Semantic Segmentation of Colorectal Polyps with DeepLab and LSTM Networks

Wei-Ting Xiao[1], Li-Jen Chang[2], and Wei-Min Liu[1], *Member, IEEE*
[1]Dept. of Computer Science and Information Engineering,
National Chung Cheng University, Chia-Yi County, Taiwan
[2]Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chia-Yi City, Taiwan

***Abstract*-- In this work we attempted to use the existing deep neural network called DeepLab_v3 to detect the polyps in colonoscopy images. Due to its large structure, the location of polyps may not be preserved and transmitted effectively. To address the issue we combined Long Short-Term Memory networks and DeepLab_v3 in parallel to augment the signal of the polyps' location. The new modification was examined with the colonoscopy image database 'CVC-ClinicDB' from MICCAI sub-challenge 2015. After training with 267 images and testing with 345 images, we got a good performance, 93.21% mean Intersection over Union (mIOU). The average computing time is 0.023 second per image. Once the model is applied to clinical colonoscopy exam videos, it could provide effective second opinions in real time to aid the diagnosis.**

I. INTRODUCTION

Colorectal cancer has high prevalence and mortality. Since 1982 there are 14000 cases of colorectal cancer diagnosed, and result in 5000 deaths each year. Currently, the physicians look for colon polyps visually through the colonoscopy exam. A standard procedure usually takes 10-30 minutes. When an abnormal polyp is found, it should be excised immediately to prevent cancer development. Fig. 1 shows an example of a polyp in a colonoscopy image and the corresponding location.

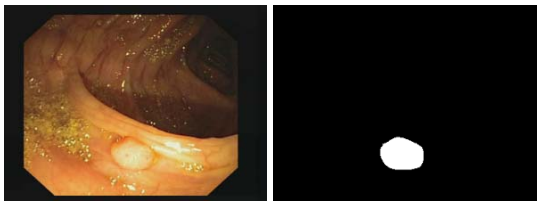

Fig. 1. Left: A polyp in a colonoscopy image; Right: its actual location

Due to the increasing incidence rate, an automatic analysis workflow will be helpful to the doctors to reduce their working load and the chance of missing detection. In this work, we propose a computer-aided segmentation system based on Convolution Neural Network (CNN). It is trained to locate the colorectal polyps from the background with normal tissue and uneven inner structure in real-time during a traditional colonoscopy exam. The core of polyps segmentation is the existing DeepLab_v3 [1] network . It uses atrous (also known as dilation) convolution to have a wider field-of-view for sensing the image content and outputting the feature response. In addition, it combines ResNet [2], multi-grid methods, and atrous spatial pyramid pooling (ASPP) in cascade to capture multi-scale feature. Since its deeper framework might lose the effectiveness to preserve and relay the signal of polyp locations, we introduced the Long-Short Term Memory networks (LSTMs) to retain the location information of polyps from DeepLab_v3. The inputs of LSTMs are feature maps from the aforementioned ResNet, multi-grid methods, and ASPP.

II. RELATED WORK

Many studies of polyp detection took machine learning based approaches [3-6]. The work in [5] tested several existing deep learning networks such as CNN-F, CNN-M, CNN-S, CNN-F MCN, CNN-M MCN, CNN-S MCN, GoogleLeNet, VGG-VD16, VGG-VD19, AlexNet and AlexNet MCN to classify colon polyps from 224x224 cropped images. Each one achieves more than 82% accuracy, where the CNN-S MCN has the highest one, 92%.

The work in [6] also used 255x255 cropped images and attempted to classify polyp as either neoplastic or hyperplastic. The database contained 1873 images for training and 284 images for validation. The authors reported a 90.1% accuracy with GoogleLeNet.

We found that most existing deep learning studies of polyp detection or classification required a cropped image as the input, and only returned a label for that image. However, the images directly from the clinical exams are always in full size. A useful computer aided system should also immediately point out the locations of polyps for the examiners.

III. METHODS

The proposed deep neural network contains two components, DeepLab_v3 and LSTMs. The former one is used to learn and extract features of polyps. The latter one is to preserve the information of polyp location from DeepLab_v3. The two networks are described as follows.

A. DeepLab_v3

DeepLab_v3 has the highest performance about mean Intersection over Union (mIoU) in PASCAL VOC 2012 competition for object detection. It has three sub-frameworks, ResNet, multi-grid methods, and ASPP in cascade. Chen *et al* [1] compared the differences between the performance of outstride 8 and outstride 16 from ResNet-101. Using the former one performs better than using the latter one because the latter one compresses more information about the image, which results in losing polyps signal. The multi-grid methods and ASPP approach result in a wider field-of-view and higher resolution in the processed feature responses.

