CONVOLUTIONAL NEURAL NETWORK AS A FEATURE EXTRACTOR FOR AUTOMATIC POLYP DETECTION

Bilal Taha, Jorge Dias, Naoufel Werghi

Department of Electrical and Computer Engineering
Khalifa University, Abu Dhabi, UAE

ABSTRACT

Colorectal cancer is one of the major causes of cancer deaths worldwide. To achieve early cancer screening, detecting the presence of polyps in the colon tract is the preferred technique. In this paper, a deep learning approach for identifying polyps in colonoscopy images is proposed. The novelty of our technique stems from the fact that it fully employs a pre-trained Convolutional Neural Network (CNN) architecture as a feature extractor. Contrary to the conventional methods which either perform fine-tuning or train the CNN from scratch, we utilize the CNN output features as an input to train the Support Vector Machine (SVM) Classifier. The efficiency of the presented framework is demonstrated on the public CVC ColonDB, in which the experimental results indicate that our methodology significantly outperforms other competitive paradigms.

Index Terms— Automatic polyp detection, Deep learning, CNN, feature extractor.

1. INTRODUCTION

According to the Centers for Disease Control and Prevention (CDC) in the United States, colorectal cancer (CRC) is the third most common cancer in the world [1]. Colorectal cancer starts with small protrusions growing inside the colorectal which could eventually lead to CRC [1]. These protrusions are known as polyps. Fig. 1 shows examples of polyps with different shapes and appearances. It is the ability to detect polyps and remove them in early stages that saves more lives and results in better prevention of CRC [2]. The most common method for this process is by visual inspection using endoscopic videos. However, clinical examination is not sufficient enough as a final judgment since there are many sources of error and false diagnosis. These sources of errors could be correlated with the medical level of expertise and the nature and appearance of the polyp itself. Indeed, the variety of shapes and sizes in which the polyps appear, and the limited field of view inside the colon, makes it difficult to the clinical examiner to keep continuous and consistent evaluations on detecting the polyps and support diagnosis. Thus, its turned out essential to develop an automated system to support the physician to detect and classify polyps. Several computer aided diagnosis (CAD) systems have developed to that end. Here we report the main works and we refer the reader to [3] for a more exhaustive survey. Some authors proposed to use texture features [4, 5], shape features [6, 7] or feature fusion [8, 9] coupled with standard classifiers. However, these methods still suffer from a high false positive rate. In addition, defining optimal descriptors proper for polyp presentation seemed to be quite complex and dependent on the correct tuning of multiple parameters. To bypass this problem, there has been a recent trend to use deep learning approaches to benefit from its powerful feature learning capacity [10]. The authors of [11] have used and trained the Convolutional Neural Network (CNN) from scratch to outperform the existing methods. However, this approach requires a huge database for the CNN to learn features from the images. In a subsequent work the same authors in [12] implemented a fine tune CNN approach demonstrating a competitive performance when compared to training a CNN from scratch. The latter work represents a neat advance in terms of efficient CNN usage. Indeed, training the CNN from scratch while it creates abstract features more related to the database, is time and computation demanding, and requires a huge labeled database which might be impractical in real applications and very costly. Fine tuning, while reducing the size of the data needed, still requires substantial data training to get the parameters accurately tuned to a specific database.

Fig. 1: Polyp samples from CVC-ColonDB database

In this work, we focus more on the CNN deployment efficiency aspect, and propose a method that capitalizes on the capacity of well-established convolution neural network architecture for producing generic features, which can be tailored for a specific classification application. The key contributions in this work are (1) Efficient employment of CNN architecture without the need for training from scratch; (2) Accommodating partial polyp appearances in the colonoscopy images; (3) Finding middle-layer features of

the computational architecture that can be more effective than the end-layer features, as would be expected; (4) Evaluating and showing the superiority of our approach when compared with competitive methods.

The remainder of the paper will be organized as follows: Section 2 introduces the proposed approach and its rational. Section 3 describes the different experiments and the related results. Section 4 concludes the paper.

