Computers and Electrical Engineering 81 (2020) 106531

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Application of deep learning for autonomous detection and localization of colorectal polyps in wireless colon capsule endoscopy☆

Esmaeil S. Nadimi[a,*], Maria M. Buijs[b,c], Jurgen Herp[a], Rasmus Kroijer[b,c], Morten Kobaek-Larsen[b,c], Emilie Nielsen[d], Claus D. Pedersen[d], Victoria Blanes-Vidal[a], Gunnar Baatrup[b,c]

[a] Group of Applied AI and Data Science, Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark
[b] Department of Surgery, Odense University Hospital, Odense, Denmark
[c] Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[d] Center for Innovative Medical Technology, Odense University Hospital, Odense, Denmark

ARTICLE INFO

Article history:
Received 12 November 2018
Revised 16 September 2019
Accepted 26 November 2019
Available online 20 December 2019

Keywords:
Colorectal polyps
Colon capsule endoscopy (CCE)
Deep learning
Machine learning
Convolutional neural networks

ABSTRACT

Recent advances in deep learning have prompted a surge of interest in analysis of medical images. In this study, we developed a convolutional neural network (CNN) for autonomous detection of colorectal polyps, in images captured during wireless colon capsule endoscopy, with risk of malignant evolution to colorectal cancer. Our CNN is an improved version of ZF-Net which uses a combination of transfer learning, pre-processing and data augmentation. We further deployed our CNN as the basis for a Faster R-CNN to localize regions of images containing colorectal polyps. We created an image database of 11,300 capsule endoscopy images from a screening population, including colorectal polyps (any size or morphology, N=4800) and normal mucosa (N=6500). Our CNN scored an accuracy of 98.0%, a sensitivity of 98.1% and a specificity of 96.3%. Our network outperforms all state-of-the-art results in autonomous detection of colorectal polyps and shows high interpretability in terms of sensitive regions.

1. Introduction

1.1. Setting the stage

Colorectal cancer (CRC) is one of the most preventable forms of cancer, if colorectal polyps are detected and removed. Several studies have shown that during screening program, adenocarcinomas have been detected in between 3% and 4.6% of those who underwent colonoscopy following a positive immunological fecal occult blood test (iFOBT) result [1]. As a result, there has been a significant increase in the number of referrals for colonoscopy to detect and treat colorectal polyps and early stage adenocarcinoma. Nationwide CRC screening is offered in many developed countries, but according to the statistics reported in [2], only 60% of people at risk participate in screening, most likely because iFOBT and Optical Colonoscopy (OC)

☆ This paper is for CAEE special section SI-mip. Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Li He.

* Corresponding author.
E-mail address: esi@mmmi.sdu.dk (E.S. Nadimi).

