Article published online: 2021-04-22

Real-time deep learning-based colorectal polyp localization on clinical video footage achievable with a wide array of hardware configurations

Authors

Jeremi Podlasek*,[1], Mateusz Heesch*,[1,2], Robert Podlasek[3], Wojciech Kilisiński[4], Rafał Filip[4,5]

Institutions

1 Department of Technology, moretho Ltd., Manchester, United Kingdom
2 Department of Robotics and Mechatronics, AGH University of Science and Technology, Kraków, Poland
3 Department of Surgery with the Trauma and Orthopedic Division, District Hospital in Strzyżów, Strzyżów, Poland
4 Department of Gastroenterology with IBD Unit, Voivodship Hospital No 2 in Rzeszow, Rzeszów, Poland
4 Department of Gastroenterology with IBD Unit, Voivodship Hospital No 2 in Rzeszow, Rzeszów, Poland
5 Faculty of Medicine, University of Rzeszów, Rzeszów, Poland

submitted 1.10.2020
accepted after revision 30.12.2020

Bibliography
Endosc Int Open 2021; 09: E741–E748
DOI 10.1055/a-1388-6735
ISSN 2364-3722

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Corresponding author
Mateusz Heesch, Department of Robotics and Mechatronics, AGH University of Science and Technology, al. Adama Mickiewicza 30, 30-059 Kraków, Poland
Phone: +794506751
heesch@agh.edu.pl

ABSTRACT

Background and study aims Several computer-assisted polyp detection systems have been proposed, but they have various limitations, from utilizing outdated neural network architectures to a requirement for multi-graphics processing unit (GPU) processing, to validating on small or non-robust datasets. To address these problems, we developed a system based on a state-of-the-art convolutional neural network architecture able to detect polyps in real time on a single GPU and tested on both public datasets and full clinical examination recordings.

Methods The study comprised 165 colonoscopy procedure recordings and 2678 still photos gathered retrospectively. The system was trained on 81,962 polyp frames in total and then tested on footage from 42 colonoscopies and CVC-ClinicDB, CVC-ColonDB, Hyper-Kvasir, and ETIS-Larib public datasets. Clinical videos were evaluated for polyp detection and false-positive rates whereas the public datasets were assessed for F1 score. The system was tested for runtime performance on a wide array of hardware.

Results The performance on public datasets varied from an F1 score of 0.727 to 0.942. On full examination videos, it detected 94% of the polyps found by the endoscopist with a 3% false-positive rate and identified additional polyps that were missed during initial video assessment. The system's runtime fits within the real-time constraints on all but one of the hardware configurations.

Conclusions We have created a polyp detection system with a post-processing pipeline that works in real time on a wide array of hardware. The system does not require extensive computational power, which could help broaden the adaptation of new commercially available systems.

Introduction

Colorectal cancer (CRC) is the second most common cancer in both sexes in Poland [1]. In the United States, it is the third most common in both occurrence and cancer-related death

* These authors contributed equally to this work.

count [2]. It has been shown that colonoscopy and endoscopic polypectomy are related to a lower occurrence and mortality rate of CRC [3–5]. This beneficial effect of colonoscopy is dependent on the quality of the endoscopic procedure [6, 7]. Adenoma detection rate (ADR) is the fraction of colonoscopic procedures in which at least one adenoma was found. It has been shown that ADR is inversely proportional to the frequency of interval cancer, the occurrence of advanced-stage CRC, and the mortality rate for this tumor [6, 7]. ADR can, among other methods, be increased by having an additional assistant or observer [8] participate in the procedure. Instead of that assistant, the help can come in the form of a computer-aided detection (CADx) system [9], which has already shown promise in gastroenterology research [10–19]. A lot of work has been done to create such software, ranging from simple hand-crafted feature-based [14] to utilizing the latest advances in computer vision, such as deep learning [12–13, 15–18]. However, the present deep learning systems have various limitations, such as requiring multiple graphics processing units (GPUs) [15] and utilizing dated neural network architectures [15–18].

This work aimed to present a polyp localization system that utilizes a state-of-the-art neural network architecture and is optimized to operate without the necessity for multiple GPUs, as well as to assess its performance on both full clinical procedure videos and multiple public image datasets. In addition, the system was benchmarked on a wide array of industrial- and consumer-grade GPUs to quantify its speed on various hardware setups.

