ONE-STAGE FIVE-CLASS POLYP DETECTION AND CLASSIFICATION

Yu Tian[†] *Leonardo Z.C.T. Pu*[‡] *Rajvinder Singh*[‡] *Alastair D. Burt*[‡] *Gustavo Carneiro*[†*]

[†] Australian Institute for Machine Learning, School of Computer Science, University of Adelaide
[‡]Faculty of Health and Medical Sciences, University of Adelaide

ABSTRACT

The detection and classification of anatomies from medical images has traditionally been developed in a two-stage process, where the first stage detects the regions of interest (ROIs), while the second stage classifies the detected ROIs. Recent developments from the computer vision community allowed the unification of these two stages into a single detection and classification model that is trained in an end to end fashion. This allows for a simpler and faster training and inference procedures because only one model (instead of the two models needed for the two-stage approach) is required. In this paper, we adapt a recently proposed one-stage detection and classification approach for the new 5-class polyp classification problem. We show that this one-stage approach is not only competitive in terms of detection and classification accuracy with respect to the two-stage approach, but it is also substantially faster for training and testing. We also show that the one-stage approach produces competitive detection results compared to the state of the art results on the MICCAI 2015 polyp detection challenge.

***Index Terms*—** Deep learning, one-stage polyp detection and classification

1. INTRODUCTION

Colorectal cancer is considered as one the most harmful cancers – current research suggests that it is the third largest cause of cancer deaths [1]. The early detection of colorectal cancer can be performed with the colonoscopy procedure for at-risk patients with symptoms like hemotochezia and anemia [2]. Colonoscopy is based on the navigation of a tiny camera into the colon in order to detect, classify and possibly remove or sample polyps, which are considered as the precursors of colon cancer [1]. The accurate detection and classification of colon polyps may improve the 5-year survival rate to over 90% [1]. Polyps can be classified into five classes, representing a range from benign to malignant [1], and such classification is imperative to determine the action to be taken by the medical practitioner during the colonoscopy procedure. Unfortunately, the accuracy of such manual classification varies substantially, leading to potentially wrong actions that have different consequences for the patient [3]. For instance, the mis-interpretation of a polyp class can lead to unnecessary endoscopic resection, which can be dangerous and costly [4]. Therefore, automated polyp detection and classification can have an important role in assisting doctors during a colonoscopy exam.

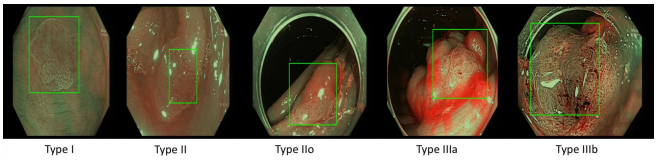

Fig. 1: Annotation of the five classes of polyps.

Such methods have traditionally been developed with a two-stage process [5], where a first stage detects one or multiple regions of interests (ROIs) that will then be classified by a second stage. Even though such division seems obvious, it introduces a few issues: 1) two models need to be trained (one for detection and another for classification), introducing un-necessary complexity given the inter-dependence between these stages; 2) the training and inference processes rely on two models that can be computationally expensive to run; and 3) the features learned for detection are not necessarily well-adapted for classification, so the detected ROIs may not be ideal for the classification process. The unification of these two stages into a one-stage detection and classification system can solve the issues above, leading to more efficient (i.e., faster) training and inference processes and potentially more accurate models given that the features learned for detection are also trained for classification. Such one-stage approaches have been studied by the computer vision community, e.g., YOLO [6], Retinanet [7] and Faster RNN [8].

In this paper, we adapt Retinanet [7] for the problem of five-class polyp detection and classification from colonoscopy images. We show that such one-stage method is not only more efficient (i.e., it shows smaller training and inference times than the two-stage approach), but it is also as accurate as state-of-the-art two-stage approaches in terms of detection and classification. These results are demonstrated on a new 5-class polyp classification, using a dataset containing 871 high-quality images of colorectal polyps. The polyps were annotated by a professional medical practitioner from the Faculty of Health and Medical Sciences of the University of Adelaide – the annotation is

*This work was partially supported by the Australian Research Council project (DP180103232) and the 2018 Northern Adelaide Local Health Network funding.

represented by a bounding box indicating the polyp location and and a 5-class label (Figure 1): hyperplastic polyp ($TypeI$), sessile serrated adenomas/polyp ($TypeIIo$), low grade adenoma/tubular adenoma ($TypeII$), high grade adenoma/tubulovillous adenoma/superficial cancer ($TypeIIIa$) and invasive cancer ($TypeIIIb$). Pu et al. [3] developed a classification system for such 5-class polyp problem, which was considered to be more effective than the 2-class [9] and 3-class [10] approaches. However, this method relies on a manual detection approach. Our experiments show that our proposed one-stage produces comparable detection and classification results when compared to the two-stage approach, while being substantially faster both in the training and inference stages. When we compare polyp detection results alone, our one-stage approach can also achieve competitive results compared to the state of the art from the 2015 MICCAI polyp detection challenge [2]. Finally, we also compare the classification results between our one-stage approach and the classifier that relies on manual detections [3], which shows that the difference in performance is mainly due to incorrect detections.