https://doi.org/10.1016/j.compeleceng.2019.106531
0045-7906/

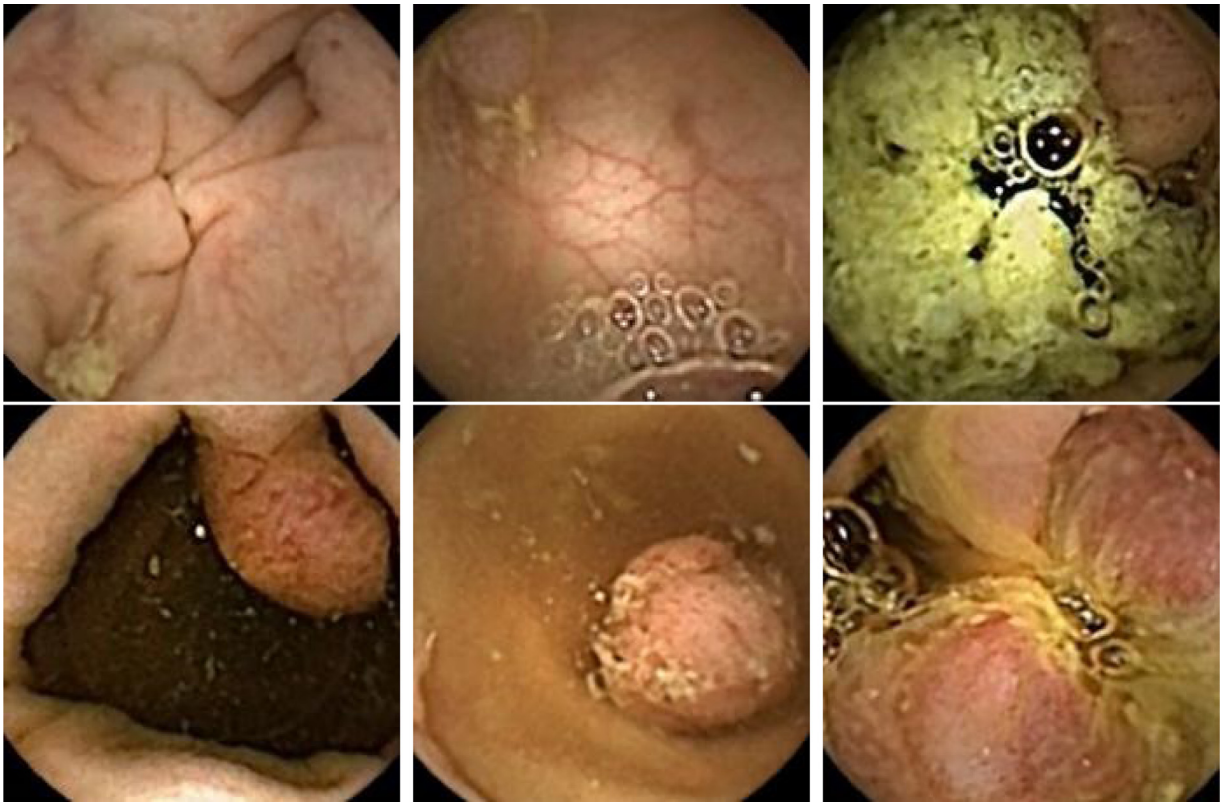

Fig. 1. Colorectal polyps in different stages of neoplasia and different grades of bowel cleanliness.

suffer from shortcomings in sensitivity, specificity, and patient acceptance due to pain and discomfort. Unlike OC, Colon Capsule Endoscopy (CCE) allows performing diagnosis using a non-invasive untethered monitoring device (also known as camera pill) carried through the GI tract via peristalsis. The camera pill is a vitamin size ingestible robot of size $31mm \times 11mm$ equipped with a processor, illumanting lights, power supply, radio communication systems and two cameras, capturing images of the inner lining of the Gastrointestinal (GI) tract.

The probability of high-grade neoplasia and of carcinomatous transformation increases with polyp size, especially when they are larger than 10 mm [1]. To improve the quality of the national screening program in Denmark in terms of 1) diagnostic accuracy, 2) polyp detection rate and 3) per-patient sensitivity for polyps only larger than 10 mm, we conducted several clinical trials in which individuals who had positive iFOBT results during screening underwent investigator blinded CCE and OC. Participants underwent repeated endoscopy if significant lesions were detected by CCE, and were considered missed by OC [2]. We refer the interested reader for a detail description of the study and the comparison between the performance of CCE and OC to our recent publications [2] and [3].

The results of our studies have indicated that in general, CCE is better than OC for polyp detection, both when comparing the total number of polyps, and when performing a paired comparison of the number of polyps detected per patient which could be contributing to OC false-negatives rather than to CCE false-positives [3]. However, the task of manually reviewing the vast number of images produced by CCE examinations (approximately 50,000 images per patient), and looking for the few images containing lesions is a laboursome task. To tackle this major drawback, several research units have suggested innovative schemes for autonomously discriminating endoscopic images based on the existence of polyps.

1.2. Previous work

Extraction of discriminative features from low-resolution medical images is one of the most challenging tasks in object recognition. Determining optimal features for a specific task at hand, based on the attributes of objects to be recognized is hard, mainly due to the ambiguity and lack of general task-independent rules for optimal feature selection. Polyp detection in CCE videos is a perfect example of such a task where tens of thousands of low-resolution images are generated [4]. The task at hand is then to detect colorectal polyps, which typically do not have common morphology, size, texture and color features (Fig 1), from a single patients video [3]. In addition, variable lighting and infrequent occurrence of polyps in a given CCE video create immense difficulties in devising a robust and data-driven method for reliable detection and segmentation process.

Although a rich body of literature surronds polyp detection in colonoscopy videos, a major study addressing similar challenges in images retrieved from CCE does not exist. Images obtained from CCE procedure are significantly lower in resolution compared to those of OC. In addition, publicly available databases such as Kvasir (A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection) had a significant imapct on the progress of polyp detection in images retrieved from OC. However, a similar database featuring colorectal polyps retrieved from CCE does not exist. Examples of modern polyp detection algorithms in colonoscopy videos are [5], [6], [7], [8] and [9].

One of the earliest publications in automatic polyp detection is the work of [10], who utilized texture spectrum with adaptive neuro-fuzzy-based classifier. The authors reported 97% sensitivity on 140 images with 70 polyp frames. Another highly relevant publication is [11] where two different shape descriptor features were compared to discriminate polyps from polyp-free colorectal regions. The algorithm was tested on 300 images out of which 150 contained polyps and accuracy of 86% was reported. In the study conducted by [12], the authors used geometrical features via ellipse fitting and multi scale rotation invariant local binary patterns and histogram of oriented Gaussian texture features. The reported 64.8% true positive rate in a total number of 27,984 frames with 12,984 polyp frames makes this study the largest in number of images, so far. The large number of polyp frames was obtained by perturbing original 541 polyp frames to obtain 12,443 extra samples for training. In a pilot study, we derived a novel algorithm based on random matrix theory for texture classification of uncalibrated images to find universality classes corresponding to the empirical spectral density of the covariance matrix of image intensity matrix [13]. Our dataset contained 344 images of the inner lining of the large bowel with at least one polyp present in the image. Classification accuracy of our algorithm was approximately 98%. Last but not least, [14] utilized a new texture feature to characterize the images by integrating advantages of wavelet transform and uniform local binary pattern with support vector machines as classifier. Their dataset contained a total of 1200 images (600 images of polyps and 600 of polyp-free tissue) and an accuracy of 91.6% was achieved. For a complete list of publications, we refer the interested reader to the excellent review paper [4].

