A NOVEL APPROACH FOR REAL-TIME COLONIC LESIONS DETECTION AND CLASSIFICATION USING INTEGRATED CNNS

Guanghui Wang& Natsumaro Kutsuna
LPixel Inc.
Bunkyo-ku, Tokyo, Japan
{oh,kutsuna}@lpixel.net

Shunsuke Kamba& Kazuki Sumiyama
Department of Endoscopy
Jikei University School of Medicine
Minato-ku, Tokyo, Japan
{kanba,kaz_sum}@jikei.ac.jp

ABSTRACT

Colorectal cancer (CRC) is one of the leading causes of cancer death. It's common to find small colorectal lesions with blurred and flat characteristics, which turn the automatic anomaly detection and classification into a very challenging problem. Previous studies have shown that deep learning is a promising way to classify lesion on static images. We use deep learning to perform detection of colorectal lesion of endoscopy images in real-time. Our approach can detect not only non-flat lesions which are easy to recognize, but also flat lesions which usually are hard to recognize by medical specialists. Because one frame can be processed in less than 0.4 seconds with 94.35% of sensitivity, it is suitable for real-time identification. The results in colonoscopy videos also shown that our model can obtain higher sensitivity than medical specialists. Deep learning is also used to classify neoplastic lesions and non-neoplastic lesions, where we achieved a sensitivity of 90.7% and an accuracy of 86.8%.

1 INTRODUCTION

Deep Learning has been used successfully in a lot of fields such as medical diagnosis(Andre et al. (2017)), self-driving(Mariusz et al. (2016)) and Robotics(Harry & Michael (2017)). Object detection is not only to detect where some objects are, but also what the object is. Object detection in real-time with deep learning has become viable since methods like *faster R-CNN*(Shaoqing et al. (2015)). By sharing features for extraction, localization and classification, this method can improve the overall accuracy with less time complexity. The approaches presented in Wei et al. (2015) and Joseph & Ali (2016) allows objects localization and classification even in computers without GPU. On the other hand, ResNet(Kaiming et al. (2015)) and Googlenet(Christian et al. (2016)) are very effective models capable to extract features in different scenarios, showing good performance even when applied on industrial applications.

Colon cancer is one of the main reasons behind cancer deaths. This cancer is usually diagnosed by endoscopy and the sensitivity varies depending on the endoscopist?s experience. Because of this, many lesions are normally missed during a colonoscopy, moreover, the shortage of endoscopists also motivates a quickly patient analysis. To overcome this problem, image-enhanced endoscopy Mingjun & Tiing (2014) has being applied to enhance mucosa surface contrast, so that lesions can be detected easily. Analysis of whole-slide histology images is another way to detect lesions[Korsuk et al. (2016)]. Recently deep learning was used to detect lesions in real-time by NEC and by the National Cancer Center in Japan. However, it is still difficult to detect flat lesions, which are easy to be overlooked by endoscopists.

This work proposes a CNN-based method for detecting all kinds of lesions including flat lesions. Because the mass data is very important to get effective features, we developed a semi-automatic annotation tool, by training 24,000 images, we got an AP (Average precision) of 0.89 for lesions. Not only static images, our model has been evaluated in video sequences and images with flat lesions, which are very hard to be recognized. In each case, the AP was higher than 0.9. Recognizing

neoplastic lesions (cancer and adenoma) and non-neoplastic lesions(nontumor) is another important issue. The proposed model achieves 86.8% of accuracy, which is higher than 60% from endoscopists.

2 DATA

The mass data is an important factor to obtain effective features. To make our annotation more accurate, we developed a semi-automatic annotation tool(Figure 1). At first, we trained 1,000 images from a pre-training by using MS COCO(Tsung-Yi et al. (2014)), and it was obtained an AP (average precision) of about 0.8 on 600 images. Because the initial data included distinct kinds of images, the domain of new data does not change the trained network drastically. In this way, instead of annotating data by hand, the only thing endoscopists need to do is to adjust the position of the bounding box. This reduced the time needed for manual annotations in 75%. Repeating the process by training the new model and collecting new data, made the annotations more and more effective.

We collected about 44,000 static endoscopy images of colon from Jikei University School of Medicine. 38,000 images of them includes at least one lesion and 6000 images are normal images without lesions. In order to obtain high quality data, protocols were stablished for annotations, including double-check. To confirm that our approach can be applied in real diagnosis, 12 videos were used as evaluation data. Furthermore, it was evaluated on 105 flat lesions, which are very difficult to distinguish.