Materials and methods

Data acquisition

To achieve our goals, we gathered 157 videos of 123 colonoscopy exams, totaling around 58 hours of footage. In total, 59 procedures had histopathological samples of the polyps examined and verified. This difference is connected to the retrospective fashion of data acquisition. All video footage had been gathered using OLYMPUS Q165 L and Q180 AL endoscopes. In addition, 2678 still photos of unique polyps were used as a supplementary source of training data, also acquired retrospectively. The data were fully anonymized and the footage was manually labeled by an experienced endoscopist (ADR > 50, procedures > 300). Special purpose software was created to speed up the process of labeling of the video material.

Data division

This material was split into training and validation sets. Together, they comprised 79,284 polyp frames extracted from 123 exams and 2678 additional polyp photos. The validation set contained 10 randomly selected exams, to avoid the presence of similar frames from a single time sequence within both sets.

The testing dataset comprised 42 videos of 34 colonoscopy procedures totaling 17 hours, separate from the training and validation footage, labeled by another experienced endoscopist (ADR > 50, procedures > 300). The histopathological samples were verified for 24 of these procedures. As an additional form of testing the system, we used four publicly available datasets: Hyper-Kvasir segmentation [20], CVC ClinicDB [21], CVC ColonDB [22], and ETIS-Larib Polyp DB [23].

▶ **Table 1** Histopathologically confirmed polyp types within the datasets.

Lesion type	Training	Testing
Tubular adenoma	57	26
Adenocarcinoma	5	0
Tubulovillous adenoma	8	2
Villous adenoma	2	0
Hyperplastic polyp	36	13
Inflammatory polyp	8	3
Mixed polyp	2	0
Serrated polyp	0	3

The distribution of histopathologically confirmed polyp types in the training (including validation) and testing datasets is presented in ▶ **Table 1**.

Data preparation

The individual frames extracted from the videos had their black edge areas cropped and were then resized to 224 by 224 pixels, as described below. The initial extraction rate was set to two frames from each second of the videos; however, at later stages, 45 videos had all individual frames extracted to increase the volume of the dataset as well as provide more instances of a naturally occurring motion blur, decreased focus, and lens contamination. The training data augmentation included rotations as well as flipping around middle vertical and horizontal axes, as well as both diagonals.

Performance constraints

The envisioned system should work alongside the medical professional to reduce examinator-related errors during a colonoscopy procedure. Bearing this in mind, we set a hard threshold of no less than 20 frames per second (FPS) to be processed by the system and fed to the video output. This number was an arbitrary choice and came from relaxing the standard 24 FPS video format to have slightly more leeway in terms of performance, bringing the max single-frame processing time to 50 ms, up from 40 ms proposed by Angermann et al. [19]. We also utilized some of the performance metrics proposed in that work, namely mean processing time and mean number of false-positives per frame (here in the form of false-positive rate).

For the system to be easy to incorporate into the broadest number of endoscopic processing units, we decided to use a single GPU as an intended hardware requirement. During this study, we successfully ran the system in real time on seven different models of Nvidia GPUs, as per ▶ **Table 2**. On a couple of models, the runtime exceeded 20 FPS more than twofold.

▶ **Table 2** Speed of the neural network on various Nvidia GPUs.

Nvidia GPU model	1050Ti	1060Ti	1080Ti	Tesla P100	Tesla K80	Tesla T4	Tesla P4
Mean processing time [ms]	40.4	29.5	22.4	29.3	55.3	17.6	35.12
Processing time standard deviation [ms]	0.23	0.33	0.49	0.29	0.16	0.27	0.08
Mean FPS	24.75	33.90	41.67	34.13	18.08	56.82	28.47

GPU, graphics processing unit; FPS, frames per second.

Localization system

We introduced a convolutional neural network (CNN) for polyp localization combining RetinaNet [24] and EfficientNet [25] architectures. The goal was to design a relatively lightweight model that achieved satisfying performance while complying with the aforementioned inference speed constraint. This was achieved by utilizing the state-of-the-art EfficientNet B4-model for the feature extraction phase within the network connected to the RetinaNet head, which offered a good trade-off between speed and object detection efficacy.