The major drawbacks of all the reported algorithms are three fold: 1) Overfitting, since the proposed methods were tuned to obtain the best possible detection accuracy results for their corresponding datasets; 2) small sample size, since 70%-80% of the images were used for training the algorithm while the reported accuracies were obtained on the remaining 20%-30% of the images and 3) lack of algorithm validation on full CCE videos featuring various grading scales of colon cleanliness. In this study, we addressed these shortcomings and developed a deep learning algorithm based on convolutional neural networks (CNN) for autonomous detection and localization of colorectal polyps of any size or morphology. The important advantages of our study are: 1) large testing dataset as 1695 images unseen by the network were allocated to the test set, 2) during the training process, the network self-determined the optimal set of features from the data, 3) using data augmentation, our algorithm is both rotation and translation invariant and 4) we tested the performance of our network on 3 full CCE videos (each featuring approximately 50,000 images) that were not seen by the network during training, validation or testing process.

The organization of the paper is as follows: Section 2 presents the methods deployed in this study, including design and the development of network architecture and construction of the database. Section 3 presents results and discussions while conclusions of the study and future works are provided in Section 4.

2. Methods

2.1. CNN architecture

The LeCun paper [15] on the recognition of hand-written digits based on a pioneering 7-level convolutional network (LeNet-5) is both remarkable and precocious. Inspired by that, we evaluated five network architectures: an AlexNet / ZF-Net [16], a GoogLeNet [17], a ResNet50 [18], a VGG16 [19] and a VGG19 [19]. These networks construct a hierarchical representation of input images where deeper layers contain higher-level features constructed using the lower-level features of earlier layers. These publicly available networks are trained on 1000 objects and therefore have learned rich feature representations for a wide range of images. However, none of the objects look like colorectal polyps and in light of that, training a network from scratch would be the most reliable strategy. However, a dataset of insufficient size is the limiting factor for this approach and instead, the pre-trained CNNs were used either as an initialization for a fixed feature extractor or transfer learning, for the task at hand.

In this study, we conducted a preliminary experiment where we allocated 10% of the database for testing the performance of different networks. We observed that in terms of sensitivity and specificity, pre-trained ZF-Net generally outperformed the other five networks (Table 1) and therefore, we presented this network architecture in greater detail. The network comprises of 25 layers where 8 of them contain learnable weights featuring 5 convolutional and 3 fully connected layers. Each convolutional layer contains 96 to 384 kernels of variable size 3×3 to 11×11 with a depth of 3 up to 256. To build robustness to intra-class deformations and to avoid overfitting, Max-Pooling kernels of size 3×3, Rectified Linear Unit (ReLU) nonlinear activation function and dropout of 50% were used. The last fully-connected (FC) layer was replaced to output two classes, namely Polyps and No Polyps. For more details, we refer the interested reader to [16] and [20].

To solve a new classification problem and depending on the use of feature extraction scheme or transfer learning, networks last layer or last three layers, respectively, must be fine-tuned. In this study, applying feature extraction scheme re-

Table 1
Performance of different networks on test dataset.

Network	Accuracy%	Sensitivity%	Specificity%
Our modified Network	98.0	98.1	96.3
AlexNet	74.1	92.3	82
GoogleNet	51.2	13.2	99.4
ResNet50	69.7	80.7	99.3
VGG16	63.5	42.4	85.6
VGG19	82.7	68.8	90.2

sulted in detection and classification accuracy of 54%, performing as poorly as a random classifier. A significant improvement in networks performance was observed when transfer learning was used (Table 1).

To improve networks performance, we further optimized networks architecture and used transfer learning. We modified the network and deployed stochastic gradient descent with momentum (SGDM) as the optimization algorithm. The learning rate was initially set to 1e-4, but adaptively modified during the training process until the validation criteria were met. The epoch size for training process was limited to max of 6, leading to 787 iterations per epoch. Validation frequency and patience were set to 3 and infinity, respectively. The dataset containing images of both the mucosa of the large bowel without polyps and those containing colorectal polyps was augmented by rotating them randomly. To achieve both rotation and translation invariance, the images were randomly rotated (4 angles) and translated (by a vector $[u, v]$, $1 \leq u, v \leq 3$, 8 directions) using data augmentation without affecting the contents or the size of the images. To enhance the contents around polyps in the images, proportional padding was used. Finally, to overcome the generalization gap while keeping the limited GPU memory in mind, mini-batch size of 10 was selected.