2. LITERATURE REVIEW

Currently, some of the dominant paradigms in object detection based on deep learning can achieve impressive detection accuracy. Two-stage detectors represent the mainstream approach, achieving the current state-of-the-art results (e.g., Fast RCNN [8] and Faster RCNN [11]). Such methods use a Region Proposal Network (RPN) to detect the object on an image without considering the class of the object and then apply a classifier on the ROI. Up until the work by Lin et al. [7], such two-stage approach was considered more accurate, but less efficient than a one-stage detection and classification approach. Lin et al. [7] argue that such discrepancy is due to the imbalance between positive and negative training samples, which is addressed with the use of the focal loss objective function. This new loss function led to the development of Retinanet [7], which shows competitive detection and classification results, compared to the aforementioned two-stage approach, while being more efficient in terms of training and testing.

A major reference for the problem of polyp detection is the MICCAI 2015 Challenge [2]. In that challenge, the most competitive methods (CUMED and OUS) are based on deep learning models. We show in the experiments that the detection stage of our proposed one-stage model has competitive results compared to these top-performing teams, even though our approach performs detection and classification, while the top methods from the challenge focus only on the detection task. Regarding polyp classification, state-of-art methods explore 2-class [9] and 3-class [10] problems. The only method that explores a 5-class problem relies on manual polyp detection [3]. In the experiments, we show that our one-stage method produces competitive results with respect to [3], when we consider only the true positive detections.

Finally, one-stage detection and classification has been explored in medical image analysis for mammograms [12], [13]. Also, Mugahed et al. [13] proposed a fully integrated 3-stage approach involving detection, segmentation and 2-class classification using YOLO [6], FrCN [13] and AlexNet [14]. However, such one-stage approach has not been applied to polyp classification from colonoscopy images. In addition, our approach is the first to be tested in a detection and classification problem with more than three classes.

3. DATASET AND METHODS

3.1. Dataset

The dataset used in this work is defined by $\mathcal{D} = \{\mathbf{x}_i, d_i, y_i, \mathbf{b}_i\}_{i=1}^{|\mathcal{D}|}$, where $\mathbf{x} : \Omega \rightarrow \mathbb{R}^3$ denotes a colonoscopy image (Ω represents the image lattice), $d_i \in \mathbb{N}$ represents patient identification [1], $y_i \in \mathcal{Y} = \{I, II, IIo, IIIa, IIIb\}$ denotes the five polyp classes, and $\mathbf{b}_i \in \mathbb{R}^4$ denotes the two 2-D coordinates of the bounding box containing the polyp. These images of colorectal polyps were obtained with the Olympus ®190 dual focus colonoscope. The distribution of this dataset is as follows: 1) Type I: 102 images (39 patients); 2) Type II: 346 images (93 patients); 3) Type IIo: 281 images (48 patients); 4) Type IIIa: 79 images (25 patients); and 5) Type IIIb: 63 images (14 patients). In total, we have 871 images (218 patients).

3.2. Methods

In this section, we present the details of the one-stage and two-stage approaches – both are based on Retinanet [7] using the ResNet-50 [15] as the underlying classifier.

Two-stage Approach: For the two-stage approach, polyp detection is achieved by predicting the bounding boxes around the tissue, where we use Resnet50 [15] as the base model for Retinanet [7]. Retinanet applies the focal loss function [7] to address the imbalance problem between foreground and background samples during training, where the focal loss is defined by $FL(p_i) = -(1-p_i)^{\gamma} log(p_i)$, where $p_i = p$ if $y_i = 1$ ($p_i = 1 - p$, otherwise), with $p \in [0, 1]$ being the model's estimated probability for the class with label $y_i = 1$ (in this context, $y_i = 1$ represents a bounding box containing a polyp); and $\gamma \in [0, 5]$ reduces the loss for well-classified examples ($p_i > 0.5$). In this two-stage approach, the Retinanet only differentiates between foreground (a bounding box containing any type of polyp) and background (normal tissue). During inference and training, the detection stage outputs bounding boxes that show a confidence score above $\tau \in \{0.05, 0.5\}$, which are merged with non-maximum suppression, i.e., boxes are merged if they have an intersection over union (IoU) above 0.5. The detected bounding boxes containing polyps are then classified with Resnet-50 [15] for the 5-class polyp classification problem. During training, we only consider the detected bounding boxes that have an IoU> 0.5 with the ground truth (the remaining samples are disregarded for training).

