Pascal Zobel, Magnus Rathke, Steffen Mühldorfer, Thomas Wittenberg*

Computer Aided Detection of Polyps in White-light-Colonoscopy Images using Deep Neural Networks

Abstract: Early detection of polyps is one central goal of colonoscopic screening programs. To support gastroenterologists during this examination process, deep convolutional neural network can be applied for computer-assisted detection of neoplastic lesions. In this work, a Mask R-CNN architecture was applied. For training and testing, three independent colonoscopy data sets were used, including 2484 HD labelled images with polyps from our clinic, as well as two public image data sets from the MICCAI 2015 polyp detection challenge, consisting of 612 SD and 194 HD labelled images with polyps. After training the deep neural network, best results for the three test data sets were achieved in the range of $recall = 0.92$, $precision = 0.86$, $F_1 = 0.89$ (data set A), $rec = 0.86$, $prec = 0.80$, $F_1 = 0.82$ (data set B) and $rec = 0.83$, $prec = 0.74$, $F_1 = 0.79$ (data set C).

Keywords: polyp detection, convolutional deep neural networks, colonoscopy

<https://doi.org/10.1515/cdbme-2019-0059>

1 Introduction

Early detection and removal of neoplastic lesions in the form of polyps is one central goal of colonoscopic screening programs. Screening colonoscopy is currently the “gold-standard” for the examination of the colon. Using a video endoscope, the surface and tissue of the colon is examined for conspicuous regions. In order to support gastroenterologists during this examination process, image-processing methods can be applied for computer-assisted detection of neoplastic lesions. In the past 20 years, many approaches from image processing have been proposed to support this task, e.g. texture and feature-based methods using various classification schemes. Recently,

new image analysis approaches have been developed, based on deep convolutional neural networks (DCNNs).

1.1 State of the Art

In the past years, various research groups have proposed different approaches for the detection of polyps in colonoscopic images and videos. These range from traditional feature-based methods in combination with a classifier as e.g. support vector machines (SVN) or k-nearest neighbor (kNN) based on small image data collections, to the use of convolutional deep neural networks with larger data sets. To the first group belong e.g. Hwang et al. [1] who make use of a geometrical approach with parametrical elliptic forms for polyp detection, evaluated on a small data collection of 27 images. 26 polyps were correctly detected, one was overseen, 5 were detected false-positively. Magoulas et al. [2] evaluate the parameterization of a hybrid neural network for the differentiation of square image regions with and without polyps. For training 1,200 and evaluation 16,876 regions were used. Achieved classification rates depend on the selected network topology and are in the range of +/-90%. Park et al. [3] proposed a multi-step variant for polyp detection of polyps in colposcopy sequences. After elimination of non-informative frames, edge information is extracted to find polyp candidates on ellipse fitting. Candidates are classified using conditional-random-field classifier based on „Eigen-Tissues” of the detected candidate regions. For training 364 sequences of 35 colonoscopy recordings are used. Total classification rate achieved on a large-scale dataset using leaving-out-one-patient cross-validation yields an area under the ROC curve of 0.89. Nowack et al. [4] applied color texture approaches together with SVNs and kNNs for the differentiation of polyps and normal tissue on 2,484 colonoscopic hand labeled images and reached a classification rate of 91%. When these images were automatically split up in regions using a super-pixel (SLIC) approach, the classification rate based on approx. 25,000 SLIC patches dropped to 81%. An extended overview of *traditional polyp detection* methods can be found in [5]. Tajbakhsh et al. [6] propose to extract edge, structure

***Corresponding author: Thomas Wittenberg:** Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany, e-mail: thomas.wittenberg@fraunhofer.de

Pascal Zobel, Magnus Rathke: Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, **Steffen Mühldorfer,** Clinics Bayreuth

and surrounding context information, which is then systematically thinned out using a voting mechanism in order to differentiate between polyps and background. On a data collection of 300 static images (CVC-ColonDB) this approach yields a sensitivity of 88%.