2. THE PROPOSED METHOD

In image analysis, designing the appropriate features for a given interpretation task has been a central problem in computer vision and medical image understanding. Explicit feature design extraction for medical images require subject-matter expertise. In this process, the visual information on which the physician relies in his assessment is not necessarily reflected into a suitable computational representation. Moreover, the practical considerations in the extraction and the usage of these features make the reproducibility of such related methods often problematic [13]. To overcome these challenges, we propose a deep learning approach, whereby Convolution Neural Network (CNN) is employed to replace hand-crafted and customized features that are often strongly sensitive to multiple parameters.

It is known that through a hierarchical unsupervised or semi-supervised feature design, CNN's can produce effective representation of the visual data [10]. Basically, a CNN architecture is composed of a sequence of cascading layers performing basic operations such as convolution, subsampling, followed by another sequence of fully connected layers, which act similarly as a classic artificial neural network.

In another hand, training a CNN network from scratch requires a large dataset. Such a process is quite tedious, in addition large datasets cannot be afforded easily in medical applications, including dataset of polyp detection. Apart of the computational resources requirements, there is no systematic guidelines as for the optimal choice of the architecture in terms of depth (number of layers) and structure.

An economic alternative is to use pre-trained CNN architectures, that are proven to have good performance through training and validation over a huge database, and then tune, via training conducted on a specific application dataset, the pre-trained weights of the architecture. This procedure, known as fine-tuning, which can be performed either across the whole CNN or at specific layers.

IIn our approach, we advocate the hypothesis that a trained CNN architecture embeds sufficiently rich feature representations that can be utilized as an input to train a standard classifier, such as the Support Vector Machine(SVM), relieving thus the system from laborious training from scratch or fine tuning. Therefore a pre-trained CNN is then deployed as a feature extractor for our specific image interpretation task of polyp detection, as depicted in the block diagram in Fig.2. There are several pre-trained CNN architectures that can be investigated, such as GoogleNet [14] and VGGNet [15]. In our method, we explored AlexNet [16]. This CNN architecture was trained with 1.2 million images for 1000 different classes, thus the learned features are expected to span a large spectrum of visual information. The main layers of the AlexNet architecture is briefly described in Table 1. As

Fig. 2: Block diagram for the CNN as a feature extractor

Table 1: Summary of AlexNet architecture

Layer	Type	Input	Kernel	Stride	Pad	Output
Data	Input image	227x227x3	N/A	N/A	N/A	227x227x3
conv1	Conv	227x227x3	11x11	4	0	96x55x55
pool1	Max pooling	55x55x96	3x3	2	0	96x27x27
conv2	Conv	27x27x96	5x5	1	2	256x27x27
pool2	Max pooling	27x27x256	3x3	2	0	256x13x13
conv3	Conv	13x13x256	3x3	1	1	384x13x13
conv4	Conv	13x13x384	3x3	1	1	384x13x13
conv5	Conv	13x13x384	3x3	1	1	256x13x13
pool5	Max pooling	13x13x256	3x3	2	0	256x6x6
FC6	fully connected	6x6x256	6x6	1	0	4096x1
FC7	fully connected	1x4096	1x1	1	0	4096x1
FC8	fully connected	1x4096	1x1	1	0	1000x1

aforementioned, features from first layers are too generic to be employed as discriminative descriptors, so we investigated features from the middle layers and onward, namely, Conv4 till FC8. The output from one of these layers will be a sort of feature encoding for a full (or partial) colonoscopy image. These features will be then fed into the subsequent classifier block, SVM, as depicted in Fig. 2. An SVM converges to a global and unique solution, and has the capacity to deal with a high-dimension input without compromising the computational complexity, and thus can map the huge number of feature vectors x_i, $i = 1...N$, generated by the CNN. When training an SVM, each feature vector is given a label either polyp or non-polyp (abnormal, normal) to create the feature-class pair $\{x, y\}$. Therefore, given L features $\{x_i, y_i\}$ such that $i = 1...L$, and $y_i \in \{1, -1\}$, $x \in \Re^D$, where D is the vector size. A hyper-plane separating the two classes could be written as

$$w^T x + b = 0 \tag{1}$$

the w is known as a weight vector which is normal to the separation hyper-plane, and b is a bias. In order to separate the two classes with a hyper-plane, equation (2) should be optimized