To further boost inference performance, Test Time Augmentation (TTA), a process in which for each frame, several slightly altered copies of it are processed and these individual predictions are then merged, was incorporated into the processing pipeline. Thanks to the power and parallelism of the GPU, it was possible to remain under the real-time limit with in-place flips and batch processing. This allowed us to take advantage of an ensemble-like strategy for each frame of the video, without linear scaling of memory consumption or inference time.

Training

The training was run on two Nvidia GeForce GTX 1080Ti GPUs with a considerably large input batch size of 96 images, which was made possible by taking advantage of gradient checkpointing. The network was trained from scratch as attempts to utilizing transfer learning resulted in subpar performance. Such training on a complete dataset with our hardware setup took approximately 2 weeks. During the process, we monitored the accuracy and F1-score with emphasis on the confusion matrix in order to gain insight not only about the broad performance of the system but also regarding class it struggled with. This was especially important due to the class imbalance within the dataset.

The neural network went through several iterations of training as we honed the training set. The first couple of iterations required adjustments to labeling, as the CNN ended up finding previously missed polyps in the training data. This also pointed to strong generalization as the net did not overfit to classify these examples per the provided labels.

Furthermore, analyzing the inference results led to the realization that the count of false positives (FP) is a major problem, and ideally should be addressed during the training stage. This was attempted with partial success via training dataset expansion and data augmentation. The expansion was possible because initially, we did not use all of the available no-polyp images because of the sheer number of them. Similar to the strategy posed by Angermann et al. [11], we sought out difficult examples among unused no-polyp images to be added to the training set. Progressively adding more of them, we ended up with a final training dataset with a 3:1 ratio of no-polyp to polyp images. To avoid destabilizing the training process by occasional batches of data containing solely no-polyp images, a data sampler was also used to ensure a certain distribution of data within every batch.

Last but not least, whenever the neural network had a satisfactory performance on the validation dataset, we ran inference on the unused no-polyp images, attempting to mine examples the system struggled with and adding them to the validation set to have a better indicator of what problems could be fixed via further augmentation and training set expansion.

Prediction signal post-processing

Additional post-processing (▶ **Fig.1**) was introduced to address three major problems:

- FP reduction
- Fading prediction flicker reduction
- Prediction size flicker reduction

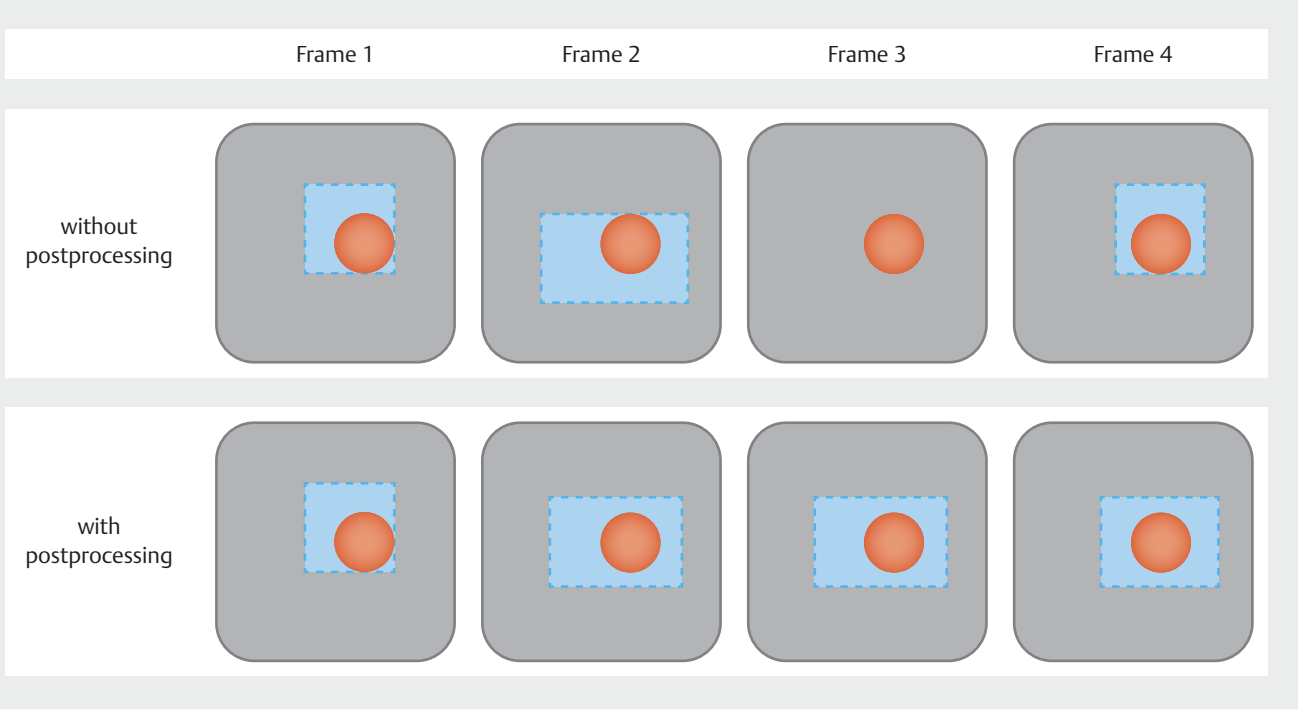
The latter two have to do with the presence of undesirable flicker leading to eye fatigue and potential irritation on the watcher's side [26]. Both of those may negatively impact the quality of the procedure, as well as the physical comfort of the operator.