2.2. Study design

This research is a sub-study of a double blinded longitudinal trial including 255 patients that were FIT-positive (national screening program in Denmark) during a period of one year (2015–2016). The participants underwent both CCE and OC in two consecutive days. Prior to the day of undergoing OC, the participants were investigated by a second-generation camera pill capsule endoscopy (PillCam COLON 2 Medtronic, Minnesota, USA). Throughout both CCE and OC investigations, polyp size, morphology and location were identified but the observers were blinded for the results of the other examination. The study was approved by the Local Ethics Committee (S20140141) and registered at clinicaltrials.gov (NCT02303756). For a more detailed study design, the interested readers are referred to our previous publications on this study listed through [2] and [3].

2.3. Image acquisition and database formation

The images used in this study were obtained from CCE videos that contained at least one colorectal polyp of any size or morphology. The videos were read in RAPID reader software (Given Imaging, Israel) and analyzed frame-by-frame by trained nurses and gastroenterologists, who identified existing polyps. The length of videos varied among patients, ranging from 40 min to 4 hours. The study resulted in an original database (before augmentation) of 1200 distinct images of colorectal polyps and 1625 images of normal mucosa.

Using data augmentation, we created a database containing 11,300 CCE images of dimension $576 \times 576 \times 3$ (RGB) with different grades of colon cleanliness, of which 4800 of them contain colorectal polyps of various sizes and morphologies. Images of normal mucosa featured different degrees of cleanliness and air bubbles. To regularize the network, reduce overfitting, ensure that it is rotation and translation invariant and help remedy the scarcity of data, we augmented the images by random rotation in four directions. We used rotation, cropping, and mirroring transformations to increase the effective size of our dataset. For each training image, we performed four random rotations (in addition to the original database) and sampled one random crop per rotation offline; effectively increasing the size of the training set by a factor of 5. We also performed random mirroring at training time. The augmented images containing polyps were all checked for contents, ensuring that polyps were not mistakenly cropped out. These augmentations are justified since masses have no inherent orientation and their diagnosis is invariant to these transformations.

We split the images randomly by patient into training, validation and testing sets (70%, 15% and 15% of the full dataset, respectively), constraining the validation and test sets to be balanced.

2.4. Region-CNN (R-CNN) network architecture

Object detection networks depend on region proposal algorithms to hypothesize object locations. The main difference between these networks is the process of selecting the regions and how they are classified. R-CNN and Fast R-CNN use a region proposal network (RPN) algorithm as a pre-processing step before running the CNN. The proposal algorithms such as EdgeBoxes [21] or Selective Search [22] are independent of the CNN and they become the processing bottleneck compared to running the CNN. Faster R-CNN addresses this issue by implementing the region proposal mechanism using the

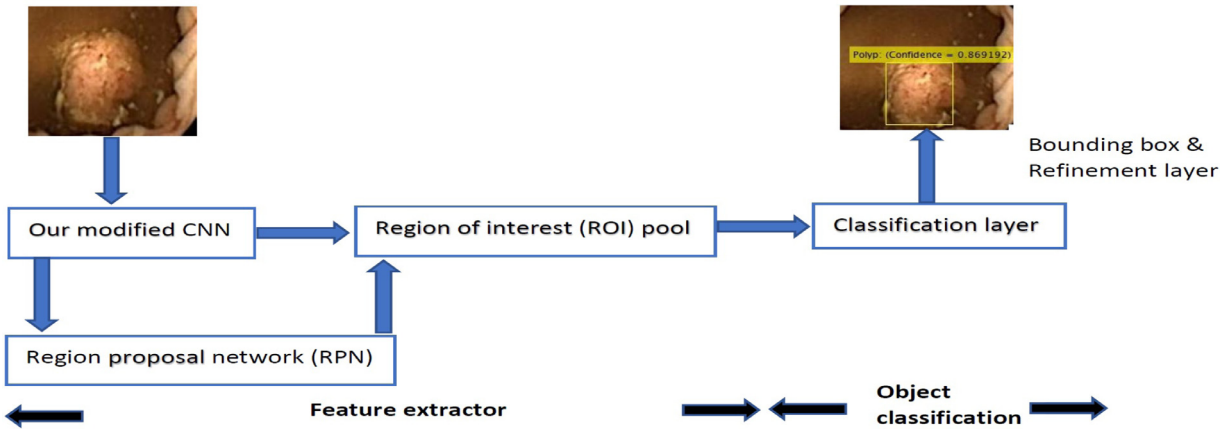

Fig. 2. Faster R-CNN detector using Region Proposal Networks.

CNN and thereby making region proposal a part of CNN training and prediction steps. For that reason, we developed a Faster R-CNN that benefits from sharing convolutional features with our developed CNN by altering the optimization process (Fig 2) [23]. The main task of this network was to localize colorectal polyps and suggest a bounding box around them. To develop our Faster R-CNN, we only selected a portion of our image database (1130 images equivalent to 10% of the database). These images were manually labeled and bounding boxes were created around each polyp that appeared in the image. We further split the images to 70% for training, 15% for validation and 15% (i.e., 170 images) for testing our Faster R-CNN network. This was due to the tedious task of labelling, drawing ground truth bounding boxes and finally creating the database for our RPNs. Each image contained bounding boxes surrounding colorectal polyps. Each bounding box was in the format [*x, y, width, height*]. The format specified the upper-left corner location and size of the object in the corresponding image.

