Polyp detection in Colonoscopy Videos Using Deeply-Learned Hierarchical Features

Sungheon Park, Myunggi Lee and Nojun Kwak

Department of Transdisciplinary Science, Seoul National University, Suwon, Korea

Abstract

This paper summarizes the method of polyp detection in colonoscopy images and provides preliminary results to participate in ISBI 2015 Grand Challenge on Automatic Polyp Detection in Colonoscopy videos. The key aspect of the proposed method is to learn hierarchical features using convolutional neural network. The features are learned in different scales to provide scale-invariant features through the convolutional neural network, and then each pixel in the colonoscopy image is classified as polyp pixel or non-polyp pixel through fully connected network. The result is refined via smooth filtering and thresholding step. Experimental result shows that the proposed neural network can classify patches of polyp and non-polyp region with an accuracy of about 90%.

Key Words: Convolutional neural network, hierarchical feature learning.

1 Introduction of Research Team

Our group, Machine Intelligence and Pattern Analysis Laboratory, aims to solve computer vision and machine learning problems. Recently, our interest lies within deep learning algorithms and its applications, especially the ones related with medical imaging. Although this is the first time we deal with endoscopy image analysis, we previously conducted research on automatic Bosniak classification problem from renal CT scan images. Our previous research experiences have inspired us to participate in this challenge.

2 Methodology

Convolutional neural network (CNN) is a main tool for the proposed method. Along with the recent improvements on deep learning, CNN is recently applied in many areas related with medical imaging [1, 2, 3]. Our approach is motivated from the work of [4]. Polyps have various colors and shapes. Especially, size of polyps are different, and even the size of the same polyp can be varied as the camera moves. Therefore, as in [4], learning multi-scale representation of polyp can improve the classification performance. In the following subsections, it is explained how multi-scale representation can be learned through CNN.

2.1 Training Data Preparation

As a first step, border lines of input images are detected and cropped using Canny edge detector. Then, triangular regions at the corner of the rectangular image are filled by reflecting nearby image regions. As a consequence, rectangular images which do not contain black boundary region are generated.
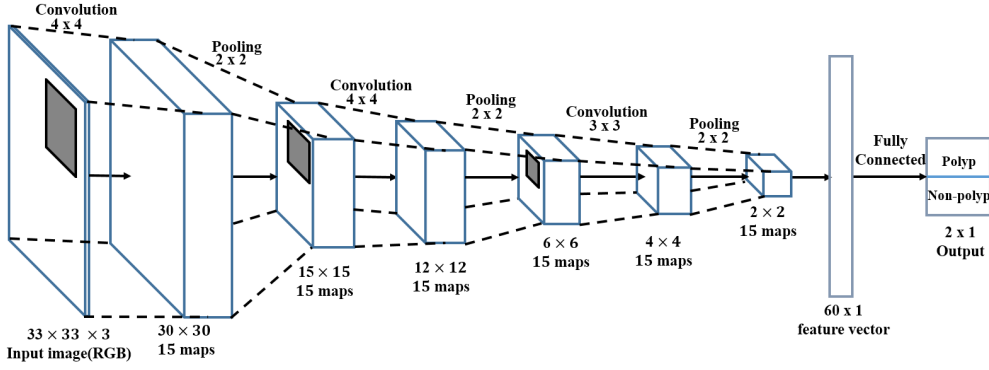
Figure 1: Structure of CNN. The network consists of 3 convolutional layers and 3 max-pooling layers. 60 dimension feature vector is generated through the network.

Training images will be cropped into small patches, and the patches are fed to the CNN in the proposed method. We used CVC-ClinicDB to organize training and test set. Among the 612 images, 550 images are selected as a training set, and the other 62 images are selected as a validation set. From the training set, 50 pixels of non-polyp region and polyp region are randomly picked and patches around the selected pixels are extracted with 3 different scales, original image, half-sized image, and quarter-sized image. 33×33 patches are extracted for each scale at the same location. Therefore, the entire training set contains 165000 image patches with the same number of polyp and non-polyp region. Each training image patch is locally normalized so that 15×15 neighbors have uniform mean and variance.