False-positive reduction

As opposed to true-positives, many FP were highly dependent on the angle of viewing, often resulting in seemingly random short "flashes" of a prediction in a certain area. Such short occurrences can be easily filtered out by setting a requirement of having persistent prediction for a set amount of time. Note that such a solution does introduce a minor detection lag depending on the desired sensitivity; however, keeping it low at several frames makes the lag minuscule while considerably lowering the amount of the aforementioned "flashes."

Fading prediction flicker reduction

Another problem associated with an imperfectly fitted network is that sometimes, mostly depending on the angle or minor noise, a polyp may not be detected despite being in plain sight. Though it may not seem like a problem that an object is detected in, say, 20 of 24 (83 %) frames per second of video, the inter-



▶ **Fig. 1** Postprocessing effect visualization. The gray box represents the frame, the red circle the tracked object, and the blue box with dashed line represents the drawn bounding boxes.

mediate missing frames cause an on/off flickering of the prediction signal. This can be remedied in a similar way to the previous problem, by requiring a longer break in the prediction signal and effectively making the drawn bounding box persist where it was. Naturally, this will have the side effect of causing the prediction to linger for several frames after the object disappears from view; however, just as in the case of FP flicker reduction, the added comfort outweighs the minor drawback.

Prediction size flicker reduction

Finally, the fact that the neural network is not perfect leads the consequent predicted bounding boxes to vary in size around the object, despite the differences between each frame being negligible. In consequence, if one were to draw the predicted bounding boxes without any processing, while the center of the prediction should remain fairly static on the object, the width and height may flicker. To alleviate this, few previous frames can be taken into account when determining the position and shape of the bounding box in the current frame, effectively giving it some inertia and more fluent movement.

Results

The videos of 34 previously unseen examinations that comprised the testing set were processed by the system, resulting in videos with overlaid prediction boxes for where the system has detected a polyp. These processed videos were evaluated by the first endoscopist (ADR > 50, procedures > 300) who initially labeled training data and had not previously seen the testing set. The results were then compared against that of the second endoscopist (ADR > 50, procedures > 300) who initially labeled the testing set. Every test video was viewed separately by two endoscopists, which gives us considerable confidence in the visible polyps being identified. The model managed to correctly detect 79 of 84 (94.05 %) of the marked polyps, as well as two additional polyps that were missed. The breakdown of per-exam detections can be found in ▶ **Table 3**. Post-processing successfully filtered out most of the flickering FPs. The remaining ones were usually present for approximately 30 to 90 frames at a time. In total, in these 34 exams, there were 682 lingering FP detections, totaling 43,005 frames. With a total frame count of 1,434,595, this resulted in a FP rate of approximately 3 %.

In addition, as mentioned previously, the system was also evaluated on external datasets: HyperKvasir, CVC ColonDB, CVC ClinicDB, and ETIS-Larib achieving F_1-scores ranging from 0.727 to 0.942. More detailed results can be seen in ▶ **Table 4**.

The speed of the proposed system is presented in ▶ **Table 2**. Every value within the table was calculated based on 1000 inferences of the network, on a single GPU without incorporating multithreading. The speed constraints for real-time processing were met on all but one GPU model, with Tesla T4 and 1080Ti exceeding 40 FPS.