The Faster R-CNN was trained in four steps, namely: Step1) Training an RPN, Step2) Training a Fast R-CNN Network using the RPN from Step1, Step3) Re-training RPN using weight sharing with Fast R-CNN, Step4) Re-training Fast R-CNN using updated RPN. The first two steps trained the region proposal and detection networks, while the final two steps combined the networks from the first two steps such that a single network was created for detection. We deployed the CNN developed in this study (series network). We further transformed the CNN into a Faster R-CNN network by adding an RPN, a Region of Interest (ROI) max pooling layer, and new classification and regression layers to support object detection.

The learning rate for the first two steps was set to 1e-3 while that of the last two steps was 1e-4. The mini-batch size was set to 1, which processed multiple image regions from one training image every iteration. The number of image regions per image were controlled by two distinct parameters, namely, Positive and Negative training samples. These two values were set to overlap with the ground truth boxes by a factor of [0.6-1.0] and [0-0.3], respectively. To consider a bounding box as a true positive containing a polyp, we selected a threshold value of 0.7 for the Intersection of Union (IoU) measure, which is a good threshold for computing Intersection over Unions for various bounding boxes.

In this study, confidence (precision) was defined as the ratio of true positive instances to all positive instances of objects in the detector, based on the ground truth, referring to the strongest detection scores. A higher score indicated higher confidence in the detection, and that the bounding box was more likely to contain a polyp.

3. Results and discussions

Empirical analyses and quantitative results of all five CNNs are presented in Table 1. We further visualized saliency maps to provide interpretability of the model to show the regions of an image that the network is sensitive to, when making predictions (Fig 3). The performance of our trained Faster R-CNN is also presented in Fig 4. The processing was carried out using NVIDIA P6000 Quadro GPUs, and MATLAB R2018a was used as the analysis software.

3.1. Empirical analysis and performance

The performance of different networks is presented in Table 1, showing the superiority of our modified network in terms of accuracy, sensitivity and specificity. Our modified ZF-Net outperformed the other four networks by a remarkable margin. A high sensitivity is of great importance, since the cost of false negatives (i.e. missed polyps) is much higher than false positives. Fig 5 presents some examples of misclassified images where either a polyp was missed (false negative or type II error) or the normal mucosa was recognized as a polyp (false positive or type I error). We observed two scenarios that frequently lead to type II error. The first group of missed polyps shared the common property of being very small in size