Furthermore, with the emerge of deep convolutional neural networks techniques several new methods for the detection of polyps have been proposed. These can be categorized into so-called “single-shot” detectors and “region-based” detectors. Among the first group, Urban et al. [7] used a variant of the “you only look once” (YOLO) technique and successfully trained a network with 8,641 images. Additionally, Shin et al. [8] were one of the first groups applying region based CNNs for polyp detection. They used a Faster R-CNN architecture and a transfer learning approach to train on 612 images from the CVC-ClinicDB [9] database (data set B, see below). Based on the Faster R-CNN architecture, He et al. [10] presented an instance segmentation framework, Mask R-CNN, predicting object masks. Mo et al. [11] also apply this approach of Faster R-CNNs [7] for polyp detection. Other groups [12,13] suggest other DCNN architectures, including a combination of Wavelets and deep learning [12], or SegNet [13]. In contrast to the fore mentioned classical approaches, the deep learning methods rely on much more training data with up to 100,000 images from 2,000 patients, yielding results up to 98% [11].

Within this work, the approach is to evaluate what accuracy can be achieved using DCNNs i.e. a Mask R-CNN architecture combined with transfer learning on colonoscopy images. We use our own training set (data set A, see below) to retrain the classification layers of the Mask R-CNN and apply the retrained DCNN on two freely available colonoscopic image data sets with polyps [9] as additional test cases.

2 Materials

As image data for training and validation of a DCNN, 2,484 still HD images from routine colonoscopy examinations (Olympus Q180 colonoscope, Exera II processor) were retrospectively extracted from video sequences and manually labeled. In this data set, in total 2513 polyps were delineated (*data set A*, “Bayreuth DB”) (see Table 1). This data set was split up in three disjunct subsets for training (70%, 1790 images), validation (15%, 350 images) and testing (15%, 344 images), see Figure 1, left column. All images in the three subsets were recorded on different days.

Additionally, from the MICCAI 2015 polyp detection challenge [9] two additional data sets were used for testing. *Data set B* (“CVC-ClinicDB”) incorporates 612 SD (388 x 284 pixels) images (obtained with a Pentax 90i colonoscope, EPKi

7000 processor). *Data set C* (ETIS-LaribPolypDB) includes 196 HD (1225 x 966 pixels) colonoscopy frames (obtained with Olympus Q160AL, Q165 colonoscopes, Exera II processor).

Table 1: Overview of data collections used

Data Set	Images	Resolution
A (“Bayreuth” DB)	2484	640 x 504 HD
B (CVC-ClinicDB)	612	388 x 284 SD
C (ETIS-LaribPolypDB)	194	1225 x 966 HD

3 Methods

For the automated detection of polyps in colonoscopy white light images, a Mask R-CNN architecture (a convolutional deep neural network for instance segmentation) [10] with a ResNet-101 [14] backbone for feature extraction was used. As code base the Mask R-CNN implementation of Matterport was used and extended. As the available data sets (Section 2) were relatively small to train the complex network architecture, a transfer learning approach was employed. Hence, the network was pre-trained on the MS-COCO image data collection with over 200,000 images and 91 classes from street-life [15]. The DCNN was then forced to a two-class problem, consisting of foreground (“polyp”) and background (“non-polyp”) classes.

The training was done in different stages, starting from re-training only the network heads to fine-tuning all layers. For training four NVIDIA Tesla P100 GPUs were used. As the available training data (from data set A) is quite small, image augmentation such as image flipping and rotation was used during the training phase.

For evaluation, the network’s sensitivity can be adjusted using a threshold to further process region of interests with lower objectiveness score. To be able to compare our results with the results of the MICCAI 2015 polyp detection challenge [11], we used the same frame-based evaluation metric. Therefore, we calculated *Recall* (*rec*), *Precision* (*prec*) and the *F₁*-score (*F₁*) for all experiments.

Polyps were rated as properly classified if the *Intersection over Union* (IoU) between the estimated mask and the ground truth mask exceeds 30%.

4 Results

In the top four rows of Table 2 the results obtained from the validation subset of data set A with varying objectiveness thresholds are presented.