Discussion

Evaluation of the system yielded satisfying results on unseen colonoscopy video footage, detecting 94 % of polyps marked by the endoscopist. In addition, it also found two additional polyps the endoscopist had missed. The performance on inde-

▶ **Table 3** Comparison of polyp detection per examination between the endoscopist and the vision system.

Exam #	System	Endoscopist
1	1	1
2	2	2
3	**1**	**0**
4	3	3
5	6	6
6	2	2
7	4	4
8	1	1
9	4	4
10	5	5
11	6	6
12	2	2
13	2	1
14	2	2
15	2	2
16	3	3
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	2	2
23	2	2
24	1	1
25	3	3
26	3	3
27	**7**	**12**
28	1	1
29	2	2
30	2	2
31	1	1
32	2	2
33	2	2
34	2	2
Total	**81**	**84**

pendent datasets was highly variable, ranging from very good results on some datasets (CVC, Kvasir) to suboptimal on others (ColonDB, Etis). The suspected culprit is the fact that these images are coming from fundamentally different distributions than the training data: varying endoscopes, resolutions, color compositions, etc. This performance drop could be remedied by specialized pre-processing to bring the incoming images closer to what the neural net has already seen; however, it is also possible that this problem has to be addressed via a more diverse training dataset that would encompass a wider array of endoscopes and colon environments, resulting in a model with better generalization capabilities.

We compared the results of the system with Wang et al. [15] and Lee et al. [17] in ▶**Table 5** due to the similarity of the methodologies. Both systems were trained solely on internal datasets with comparable patient volume, in our case, videos from 123 exams (79284 polyp frames) and 2678 still photos of unique polyps; 5545 photos from 1290 patients for Wang et al. [15]; and 8075 photos of 503 polyps for Lee et al [17]. The models were also all evaluated on an independent, publicly available dataset, CVC-ClinicDB. As can be seen, the comparison of this dataset favors our system across the board, except for FPs, where Lee et al. measured 10 as compared to our 16.

Urban et al. [16] and Guo et al. [18] have published great results on polyp localization, with the former obtaining 96 % and the latter 100 % accuracy on their internal datasets. Unfortunately, these systems are not benchmarked on any public datasets, hence it is hard to make a meaningful comparison between ours and those systems.

Although the comparison of the results on internal datasets is not conclusively indicative of differences in performance of the systems, we do note that our system achieved a lower FP rate than most of the state-of-the-art, as can be seen in ▶**Table 6**. The proposed system showed higher FPR than Guo et al. [18]. The probable reason for this is the use of full examination footage for evaluation, instead of only the withdrawal part of the procedure. However, to have conclusive evidence of performance differences on such large-scale clinical data, all the systems would have to be evaluated on the same material.

When it comes to composing the training dataset for such systems, it is important to pay attention to both the volume and variety of data. Using multiple images of each unique polyp reduces the number of procedures needed to generate a given volume of data, and provides the system with multiple-angle views of those polyps. However, it is important to keep in mind that sufficient diversity of polyps captured is also necessary for good generalization capabilities. In our case, the initial dataset was expanded by additional database of polyp photos to ensure that diversity.

There are many published polyp detection and localization systems [14, 27–33], courtesy of GIANA Challenge [34], and the datasets that were made available to the public after its conclusion. However, these systems were all trained and validated on the same limited group of small datasets, usually CVC-ClinicDB, CVC-ColonDB, and ETIS-LARIB. Although this makes new architecture design comparison easy and meaningful, the same it not true when comparing them to systems created for clinical use. These are trained on a substantial amount of clinical data from an independent dataset, which may have a very different data distribution. These two groups are solving similar, although subtly different problems. The former attempts

▶ **Table 4** Results with various public polyp localization datasets.

Dataset	Images	Polyps	True positives	False negatives	False positives	Recall	Precision	F1-score
CVC ClinicDB	612	645	588	57	16	0.912	0.974	0.942
Hyper-Kvasir	1000	1063	938	125	24	0.882	0.975	0.926
CVC ColonDB	380	379	281	98	23	0.741	0.924	0.823
ETIS-Larib Polyp	192	208	140	68	37	0.673	0.790	0.727

▶ **Table 5** Comparison of results with the CVC ClinicDB dataset.

Model	True Positives	False negatives	False positives	Recall	Precision	F1-score
Wang et al.	570	76	42	0.882	0.931	0.898
Lee et al.	577	63	10	0.902	0.982	0.941
Podlasek et al.	588	57	16	0.912	0.974	0.942

▶ **Table 6** Comparison of false-positive rates.