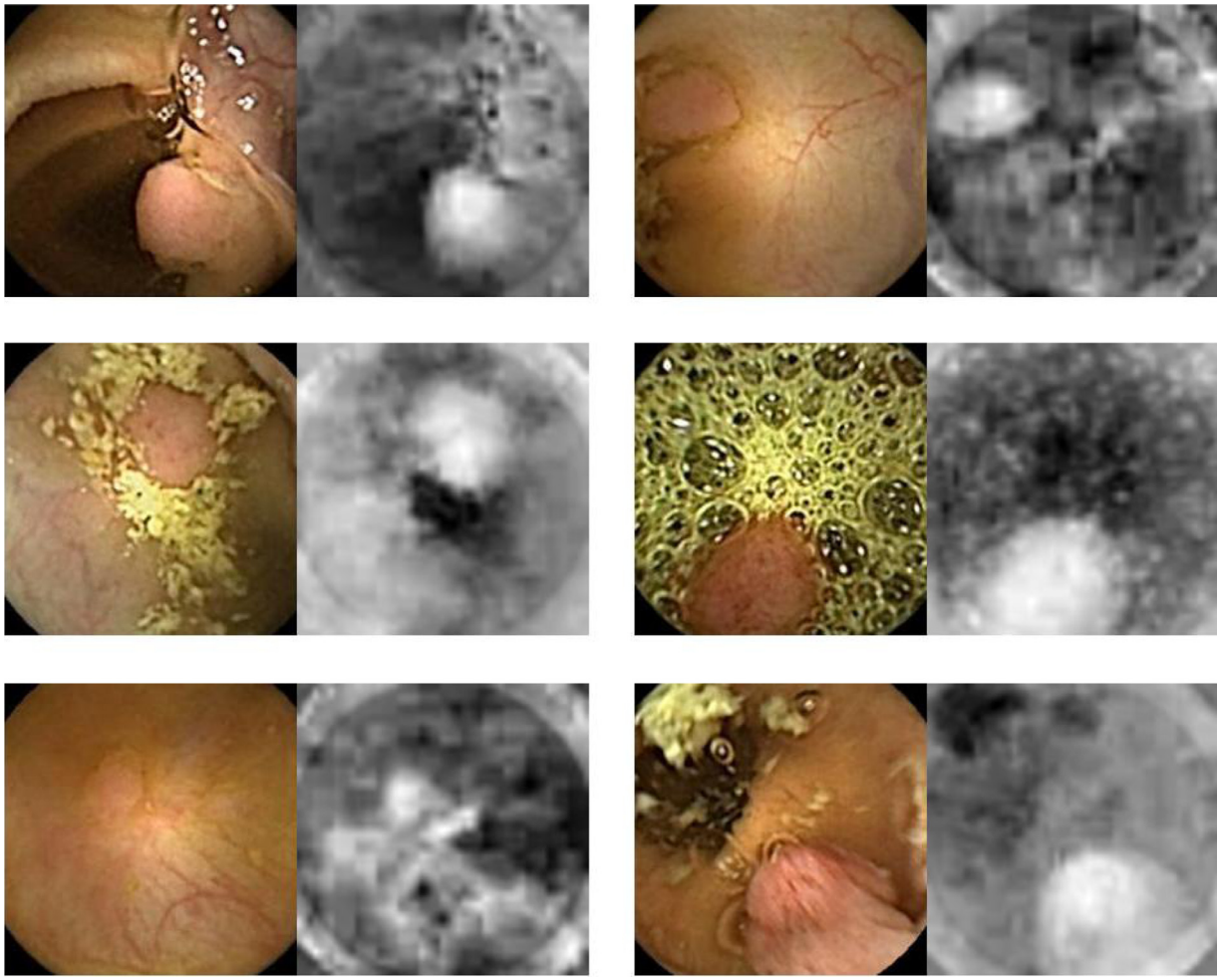

Fig. 3. Saliency map of the network for some polyps, where brighter regions indicate higher contribution.

(Fig 5: top-left and bottom-right). This would be of less concern since small polyps, i.e. those smaller than 6mm in size are less likely to harbor advanced neoplasia and are often missed by OC [24] and [25]. The second group of missed polyps shared the feature of being so large (larger than 10 mm) in the image, that they caused perspective distortion (Fig 5: bottom-left).

Although this type II error could raise concerns over the performance of the network in detecting very large polyps, these polyps were also detected in previous frames where they looked smaller, which mitigate the consequences of the obstruction of cameras field of view.

We further tested the performance of our CNN on 3 full CCE videos (each featuring approximately 50,000 images) that were not seen by the network during training, validation or testing phase. Each video contained at least one polyp which were successfully picked by our CNN. However, we observed that the network, in multiple occasions, classified the normal mucosa as colorectal polyps (specificity of 86% over the 3 videos), which could be due to the specificity measure of the detection network (96.3%).

3.2. Interpretability, saliency maps and Faster R-CNN

To visualize the saliency maps of our network and interpret its outcome, we computed the gradient of the image with respect to the unnormalized class scores. Fig 3 represents the activation map in which regions with larger gradient (brighter in the image), indicate higher contribution to the prediction. The network learnt to attend to the edges of the polyps, which is a high-signal criterion for diagnosis, while also paying attention to the context. The performance of our Faster R-CNN network in terms of accuracy, sensitivity and specificity over 170 annotated test images not previously seen by the network was 94.6%, 95.3% and 92.8%, respectively. Examples of detected and localized polyps using our developed Faster R-CNN are presented in Fig 4.

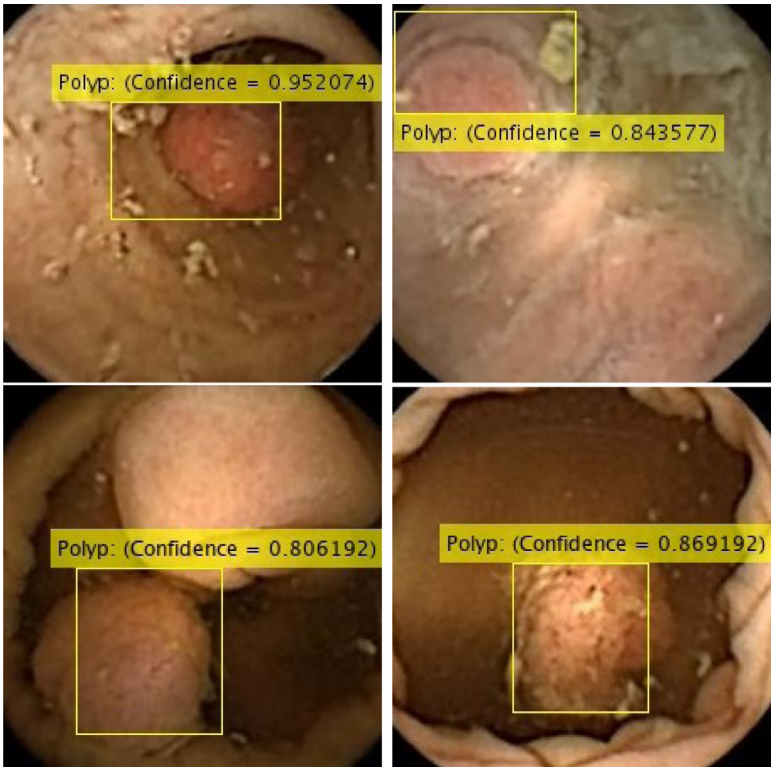

Fig. 4. Colorectal polyp localization using Faster R-CNN.