$$min(\frac{1}{2}w^T w + C\sum_{i=1}^{L}\xi_i) \tag{2}$$

subject to the constrain $y_i((w^T x_i) + b) >= 1 - \xi_i$, where $\xi_i >= 0$ for $i = 1, ..., L$, and C is the penalty parameter. This will lead to the optimal hyper-plane that minimizes the

distance between itself and all the training examples. The optimal hyper-plane, allows a classification to be done according to a decision function such as:

$$f(x) = \text{sgn}(w^T x + b) \quad (3)$$

3. EXPERIMENTATION

To evaluate the performance of our method, CVC-ColonDB [17] was used for training and testing. This database consists of 15 short colonoscopy videos for different 15 cases. It includes different polyp sizes, appearances and colors. We conducted a series of extensive experiments on the specified database that aimed to assess the performance of the CNN as a feature extractor and its effectiveness in the detection scenarios. In these experiments we studied the effects of 1) Selecting features from different layers of the CNN, 2) The image patch size, and 3) The polyp appearance in each patch, that is the minimum portion of polyp area visible in a patch to be considered as genuine case (true positive).

In the first experiment, the main focus is on the quality of extracted features from the CNN. The features were employed from different layers from the pre-trained CNN. AlexNet was trained using the ImageNet database which consists of non-medical images, therefore there is a need to know the best layer that will provide the best features discriminating polyp from non-polyp cases. As we mentioned earlier, we considered only deep layers, starting from Conv4. In this transfer learning scheme, the layers up to the output features layers are frozen and the output features are used to train the SVM classifier. For example, considering Conv5, as the feature output layer, we keep the weights across the layers conv1 to conv5 at their pre-trained values, while training the SVM classifier. While this scheme reduces the number of trained entities, the number of features remain large, as an example, the dimension of the obtained feature vector from the fourth convolution layer (C4) is 13x13x384 which is equal to 64896 features. In this experiment, the patch size was fixed to 16 patches/image, each patch is 100x100, making a total of 4800 patches, and we considered any polyp appearance to be a positive case meaning no threshold as for the minimum size of its partial appearance. For training protocol, we adopted the 70%, 30% for training and testing, respectively. Fig.4.a depicts the obtained ROC curves related to the different output layers. For instance, C5 refers to the features coming out from the layer Conv5. Table 2 reports the best recall and precision performance obtained for each layer. It is interesting to notice that the top performance is obtained with features coming out from a middle layer (conv5), which are less descriptive than their deeper layered counterparts (e.g. FC8 - see Table 1).

In the second experiment, the effect of patch size on the performance of the CNN as a feature extractor is investigated. The patches are constructed by utilizing a sliding window vertically and horizontally without any overlapping, dividing thus the image into patches. The choice of optimal patch size is a bit problematic. While reducing the patch size increases the volume of samples used for training and testing, which is good for the over-fitting problem, it increases also the number of small partial polyps, and thus jeopardizing the ability to extract good features. On the other hand, the big patch size reduces the probability of partial polyps in each patch but, at the same time it reduces the size of the training samples. To address this issue, we investigated the optimal size empirically by experimenting three patch sizes 200x200, 100x100, and 50x50. Fig. 3. (a) depicts the three patch sizes on the polyp images, respectively. Fig. 4.b shows the ROC curves related to each patch size whereas the best recall and precision values for experiment 2 are reported in Table 2.

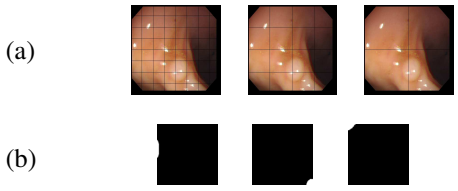

Fig. 3: (a) samples of different patch size, (b) masks corresponding to patch samples with small polyp portion.

In the third experiment, we studied the effect of the threshold on the polyp area portion for considering a patch as containing a polyp (positive sample) or not (negative sample). The motivation behind this experiment is that in practical situations, these small polyp parts are not noticeable by the physicians and thus should be considered rather than as a negative sample. The mask samples already reported in Fig. 3. (b) illustrate examples of odd small polyp portion in a patch. In this experiment, the patch size was fixed to 100x100 and the features were taken after the fifth convolution layer (C5). ROC curves obtained with different thresholds are depicted in Fig 4.c. We notice that the best ROC curve corresponds to a 7% threshold. This is also reflected in the recall and precision scores in Table 2.