Model	False-positive rate [%]
Guo et al.	1.6
Wang et al.	4.6
Lee et al.	6.3
Urban et al.	5.0–7.0
Podlasek et al.	3.0

to achieve the best performance on an isolated group of datasets, which often lack negative examples that outweigh the positive ones by a large margin in the clinical setting. The latter aims for good performance on its source of clinical data, with the performance on these public datasets not being the optimization goal, but rather, a side-effect of training data distribution and model generalization capabilities. Therefore, these between-group comparisons ultimately only provide context about how well these systems generalize.

Despite the efforts to eliminate FP predictions, some are still present. The cases the system struggled the most with were colon folds and overly illuminated areas, as well as bubbles and fecal matter. The majority of FP from our system came from the early phase of the colonoscopy before the examiner reached the cecum and inflated the colon to the desired, easier-to-inspect volume. While working alongside the endoscopist, the system would be engaged only after this stage, reducing the already low number of false alarms. On top of that, many of these false detections are ones that are signaled from a relatively far distance and are not repeated once the camera gets closer to the suspect area, bearing resemblance to the way humans approach such tasks: marking suspect areas from distance and ultimately making a decision based on closer inspection. This has also been observed by Wang et al. [15].

The speed constraint defined previously in this paper has been satisfied, with the system being able to run at over 24 frames per second even on comparatively slow GPUs (aside from Nvidia Tesla K80), up to around 56 frames per second on inference-optimized Nvidia Tesla T4.

The fact that the system missed some polyps that were found by the endoscopist while also correctly identifying ones that they missed reinforces the idea that this sort of software would be best utilized working in tandem with a medical professional, to ensure more effective overall detection. The immediate effects of using similar systems during procedures on the resultant ADR have already been investigated and show promising results; however, there are still many aspects that remain unaddressed, such as impact on the duration of the procedures, lasting effects on ADR, effects on work comfort, and effectiveness as a supervisory tool for trainees.

This work will be extended in several directions, both to check the feasibility of the system to give instant feedback to make the operator more proficient over time [35] and possibly to apply the same architecture to problems within other domains, such as urology. In its current iteration, the system can run in real time on a wide array of GPUs; however, filter pruning, weight quantization, and other techniques could make the system quick enough to not require a GPU in the first place. The next step is to verify whether the ADR of endoscopists using this system is increasing over time and to examine the precise long-term after-effects of such human-machine cooperation in a prospective clinical study. In fact, this study was the first step in such a planned study, which has already been approved by the Bioethics Committee.

As for the limitations of the study, the data were acquired in a retrospective fashion, and in effect, not every polyp underwent histopathological assessment. In addition, the presence of a polyp and later detection by the system were both overseen by a single endoscopist.

Conclusions

In conclusion, despite the aforementioned limitations, we created a polyp detection system with a post-processing pipeline to improve user comfort. The system ran within predefined constraints of real-time video processing and worked well even with low-resolution input and consumer-grade hardware. The system achieved a satisfying performance on unseen videos, detecting 94 % of the marked polyps, as well as finding additional ones. The performance on public polyp localization datasets ranged from F1-scores of 0.727 to 0.942 because the images in these datasets came from varied distributions. We believe this work can help with the creation and adaptation of new, commercially available CAD systems for polyp detection within the large intestine.

Acknowledgments

The authors thank Stanisław Lakoma, MD, and Mikołaj Podlasek, MD, for providing help with building a training database of polyps. We plan to release the dataset that was used in a standardized and anonymized format. Currently the de-identified data are available from the authors upon reasonable request and with permission of the Institutional Review Board.