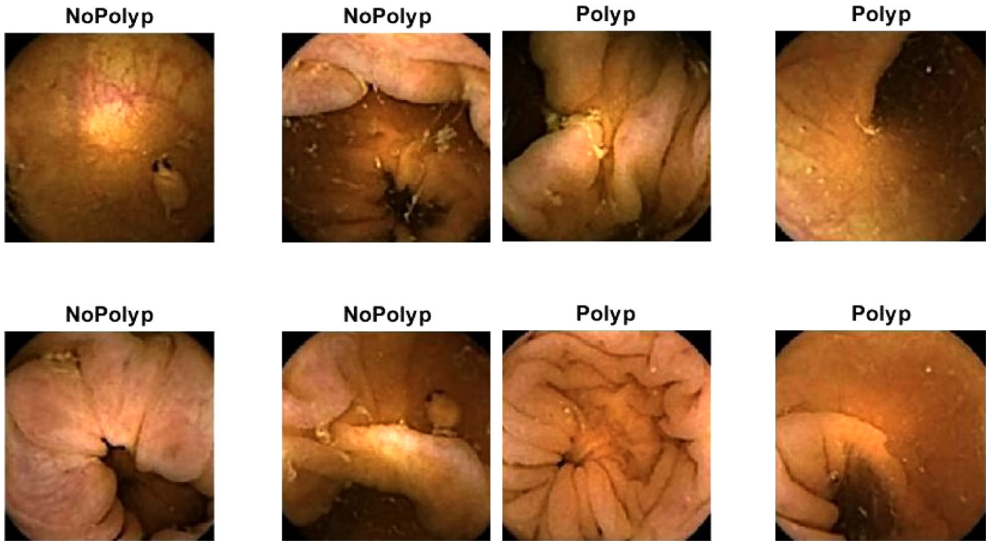

Fig. 5. Misclassification examples, left: false negatives, right: false positives.

4. Conclusions and future works

In this study, we proposed a novel and an autonomous end-to-end deep learning model to detect colorectal polyps of any size and morphology in CCE videos. We showed that our network architecture benefitting from the combination of data augmentation and transfer learning, overcame the ambiguity and lack of general task-independent rules for optimal feature selection common to medical computer vision tasks. Consequently, our approach outperformed all other state-of-the-art results in polyp detection by a wide margin.

Additionally, our method gave more interpretability to network predictions by studying saliency maps where the contribution of polyps in terms of activation was significant. The saliency map could assist the medical staff to interpret and detect the regions of the image, which due to a high contribution, have been marked as polyp. This is especially of high importance when an image is misclassified (type I or type II error). The interpretability enables a smooth adoption of this method in clinical settings, paving the road for the deployment of this network. This is because currently, highly experienced nurses manually perform detailed pre-reading and analysis of CCE images. This process is tedious, as up to 50,000 CCE retrieved images of each patient need to be thoroughly investigated. Our detection, localization and characterization algorithm mostly automatize this process and therefore significantly reduces the burden on medical staff.

Future work includes exploring architectures such as Capsule Networks based on ensemble learning, and integration of attention mechanisms, which are more difficult to train, but could provide even more concrete interpretability. Our goal is to use a CNN to encode the image, characterize the size, morphology and location of the polyps, and a Recurrent Neural Network with attention mechanisms to generate a description of histological features. This will be achieved by performing semantic segmentation at pixel-level using handcrafted feature-based approaches.

Declaration of Competing Interest

The authors do not have any conflict of interest to report.

Acknowledgements

This research was financially supported in part by a research grant from the University of Southern Denmark, Odense University Hospital, Danish Cancer Society and Region of Southern Denmark, through the Project EFFICACY. The Quadro P6000 GPU card used for this research was generously donated by NVIDIA Corporation.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compeleceng.2019.106531.