The best detection rate $rec = 0.92$ relates to a precision of 0.80, a F_1 score of 0.86 and a threshold of 0.3, while the highest values for precision $prec = 0.87$ relates to a recall value of $rec = 0.90$, a F_1 score of 0.88 and a threshold of 0.8.

As the objectiveness threshold of 0.3 indicates a good tradeoff between a low rate of missed polyps (FN) and a reasonable false-alarm rate (FP), we used this threshold for evaluation on the additional test data sets.

Table 2: Results with various objectiveness thresholds obtained from the *validation* subset (top four rows) and on the *test subset* (row 5) of *data set A*. The bottom two rows show reference measurements from [4] using kNN and SVM approaches on automatically computed SLIC patches on the same data set.

Subset	Thresh	TP	FN	FP	Rec	Prec	F_1
Val.	0.8	333	37	48	0.90	0.87	0.88
Val.	0.5	336	32	68	0.91	0.83	0.87
Val.	0.3	337	29	82	0.92	0.80	0.86
Val.	0.1	337	27	118	0.93	0.74	0.82
Test	0.3	326	26	52	0.93	0.86	0.89
Ref, kNN					0.74	0.83	0.79
Ref, SVN					0.82	0.84	0.83

Table 3: Results using the trained DCNN on data set B and C from the MICCAI 2015 polyp detection challenge (rows 1-2). The bottom row provides the challenge's best result (CUMED) on the ETIS data set for comparison [9] and recent findings by [17].

Subset	Thresh	TP	FN	FP	Rec	Prec	F_1
CVC	0.3	537	89	132	0.86	0.80	0.82
ETIS	0.3	170	34	59	0.83	0.74	0.79
MICCAI-ETIS		110	98	226	0.69	0.72	0.70
Kang – ETIS					0.743	0.738	0.667

Results on the corresponding *test data set A* were $rec = 0.92$, $prec = 0.86$, $F_1 = 0.89$ (see Table 2, row 5). Furthermore, at the bottom of Table 2, reference results from previous experiments on the same data set A using texture analysis and a kNN classifier [4] are provided. Example images of detect polyps from data set A can be found in Figure 1, center column. On test set B (CVC) the best results with a threshold of 0.3 are $rec = 0.86$, $prec = 0.80$, and $F_1 = 0.82$. On test set C (ETIS) best

results of $rec = 0.83$, $prec = 0.74$, and $F_1 = 0.67$ were obtained. Compared to the 2017 published values of the MICCAI challenge's winner [9], our results are considerably better on data set C.

Processing time per frame on a system with NVIDIA Tesla P100 GPU is at 500ms.

5 Discussion

In comparison to our previous published results on *test set A* [4] using a combination of SLIC super-pixels for image segmentation, color-texture features for SLIC-patch characterization and a k-nearest-neighbor ($rec = 0.74$, $prec = 0.84$, $F_1 = 0.78$) as well as a support vector machine classifier ($rec = 0.82$, $prec = 0.83$, $F_1 = 0.83$), the results achieved with the proposed DCNN approach ($rec = 0.93$, $prec = 0.86$, $F_1 = 0.89$) already improve our earlier work, see Table 2.

It can be seen, that the applied DCNN model is able to perform accurate detection and segmentation on completely independent test data sets of the MICCAI polyp detection challenge. We were able to outperform the challenges winner [9] on data set C ($rec = 0.69$, $prec = 0.72$, $F_1 = 0.71$) with our approach (trained on data set A) ($rec = 0.83$, $prec = 0.74$, $F_1 = 0.71$) as well as recent work by Kang [17], see Table 3.

However, our training data set of 1,790 images is rather small for training a Mask R-CNN with a ResNet-101 backbone. Experimenting more on extended data augmentation in order to increase the amount of training data may improve the results. Shin et al. [16] experimented with advanced data augmentation techniques for polyp object detection on the ETIS data set and considerably improved their results.