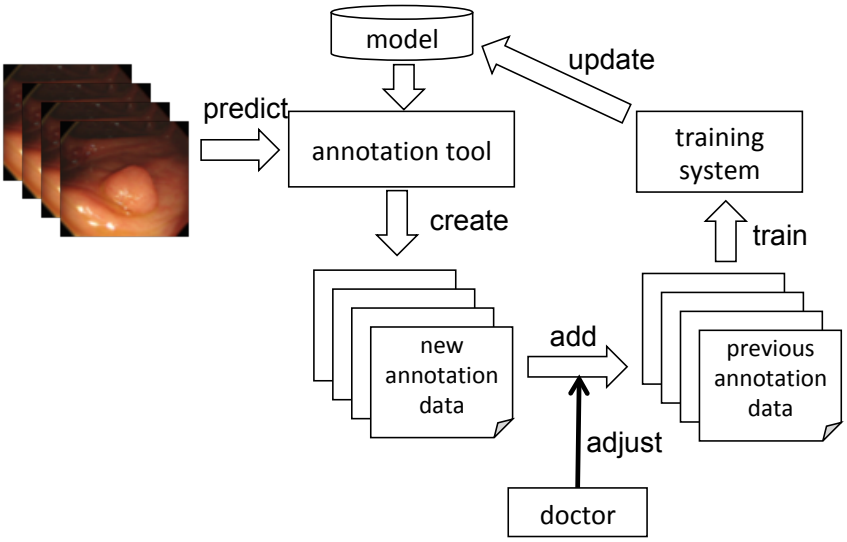


Figure 1: Semi-automatic annotation tool

3 METHODOLOGY

At present, many deep learning approaches (Faster R-CNN(Shaoqing et al. (2015)), SSD(Wei et al. (2015)), YOLO2(Joseph & Ali (2016)), PVANet(Kye-Hyeon et al. (2016))) can be used to detect objects. They got a high accuracy on normal images such as persons, cats or unanimated objects. Since exist a large number of images for training and testing with clear references of each class, a high accuracy can be achieved easily. To detect flat lesions, we included images with flat lesions for training using faster R-CNN with Inception resnet v2 as a basenet.

However, there are many normal areas(Figure 2), similar with lesions, which lead to the problem of high false positive. To overcome this problem, we use the trained model on the prediction of images without lesions, obtaining known false positives. By changing the labels of these false lesions into their corresponding class, we could increase the number of non-normal examples for training.

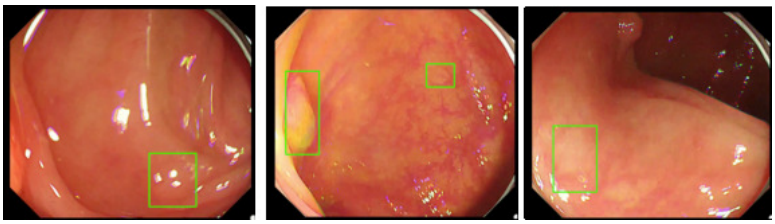


Figure 2: Normal areas similar with lesion

The deeper the neural network, the more abstract features can be extracted such as shape(Figure 3). Unlike classifying persons, cats and dogs, which have particular abstract features, in case of colorectal lesions, many normal areas have the same shape than true positive lesions, which means these features are not very suitable to classify lesions and normal areas. We assumed that low-level features that can be obtained from upper layers are more effective features.

Inspired by Resnet (Kaiming et al. (2015)) DenseNet (Huang et al. (2017)), two variations of Resnet101(Figure 4) were proposed to verify our assumption. Resnet101-v1 concatenate features from upper layers to features from deep layers. As a comparison, Resnet101-v2 enhance features from deep layers.