B. Long Short-Term Memory networks

The Long Short-Term Memory networks (LSTMs) [7] preserve information for long periods. The memory cell (Ct, the horizontal path of Image_c in Fig. 2) is changing with the input gate, forget gate, and output gate. Input gate decides what information to be thrown away from the cell state. Forget

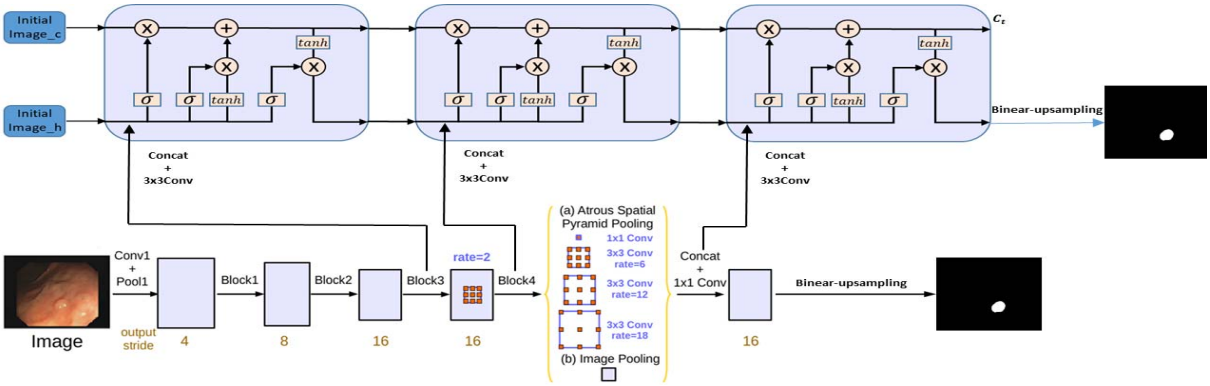

Fig.2 Parallel modules with DeepLab_v3 and LSTMs (reworked from [1,9]).

gate decides what new information to be stored in the cell state. Output gate decides what to be output. The LSTMs have the ability to remove or add information to Ct, which ensures keeping important information for longer period.

In DeepLab_v3, the ResNet, multi-grid methods, and the ASPP acquire different information. Combining several LSTMs with DeepLab_v3 in parallel (Fig. 2) is to preserve the information about polyps' position. Image_j is the result of last operation. Image_c is the preservation of previous information.

IV. EVALUATION AND EXPERIMENTAL RESULTS

'CVC-ClinicDB' is a public database released by MICCAI sub-challenge 2015, and contains 612 384x288 colonoscopy images. The boundaries of polyps in these images have been delineated as the ground truth, which made the evaluation of automatic polyp detection feasible. We selected 267 images for training and 345 images for testing. Each pixel was classified into two classes: polyp or background. Then the mIoU, conceptually an averaged classification rate measurement, can be calculated according to the ground truth. Our implementation was built on Keras backend TensorFlow.

We first applied LSTMs input for the 3x3 convolution to compress the filter size to 64, and performed batch normalization to prevent vanishing or exploding gradient. The three gates of LSTM are all 1x1 convolution as input to the horizontal path Ct. The Image_h and Image_c were initialized with zeros. They acquired and preserved important information from different layers with the union of three gates. Since the dimension of Image_h and Image_c are different from the dimension of original feature maps in DeepLab_v3, their size is reduced by the max-pooling operation, and then bilinearly upsampled to the desired spatial dimension in the end.

In Table 1 we compared the performance with two existing methods, SegNet [8] and the pure DeepLab_v3 without modification. Our proposed modification showed a higher accuracy of polyps segmentation in terms of mIOU, and maintained similar computation cost.

Table1. Performance on CVC-ClinicDB test set

Method	mIoU	Computing time per image
SegNet [8]	77.67%	0.045 (s)
DeepLab_v3 [1]	89.23%	0.02 (s)
Proposed network	93.21%	0.023 (s)

We also observed that applying LSTMs to different parts of DeepLab_v3 (referred to Fig. 2) could vary the mIOU slightly (Table 2), but the significance needs to be further verified.

Table 2. Inference strategy on the CVC-ClinicDB test set

Method	Input	ResNet	multi-grid	ASPP	mIoU
DeepLab_v3 + LSTMs			✓	✓	93.27%
		✓	✓	✓	93.21%
	✓	✓	✓	✓	93.21%

V. CONCLUSIONS

We proposed a modified framework that unites DeepLab_v3 and LSTMs to extract the dense maps. It can preserve the information about the polyps' position. A good performance 93.21% mIoU is achieved, and requires only 0.023 (s) computing time per image. Such speed is acceptable for processing image streams like classical colonoscopy exam videos. Currently we are working on integrating the system with the instruments in a clinical colonoscopy exam room to test if it could raise the detection capability of physicians.

VI. ACKNOWLEDGEMENT

We appreciate the funding support from Ministry of Science and Technology (106-2221-E-194 -036), Taiwan.

REFERENCES

[1]. CHEN, Liang-Chieh, et al. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587, 2017.

[2]. HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

[3]. MAROULIS, Dimitrios E., et al. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. Computer Methods and Programs in Biomedicine, 2003, 70.2: 151-166.

[4]. PARK, Sun Young, et al. A colon video analysis framework for polyp detection. IEEE Trans. on Biomedical Engineering, 2012, 59.5: 1408-1418.

[5]. RIBEIRO, Eduardo, et al. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification. Computational and mathematical methods in medicine, 2016.

[6]. CHEN, Peng-Jen, et al. Accurate Classification of Diminutive Colorectal Polyps Using Computer-aided Analysis. Gastroenterology, 2017.

[7]. HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. Neural computation, 1997, 9.8: 1735-1780.

[8]. BADRINARAYANAN, Vijay, et al. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv: 1511.00561, 2015.

[9]. Understanding LSTM Networks, 2015. http://colah.github.io/posts/2015-08-Understanding-LSTMs/