2.2 Hierarchical Feature Learning Using Convolutional Neural Network

To learn the scale-invariant features through CNN, image patches that are used as an input to CNN are cropped in different scales. Same size of image patches with different scales are fed to the single CNN. The structure of the CNN is illustrated in Figure 1. The CNN consists of 3 convolutional layers and max-pooling step is performed after every convolutional layer. 15 filter banks are used in each convolutional layer. As a result, 60 dimension feature is generated through the CNN for each image patch. The cost function of the classification layer is the cross entropy loss function, which is defined as

$$E(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\mathbf{1}\{y^{(i)} = k\} - \mathrm{P}(y^{(i)} = k|x^{(i)}; \theta)), \tag{1}$$

where m is the number of training image patches, $x^{(i)}$ and $y^{(i)}$ are ith training images and labels respectively, $\mathbf{1}\{\cdot\}$ is an indicator function, and $k = 0$ or 1 for two-class problems. Sigmoid function is used for an activation function of the CNN. The parameters of CNN is optimized via stochastic gradient method with minibatch size of 256 and with no additional regularization cost. Three different scales are used to extract multi-scale image patch as explained in Section 2.1.

By using different scales of an image as an input, it can be expected that the CNN will learn the features that represents the shape of polyps in various scales. Hence, the robustness to scale change can be imposed on the trained CNN. After the CNN is trained, the classification layer is discarded, and 60 dimension features will be used as an input to another neural network, which will be explained in the next subsection.

2.3 Classifier Using Multi-scale Features

Through the CNN trained in Section 2.2, single image patch returns 60 dimension feature vector. To exploit multi-scale representation, features from the patch of the same location with different scales are concatenated.
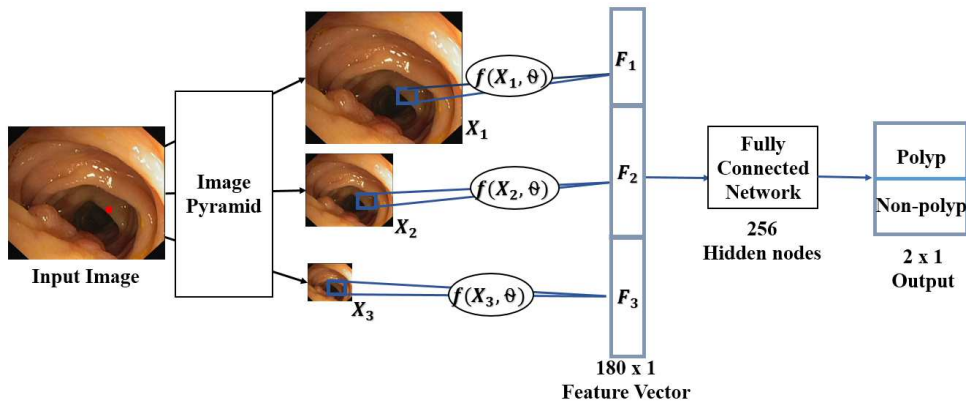
Figure 2: Overview of testing algorithm. To classify a pixel in the input image (red dot), image patches at different scales pass the same CNN. The resulting features are concatenated and used as an input to the fully connected network which classifies the pixel as polyp or non-polyp region.

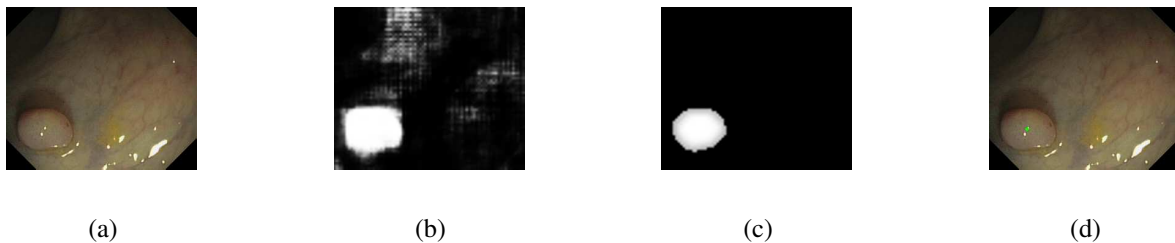|			
(a)	(b)	(c)	(d)

Figure 3: An example of polyp detection result. (a) Input image. (b) Grayscale probability map. Black is mapped to probability of 0 and white is mapped to probability of 1. (c) Smoothed probability map after post-processing (d) Final result. The location of the polyp is marked with a green dot.