Competing interests

J. Podlasek and M. Heesch are co-founders of moretho Ltd.

References

[1] Dahms S, Nowicki A. Epidemiology and results of treatment of colorectal cancer in Poland. Polish J Surgery 2015; 87: 598–605

[2] Siegel R, DeSantis C, Jemal A. Colorectal cancer statistics, 2014. CA Cancer J Clin 2014; 64: 104–117

[3] Brenner H, Chang-Claude J, Jansen L et al. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. Gastroenterology 2014; 146: 709–717

[4] Winawer SJ, Zauber AG, Ho MN et al. Prevention of colorectal cancer by colonoscopic polypectomy. N Engl J Med 1993; 329: 1977–1981

[5] Săftoiu A, Hassan C, Areia M et al. Role of gastrointestinal endoscopy in the screening of digestive tract cancers in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2020; 52: 293–304

[6] Kaminski MF, Regula J, Kraszewska E et al. Quality indicators for colonoscopy and the risk of interval cancer. N Engl J Med 2010; 362: 1795–1803

[7] Corley DA, Jensen CD, Marks AR et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014; 370: 1298–1306

[8] Buchner AM, Shahid MW, Heckman MG et al. Trainee participation is associated with increased small adenoma detection. Gastrointest Endosc 2011; 73: 1223–1231

[9] Wang P, Berzin TM, Brown JRG et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019; 68: 1813–1819

[10] Yang YJ, Bang CS. Application of artificial intelligence in gastroenterology. World J Gastroenterol 2019; 25: 1666

[11] Angermann Q, Aymeric H, Olivier R. Active learning for real time detection of polyps in videocolonoscopy. Medical Image Understanding and Analysis Conference 2016; 90: 182–187

[12] Otani K, Nakada A, Kurose Y et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. Endoscopy 2020; 52: 786–791

[13] Cho B-J, Bang CS, Park SW et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019; 51: 1121–1129

[14] Pogorelov K, Ostroukhova O, Jeppsson M et al. Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS) 2018: 381–386

[15] Wang P, Xiao X, Brown JRG et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nature Biomed Engineer 2018; 2: 741–748

[16] Urban G, Tripathi P, Alkayali T et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018; 155: 1069–1078

[17] Lee JY, Jeong J, Song EM et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. Sci Rep 2020; 10: 1–9

[18] Guo Z, Nemoto D, Zhu X et al. A polyp detection algorithm can detect small polyps: An ex vivo reading test compared with endoscopists. Digest Endosc 2020: doi:10.1111/den.13670

[19] Angermann Q, Bernal J, Sánchez-Montes C et al. Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis. In: Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures. Springer; 2017: 29–41

[20] Borgli H, Thambawita V, Smedsrud PH et al. Hyper-kvasir: A comprehensive multi-class image and video dataset for gastrointestinal endoscopy. Dec 2019. [Online]. Available from: osf.io/mkzcq, Accessed 2020 Jun 3

[21] 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 Graphics 2015; 43: 99–111

[22] Bernal J, Sánchez J, Vilarino F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 2012; 45: 3166–3182

[23] Silva J, Histace A, Romain O et al. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int J Comput Assist Radiology Surgery 2014; 9: 283–293

[24] Lin T-Y, Goyal P, Girshick R et al. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision 2017: 2980–2988

[25] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:190511946 2019.

[26] Wilkins AJ. Visual stress. Oxford University Press; 1995

[27] Yu L, Chen H, Dou Q et al. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. IEEE J Biomed Health Informat 2016; 21: 65–75

[28] Mohammed A, Yildirim S, Farup I et al. Y-net: A deep convolutional neural network for polyp detection. arXiv preprint arXiv:180601907 2018.

[29] Fan D-P, Ji G-P, Zhou T et al. Pranet: Parallel reverse attention network for polyp segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2020: 263–273

[30] Guo YB, Matuszewski B. Giana polyp segmentation with fully convolutional dilation neural networks. Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer

Graphics Theory and Applications: SCITEPRESS-Science and Technology Publications 2019: 632–641

[31] Sornapudi S, Meng F, Yi S. Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps. Applied Sci 2019; 9: 2404

[32] Qadir HA, Balasingham I, Solhusvik J et al. Improving automatic polyp detection using cnn by exploiting temporal dependency in colonoscopy video. IEEE J Biomed Health Informat 2019; 24: 180–193

[33] Dijkstra W, Sobiecki A, Bernal J et al. Towards a single solution for polyp detection, localization and segmentation in colonoscopy images. In: VISIGRAPP (4: VISAPP). 2019: 616–625

[34] Bernal J, Tajkbaksh N, Sánchez FJ et al. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. IEEE Transact Med Imaging 2017; 36: 1231–1249

[35] Gurudu SR, Boroff ES, Crowell MD et al. Impact of feedback on adenoma detection rates: Outcomes of quality improvement program. J Gastroenterol Hepatol 2018; 33: 645–649