[1]Note that the dataset has been de-identified – d_i is useful only for splitting $\mathcal{D}$ into training, testing and validation sets ni a patient-wise manner.

Table I: Training and inference running times of the one- and two-stage approaches.

method	training time	inference time
1-stage	13hrs-14hrs	0.067s per image
2-stage	23hrs-24hrs	0.221s per image

One-stage Approach: A more efficient alternative to the two-stage approach detailed above is a method that can detect and classify the polyps simultaneously. The approach consists of a single-model that can detect and classify polyps – this model is trained in an end-to-end fashion [7]. During inference, the bounding boxes with confidence scores larger than a threshold τ are merged with non-maximum suppression – specifically, we merge all bounding boxes that have the same class with an IoU>0.5, and the confidence of the merged bounding boxes is the maximum classification confidence of the merged bounding boxes.

4. EXPERIMENTAL RESULTS

In our experiments, we aim to show that the one-stage approach produces competitive detection and classification accuracy, when compared to the two-stage approach, while being faster to train and to test. In addition, we also aim to demonstrate that our proposed one-stage approach produces competitive detection results when compared to the current state of the art methods for polyp detection. The detection and classification experiments are based on a 5-fold cross validation experiment, using the dataset detailed in Sec. 3.1, where the training set $\mathcal{T} \subset \mathcal{D}$ contains images from 60% of the patients, the validation set $\mathcal{V} \subset \mathcal{D}$ has images of 20% of patients and the test set $\mathcal{U} \subset \mathcal{D}$ contains images of the remaining 20% of the patients, where $\mathcal{T} \bigcap \mathcal{V} \bigcap \mathcal{U} = \emptyset$. All experiments are carried out on a desktop computer with Intel i7-8700k processor, 16GB of DDR4 RAM and 11GB Nvidia GTX 1080Ti. The Resnet-50 is pre-trained on Imagenet [16]. During training, we use data augmentation (small rotations, translations, shears, scaling and random flipping), increasing the training set by six fold. Results are shown in terms of mean average precision (mAP) for the detection results, and accuracy and area under the ROC curve (AUC) for classification. The comparison with the state of the art is based on the measures and results from the MICCAI 2015 polyp detection challenge [2].

4.1. Training and Inference Running Time

Table I displays the training and testing running times for both the one- and two-stage methods, which clearly shows that the one-stage approach is more efficient.

4.2. Detection and Classification Results

Figure 2 shows the detection results for the one- and two-stage methods. The results suggest that the 2-class detection used in the two-stage method achieves comparable mAP to the 5-class detection used by the one-stage system. We also show the detection results of our one-stage approach on the 2015 MICCAI polyp detection challenge [2] in Table II – these results show that even though implemented for detecting and classifying polyps (into five classes), our method is competitive with the state of the art, which are all designed specifically for detecting polyps.

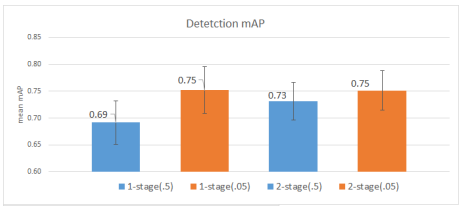

Fig. 2: Comparison between the one- and two-stage detectors with score threshold $\tau \in \{0.05, 0.5\}$ (shown in parenthesis) in terms of mean and standard deviation of the mAP over the 5-fold cross validation experiment.

Table II: Comparison between our one-stage detector and the state of the art from the MICCAI 2015 Polyp detection Challenge [2].

	TP	FP	Prec	Rec	F1	F2
CUMED	144	55	72.3	**69.2**	**70.7**	**69.8**
CVC-CLINIC	102	920	10	49	16.5	27.5
ETIS-LARIB	103	1373	6.9	49.5	12.2	22.3
OUS	131	57	69.7	63	66.1	64.2
PLS	119	630	15.8	57.2	24.9	37.6
SNU	20	176	10.2	9.6	9.9	9.7
UNS-UCLAN	110	226	32.73	52.8	40.4	47.1
1-stage Detector	134	**48**	**73.6**	64.42	68.72	66.07

We also assess the classification performance by comparing the one- and two-stage approaches, using as baseline a classifier that relies on the manually detected polyps [3]. This baseline consists of a Resnet-50 classifier trained and tested on manually detected polyps, so it only handles correct detections, which means that it represents an upper-bound to the performance of the one- and two-stage approaches. Figure 3 shows a comparable performance between the one- and two-stage methods and a superior performance of the Manual method. In order to isolate the performance of the classifier in the one- and two-stage approaches, in Fig. 4 we show the classification results of the one- and two-stage approaches taking into account only the true positive detections. This makes the performance of the three methods comparable, showing that the worse performance of the one- and two-stage approaches in Fig. 3 is mainly due to the incorrect detections.