Consequently, 180 dimension feature vector is generated. Using this multi-scale feature vector, 2-layer fully connected network is trained. The fully connected network has one hidden layer with 256 hidden nodes. The cost function of the classification layer is the same as (1). It is trained via stochastic gradient method with no additional regularization cost as in the case of CNN. The whole structure of the network including the CNN and the fully connected network is illustrated in Figure 2.

2.4 Testing and Post-Processing

In the test phase, each pixel of the test image should be classified into polyp or non-polyp region. However, the computation cost for classifying the whole pixels of the image is expensive, and it is unnecessary since patches of neighboring pixels are almost the same and output similar results. Thus, we extract patches of a test image in a stride of 4. Resulting probability map has smaller size than the original image, and we simply upsample the probability map to the equal size of the test image. As a result, we get the dense probability map of a test image, whose value means the probability that the pixel is classified as a polyp.

However, the resulting probability map is usually noisy (Figure 3(b)). Therefore, as a post-processing step, the probability map is smoothed to via 5×5 gaussian filter and 9×9 median filter, yielding smoothly varying probability map. Next, the smoothed probability map is thresholded to find the candidate polyp regions. Pixels with the probability lower than 0.65 is set to 0. Lastly, connected component of non-zero regions are found, and the center of mass of each connected component is determined as the center of polyp. Figure 3 shows an example of polyp detection process and the result.

True positive	False positive	False negative	Precision	Recall	F-score
48	25	10	0.6575	0.8276	0.7328

Table 1: Detection performance on the validation set from CVC-ClinicDB.

3 Preliminary Results

In this section, preliminary results on CVC-ClinicDB database is illustrated. We are unable to conduct the experiments on ASU-Mayo database due to lack of time.

As mentioned before, the validation set of CVC-ClinicDB consists of 62 images. First, to check the performance of our CNN and the classifier, we randomly select 50 pixels of non-polyp region and polyp region in each image of the validation set and test the patches to classify whether the selected pixel is a polyp region or not. Among the 6200 test pixels, 91.60% of the test pixels are classified correctly as polyp or non-polyp region.

Next, polyp detection is performed on each image of the validation set following the method explained in Section 2.4. True positive, false positive, and false negative results are counted according to the rules of the challenge. The result is presented in Table 1. Precision, recall, and F-score are also calculated. The proposed method detected 48 polyps among 64 polyps in 62 images, which shows about 75% detection rate. Despite the high accuracy in pixel classification test, F-score is not so high due to the large number of false positives. We expect that the number of false alarms can be reduced by appropriate post-processing using the noisy probability map. We will also test the performance of our algorithm on the ASU-Mayo database.

4 Future Work

We are planning to modify our algorithm in a number of ways to improve the detection quality until the competition. First, tuning the parameter of the CNN can affect the performance. Adding regularization cost can be also considered for better feature learning. Next, the post-processing step can be improved by using the labeling method in [4, 5] or using graph cut optimization [6]. We expect this modification will reduce number of false alarms in test images. Also, the proposed detection algorithm can be easily extended to polyp segmentation.

References

[1] D. Ciresan, A. Giusti, L. M. Gambardella, J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images", *Advances in neural information processing systems,* 2843-2851, 2012.

[2] D. C. Cirean, A. Giusti, L. M. Gambardella, J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks", *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013,* 411-418, 2013.

[3] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R. M. Summers, "A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations", *Medical Image Computing and Computer-Assisted Intervention,* 520-527, 2014.

[4] C. Farabet, C. Couprie, L. Najman, Y. LeCun, "Learning hierarchical features for scene labeling", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35.8:1915-1929, 2013.

[5] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs", *arXiv preprint,* 1412.7062, 2014.

[6] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 23.11:1222-1239, 2001.