Comparing our work to the results of other groups we can observe the following: Shin et al. [8] obtained results on data set C (ETIS) with a precision of $prec = 0.91$ ($rec = 0.71$, $F_1 = 0.80$) which has been trained on data set (CVC) using Faster R-CNNs, yielding a detection of polyps, but no segmentation. With over 100,000 training images, Mo et al. [11] achieved an F_1 score of 0.98 for polyp detection. Wang et al. [13] achieved a $recall = 0.88$ ($prec = 0.93$, $F_1 = 0.90$) using 5,545 (3,634 with polyps) images from 1,290 patients for training. Hence, it can clearly be seen that using more data for training will improve the detection rate.

6 Conclusion

For real clinical value, still too many false positive areas are detected and some false negatives are missed (see Fig. 1

bottom row). One explanation for the currently achieved values is the small training data set, as it is known from literature [11,12, 13] that more and diverse training data in the range of 100,000 [11] images and more will provide better results. Thus, we are currently extending our training data set in order to achieve better results.

Moreover, for clinical usage real-time processing is strongly desired. A step in this direction would be pruning the heavy ResNet-101 backbone to a smaller architecture. This would significantly reduce the computation time.

Figure 1: Examples of automatic detected polyps from data set A. Left column: original data; right column: manually annotated ground truth; center column: detected polyps. From top to bottom: elevated polyp, small polyp, two small polyps (one not annotated in ground truth), flat polyp, false-negative polyp (not detected).

This publication has partially been funded by the Federal Ministry of Education and Research under the project reference numbers 16FMD01K, 16FMD02 and 16FMD03.

References

- [1] Hwang et al. Polyp detection in colonoscopy video using elliptical shape feature. *Proc's ICIP 2007*, pp. 465-468.
- [2] Magoulas et al. Neural network-based colonoscopic diagnosis using on-line learning and differential evolution. *Applied Soft Computing* 4(4) 2004, pp. 369–379.
- [3] Park et al. A colon video analysis framework for polyp detection. *Trans. Biomed. Eng.* 59(5) 2012, pp. 1408-1418.
- [4] Nowack et al: Unterscheidung von Polypen und Hintergrundgewebe in statischen Koloskopieaufnahmen. In Hahn et al. (Eds.): *Proc's CURAC 2015*, pp.201-204
- [5] Ameling et al. Detection of lesions in colonoscopic images: A review. In *Proc's World Cong. Med. Physics & Biomed. Eng.* 25 (IV) 2009 pp. 995–998
- [6] Tajbakhsh et al. Automated polyp detection in colonoscopy videos using shape and context information. *Trans Medical Imaging* 35(2), 2016. pp. 630-643.
- [7] Urban et al: Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018;155:1069-78
- [8] Shin et al: Automatic colon polyp detection using region based deep CNN and post learning approaches. 2018(6), *IEEE Access* 40950—40962. 10.1109/ACCESS.2018.2856402
- [9] Bernal J et al. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. *Trans Medical Imaging* 36(6), 2017: 1231-1249
- [10] He K et al: Mask R-CNN (2011). Online <http://arxiv.org/pdf/1703.06870v3>.
- [11] Mo X et: An efficient approach for polyps detection in endoscopic videos based on Faster R-CNN (2018). *arXiv:1809.01263*
- [12] Billah et al: An automatic gastrointestinal polyp detection system in video endoscopy using fusion of color wavelet and convolutional neural network features. *Int. J. Biomed. Imaging* 2017, Article ID 9545920.
- [13] Wang et al: Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2018: doi:10.1136/gutjnl-2018-317500
- [14] Waleed, A (2017): Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow, https://github.com/matterport/Mask_RCNN
- [15] Lin et al.: Microsoft COCO: Common Objects in Context in ECCV, 2014
- [16] Shin et al: Abnormal colon polyp image synthesis using conditional adversarial networks for improved detection performance. *IEEE Access* 2018. DOI:10.1109/ACCESS.2018.2872717
- [17] Kang & Gwak: Ensemble of Instance Segmentation Models for Polyp Segmentation in Colonoscopy Images. *IEEE Access* March 2019. DOI: 10.1109/ACCESS.2019.2900672