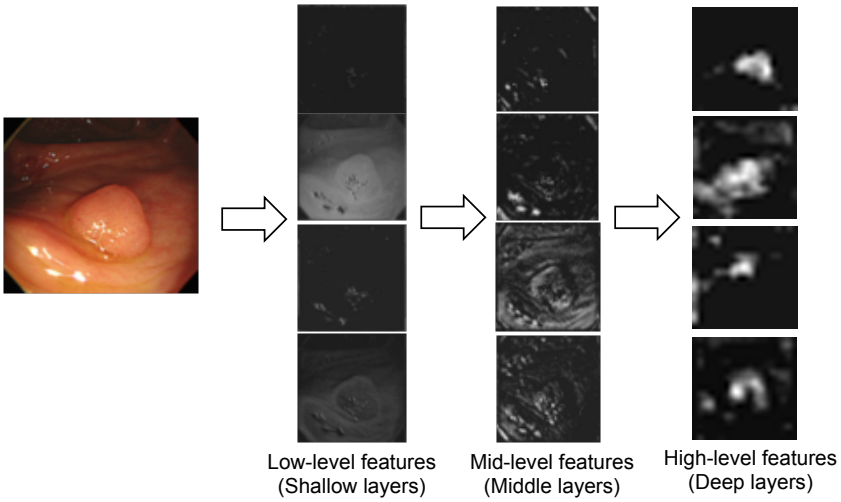


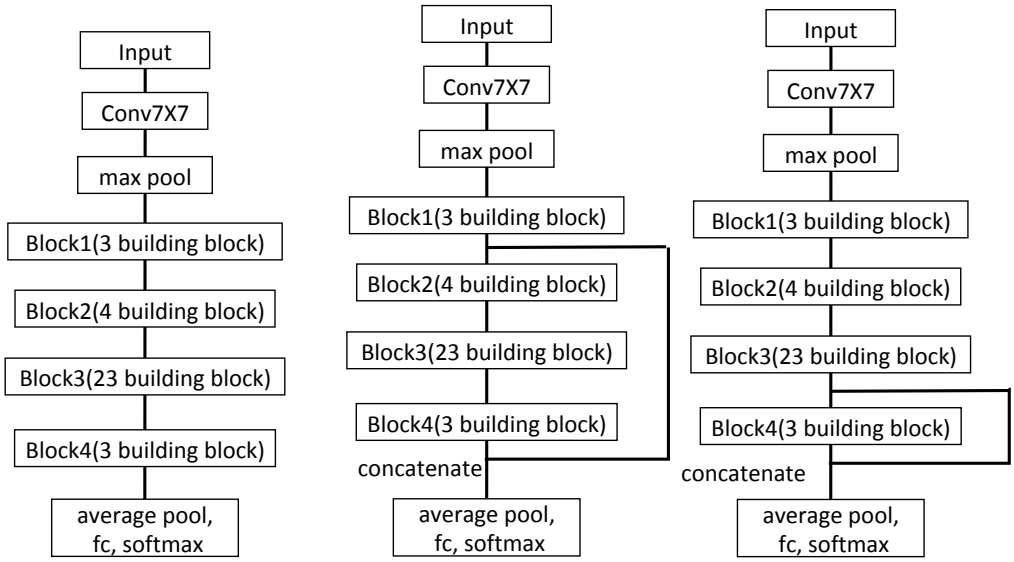
Figure 3: Normal areas similar with lesion

To classify neoplastic lesions and non-neoplastic lesions, we clipped about 46,500 lesion areas from endoscopy images to be used for training. However, there are few samples of non-tumor cancer (adenoma), which can lead to Class Imbalanced Problem (Krawczyk (2016)). To avoid this problem, we just downsampled adenoma data and make it have the same level than the amount of data. An inception resnet v2 was trained with our data and it presented a very positive performance by extracting effective features from imagenet (Deng et al. (2009)).

4 EXPERIMENTAL RESULT

4.1 DETECTION

We evaluated our model on static images. Every image includes at least one lesion, static images with flat lesions, and real videos. In all cases, it was proven that our approach using deep learning is an effective way to detect colonic lesions.



(a) Resnet101 (b) Resnet101-v1 (c) Resnet101-v2

Figure 4: Resnet101 and Variations of Resnet101. (a) is the original Resnet101. (b) is a variation of Resnet101 enhancing low-level features from shallow layers. (c) is a variation of enhancing high-level features from deep layers.

Table 1: Evaluation on Faster R-CNN using normal images

Data	Class Number	SENSITIVITY	SPECCIFITY	ACCURACY
only lesion images	1	0.9435	0.4923	0.5941
lesion images+100 normal images	2	0.9212	0.6855	0.7411
lesion images+500 normal images	2	0.7506	0.8731	0.8451
lesion images+1000 normal images	2	0.7733	**0.8863**	**0.8596**

On static images, including all lesions types and sizes, was obtained 94.45% of sensitivity. On 12 original videos, endoscopists detected 34 lesions, whereas with our method, were detected 41 lesions (7 lesions were overlooked by endoscopists). Besides that, our system did not miss any lesion detected by endoscopists.

4.2 Improvement of false positive problem with normal images

About 6,000 images were predicted by our model. In the predicted result, nearly 1,000 images included false positive lesions. We renamed those false positive lesions as "normal" label. We added 1,000 images for training with "normal" label and 24,000 images with lesions separately. By training the model with faster R-CNN+Inception resnet v2, we got the result presented in result(Table 1), which showed that using data with false positive lesions overcomes the problem of low specificity.

By enhancing features from upper layers, we got a higher specificity and accuracy comparing to the original Resnet101 and Resnet101-v2, which enhanced features from deep layers (Table 2).

Since flat lesions are very similar with normal area, more effective features were extracted by our deep neuronetwork, the sensitivity on images with flat lesions was 94.44%.