5. CONCLUSIONS

In conclusion, our work shows that one-stage detection and classification can achieve comparable performance with higher efficiency compared to two-stage approaches on the new 5-class polyp classification problem. The results also show that the gap between the one-stage approach and the manual method [3] is mainly due to the mis-detected polyps.

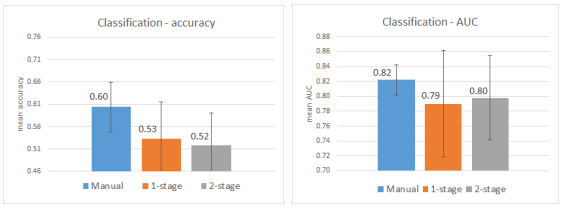

Fig. 3: Comparison between our proposed one- and two-stage methods and a classification method that uses manual detection of polyps (Manual) [3] using the mean and standard deviation of the classification accuracy (left) and AUC (right) over the 5-fold cross validation experiment.

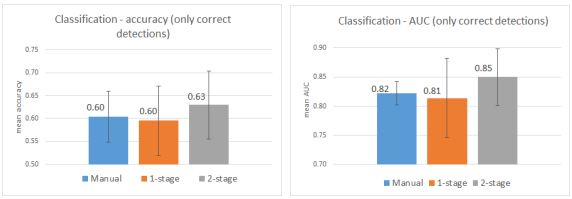

Fig. 4: Comparison between our proposed one- and two-stage methods and a classification method that uses manual detection of polyps (Manual) [3] using the mean and standard deviation of the classification accuracy (left) and AUC (right) over the 5-fold cross validation experiment. In this figure, we only consider the true positive detections to isolate the performance of the classifier.

In future work, we plan to develop methods to improve the detection approach for the one-stage approaches in order to reduce the gap with respect to the manual method.

6. REFERENCES

[1] Rebecca Siegel, Carol DeSantis, and Ahmedin Jemal, "Colorectal cancer statistics, 2014," *CA: a cancer journal for clinicians*, vol. 64, no. 2, pp. 104–117, 2014.

[2] Jorge Bernal, Nima Tajkbaksh, Francisco Javier Sánchez, et al., "Comparative validation of polyp detection methods in video colonoscopy: results from the miccai 2015 endoscopic vision challenge," *IEEE transactions on medical imaging*, vol. 36, no. 6, pp. 1231–1249, 2017.

[3] Leonardo Zorron Cheng Tao Pu, Brock Campbell, Alastair D Burt, Gustavo Carneiro, and Rajvinder Singh, "Computer-aided diagnosis for charaterising colorectal lesions: Interim results of a newly developed software," *Gastrointestinal Endoscopy*, vol. 87, no. 6, pp. AB245, 2018.

[4] Jeroen C Van Rijn, Johannes B Reitsma, Jaap Stoker, et al., "Polyp miss rate determined by tandem colonoscopy: a systematic review," *The American journal of gastroenterology*, vol. 101, no. 2, pp. 343, 2006.

[5] Mugahed A. Al-antari, Mohammed A. Al-masni, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim, "A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification," *International Journal of Medical Informatics*, vol. 117, pp. 44 – 54, 2018.

[6] Joseph Redmon and Ali Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[7] Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[8] Ross Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[9] Yoriaki Komeda, Hisashi Handa, Tomohiro Watanabe, et al., "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience," *Oncology*, vol. 93, no. Suppl. 1, pp. 30–34, 2017.

[10] Eduardo Ribeiro, Michael Häfner, Georg Wimmer, et al., "Exploring texture transfer learning for colonic polyp classification via convolutional neural networks," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. IEEE, 2017, pp. 1044–1048.

[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[12] Dezső Ribli, Anna Horváth, Zsuzsa Unger, Péter Pollner, and István Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Scientific reports*, vol. 8, no. 1, pp. 4165, 2018.

[13] Mugahed A Al-antari, Mohammed A Al-masni, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim, "A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification," *International Journal of Medical Informatics*, vol. 117, pp. 44–54, 2018.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.