] B. Efron and R. Tibshirani, "Improvements on cross-validation: The .632+ bootstrap method," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. pp. 548–560, 1997. [Online]. Available: http://www.jstor.org/stable/2965703

[26] M. Lux and O. Marques, *Visual Information Retrieval Using Java and LIRE*. Morgan & Claypool, 2013.

[27] T. de Lange, S. Larsen, and L. Aabakken, "Image documentation of endoscopic findings in ulcerative colitis: photographs or video clips?" *GE*, vol. 61, no. 6, pp. 715–720, 2005.

[28] M. Riegler, K. Pogorelov, J. Markussen, M. Lux, H. K. Stensland, T. de Lange, C. Griwodz, P. Halvorsen, D. Johansen, P. T. Schmidt, and S. L. Eskeland, "Computer aided disease detection system for gastrointestinal examinations," in *Proc. of MMSys*, 2016.

[29] A. Khaleghi and I. Balasingham, "Wireless communication link for capsule endoscope at 600 mhz," in *Proc. of EMBC*, 2015, pp. 4081–4084.

[30] K. Pogorelov, M. Riegler, P. Halvorsen, P. T. Schmidt, C. Griwodz, D. Johansen, S. L. Eskeland, and T. de Lange, "GPU-accelerated real-time gastrointestinal diseases detection," in *Proc. of CBMS*, 2016.