References

[1] Bujanda LA, Cosme A, Gil I, Arenas-Mirave JI. Malignant colorectal polyps.. World J Gastroenterol 2010;16:3103–11.
[2] Kobaek-Larsen M, Kroijer R, Dyrvig AK, Buijs MM, Steele RJC, Qvist N, et al. Back-to-back colon capsule endoscopy and optical colonoscopy in colorectal cancer screening individuals.. Colorectal Diseas 2018;20:317–30.
[3] Blanes-Vidal V, Nadimi ES, Buijs MM, Baatrup G. Capsule endoscopy vs. colonoscopy vs. histopathology in colorectal cancer screening: matched analyses of polyp size, morphology and location estimates. Int J Colorectal Diseas 2018;In Press.
[4] Surya-Prasath V.B.. Polyp detection and segmentation from video capsule endoscopy: a review. arXiv:160901915 2018.
[5] Pogorelov K, Ostroukhova O, Jeppsson M, Espeland H, Griwodz C, Lange T, et al. Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS); 2018.
[6] Yu L, Chen H, Dou Q, Qin J, Hengand PA. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos.. IEEE J Biomed Health Inform 2017;21:65–75.
[7] Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy.. Gastroenterology 2018;In Press.
[8] Puig I, Kaltenbach T. Optical diagnosis for colorectal polyps: a useful technique now or in the future? Gut Liver 2018;12:385–92.
[9] Mohammed A., Yildirim S., Farup A., Pedersen M., Hovde O.. Y-Net: a deep convolutional neural network for polyp detection. arXiv arXiv:1806:01907 2018.
[10] Kodogiannis V, Boulougoura M. An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging.. Int J Inf Technol 2007;13:46–56.
[11] Li B, Fan Y, Q MM, Qi L. Intestinal polyp recognition in capsule endoscopy images using color and shape features.. In: IEEE International Conference in Robotics and Biomimetics (ROBIO); 2009. p. 1490–4.
[12] Jia J, Sun S, Terrence T, Wang P. Accurate and efficient polyp detection in wireless capsule endoscopy images.. US Patent 2014;14/471:143.
[13] Nadimi ES, Herp J, M BM, Blanes-Vidal V. Texture classification from single uncalibrated images: random matrix theory approach. 27th IEEE International Workshop on Machine Learning for Signal Processing (MLSP); 2017. doi:10.1109/MLSP.2017.8168115.
[14] Li B, Meng MQH. Automatic polyp detection for wireless capsule endoscopy images.. Expert Syst Appl 2012;39:10952–8.
[15] LeCun Y, Bottou L, Benigo Y, Haffner P. Gradient-based learning applied to document recognition.. Proc IEEE 1998;86:2278–324.
[16] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks.. In NIPS 2012.
[17] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions.. In: In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 1–9.
[18] He K., Zhang X., Ren S., Sun J.. Deep residual learning for image recognition. arXiv:151203385v1 2018.
[19] Simonyan K., Zisserman A.. Very deep convolutional networks for large-scale image recognition. arXiv:14091556v6 2018.
[20] Zeiler M, Fergus R. Visulization and understanding convolutional networks.. ICCV 2012.
[21] Zitnick CL, Dollar P. Edge boxes: locating object proposals from edges. Eur Conf Comput Vision 2014:391–405.
[22] Uijlings JRR, van de Sande KEA, Gevers T, Smeulders AWM. Selective search for object recognition.. Int J Comput Vision 2013;104:154–71.
[23] Shaoqing R, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 2015;28.
[24] Heresbach D, Barrioz T, Lapalus MG, Coumaros D, Bauret P, Potier P, et al. Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies.. Endoscopy 2008;40:284–90.
[25] Murino A, Hassan C, Repici A. The diminutive colon polyp: biopsy, snare, leave alone? Curr Opin Gastroenterol 2016;32:38–43.

Esmaeil S. Nadimi is a Professor of machine learning (ML) and artificial intelligence (AI) and the leader of the group of clinical machine learning at the University of Southern Denmark. His research interests mainly include applications of ML and AI in health, for diagnostics and personalized treatments.

Maria M. Buijs is a Ph.D. student and a medical doctor of Gastroenterology at center for Advanced Adenomas and Early Colorectal Cancers (SATC center) in Svendborg hospital. Her research interests mainly include diagnostics of colorectal diseases.

Jurgen Herp is an Assistant Professor of machine learning (ML) and artificial intelligence (AI) at the group of clinical machine learning at the University of Southern Denmark. His research interests mainly include applications of ML and AI in health, for diagnostics and personalized treatments.

Rasmus Kroijer is a Ph.D. student and a medical doctor of Gastroenterology at center for Advanced Adenomas and Early Colorectal Cancers (SATC center) in Svendborg hospital. His research interests mainly include diagnostics of colorectal diseases.

Morten Kobaek-Larsen is an Associate Professor of Gastroenterology at the center for Advanced Adenomas and Early Colorectal Cancers (SATC center) in Svendborg hospital. His research interests mainly include diagnostics of colorectal diseases.

Emilie Nielsen is a specialist advisor for clinical development of innovative solutions in health care at the Center for Innovative Medical Technologies (CIMT). Her research interests mainly include development of innovative solutions for diagnosis of colorectal cancer.

Claus D. Pedersen is a chief consultant and the leader of the group of innovative solutions in health care at the Center for Innovative Medical Technologies (CIMT). His research interests mainly include development of innovative solutions for diagnosis of colorectal cancer.

Victoria Blanes-Vidal is an Associate Professor of machine learning (ML) and artificial intelligence (AI) at the group of clinical machine learning at the University of Southern Denmark. Her research interests mainly include applications of ML and AI in health, for diagnostics and personalized treatments.

Gunnar Baatrup is a Professor of Gastroenterology and the leader of the center for Advanced Adenomas and Early Colorectal Cancers (SATC center) in Svendborg hospital. His research interests mainly include diagnostics of colorectal diseases.