Table 2: Evaluation on Variations of Resnet101

Data	SENSITIVITY	SPECCIFITY	ACCURACY
Resnet101	0.8185	0.4506	0.5341
Resnet101-v1	0.7858	**0.5023**	**0.5655**
Resnet101-v2	0.8219	0.4303	0.5163

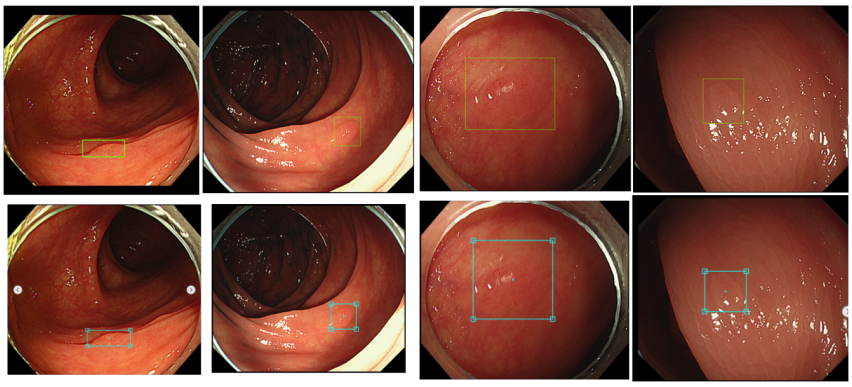


Figure 5: Detection in images with flat lesion

4.3 CLASSIFICATION

We clipped about 46,500 lesion areas from endoscopy images for classification, which were separated into 2 classes: cancer, adenoma, nontumor. We got a training data including 1,762 cancer samples, 1,246 nontumor samples, 1,422 adenoma samples. 73% accuracy was archived on 280 testing data. In the case of classifying neoplastic lesions and non-neoplastic lesions, was achieved 90.7% sensitivity, 65.9% specificity and 86.8% accuracy, which is higher than the score from endoscopists.

5 CONCLUSIONS AND FUTURE WORK

We proposed a deep learning method to detect all kinds of colonic lesions including, flat lesions and was achieved a high sensitivity. By using false positive lesions, the problem of high specificity has been overcome.However, we would like to figure out how false positive lesions improve the performance and how many false positive lesions in training data is the best to get the better performance.

We have shown that low-level features are important to classify lesions and normal areas similar with lesion and enhancing low-level features is an effective way to overcome the problem of high specificity. The concatenating method will consider more sophisticated models and deep learning approaches.

ACKNOWLEDGMENTS

This research is supported by ”The Practical Research for Innovative Cancer Control” from Japan Agency, Medical Research and development, AMED.

REFERENCES

Esteva Andre, Kuprel Brett, A. Novoa Roberto, Ko Justin, M. Swetter Susan, M. Blau Helen, and Thrun Sebastian. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017. doi: 10.1038/nature21056.

Szegedy Christian, Ioffe Sergey, and Vanhoucke Vincent. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. ImageNet: a large-scale hierarchical image database. In *CVPR*, pp. 248–255. IEEE Computer Society, 2009. ISBN 978-1-4244-3992-8.

A. Pierson Harry and S. Gashler Michael. Deep learning in robotics: A review of recent research. *CoRR*, abs/1707.07217, 2017.

Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Redmon Joseph and Farhadi Ali. YOLO9000: better, faster, stronger. *CoRR*, abs/1612.08242, 2016.

He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

Sirinukunwattana Korsuk, E Ahmed Raza Shan, Tsang Yee-Wah, R. J. Snead David, A. Cree Ian, and M. Rajpoot Nasir. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 2016.

Bartosz Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4):221–232, Nov 2016. ISSN 2192-6360. doi: 10.1007/s13748-016-0094-0.

Kim Kye-Hyeon, Cheon Yeongjae, Hong Sanghoon, Roh Byung-Seok, and Park Minje. PVANET: deep but lightweight neural networks for real-time object detection. *CoRR*, abs/1608.08021, 2016.

Bojarski Mariusz, Del Testa Davide, Dworakowski Daniel, Firner Bernhard, Flepp Beat, Goyal Prasoon, D. Jackel Lawrence, Monfort Mathew, Muller Urs, Zhang Jiakai, Zhang Xin, Zhao Jake, and Zieba Karol. End to end learning for self-driving cars. *CoRR*, abs/1604.07316, 2016.

Song Mingjun and Leong Ang Tiing. *Early detection of early gastric cancer using image-enhanced endoscopy: Current trends*, volume 3. Elsevier, 2014.

Ren Shaoqing, He Kaiming, B. Girshick Ross, and Sun Jian. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

Lin Tsung-Yi, Maire Michael, J. Belongie Serge, D. Bourdev Lubomir, B. Girshick Ross, Hays James, Perona Pietro, Ramanan Deva, Dollár Piotr, and Zitnick C. Lawrence. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.

Liu Wei, Anguelov Dragomir, Erhan Dumitru, Szegedy Christian, E. Reed Scott, Fu Cheng-Yang, and C. Berg Alexander. SSD: single shot multibox detector. *CoRR*, abs/1512.02325, 2015.