Normally, one should test all the combinations of the three parameters (CNN layer, patch size and polyp portion threshold) to come out with a combination that achieves the best performance. However, performing such exhaustive procedure needs to conduct $5 \times 3 \times 4 = 60$ different training. As a less demanding alternative, though sub-optimal, we considered the best parameter in each of the previous three experiments (Conv5 layer, 50×50, 7%), then re-evaluated the performance of the system. We compared our method with six state of the art methods that used the same database [17, 18, 19, 20, 21, 22]. In [17] the authors proposed an algorithm based on the polyp distinct shape and used a segmentation algorithm, to minimize the number of most likely polyp. Then, they utilized Sector Accumulation-Depth of Valleys Accumulation (SA-DOVA) as a descriptor for the

detection process. In their subsequent work [20], they have improved their methodolgy by focusing more on the pre-processing stage where they tackled the effect of specular lights and blood vessels. Furthermore, authors of [21] proposed a system to handle the imbalance size between the polyp and non-polyp samples. They have employed least-squares analysis to learn different types of features. In [18], the authors used Haar features, one layer of classification, and a voting method to detect polyps. Then, in [19] they implemented a two stage edge classification scheme to obtain a refined edge map and the direction of the normal for the polyp-like edges. Afterwards, a new voting scheme is applied to the refined edge map to localize polyps by detecting curvy boundaries. In a recent work, the authors in [22] augmented their previous shape-based approach with context-clues information derived around the polyp boundaries. Table 3 reports the performances of the seven methods. We found that our approach outperforms all the existing paradigms in terms of recall with a score of 96%, concurrently, illustrating a very close score in term of precision compared with the best value, where the difference is only 0.3%.

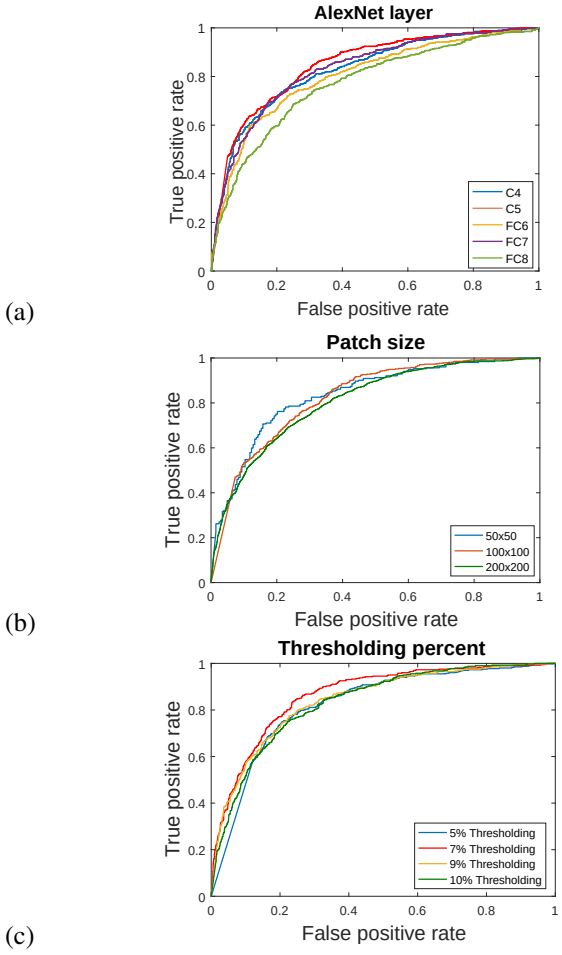

Fig. 4: ROC curves related to experiments 1(a), 2(b) and 3(c).

Table 2: The recall and precision values for each experiment

		Recall	Precision
Experiment 1			
CNN feature	Conv4	74%	90.8%
	Conv5	**89.6%**	**86.3%**
	FC6	98.9%	76.5%
	FC7	99.1%	76.1%
	FC8	97.2%	77.9%
Experiment 2			
patch size	200×200	79.8%	72.3%
	100 × 100	65.3%	91.4%
	50 × 50	**99.4%**	**85.2%**
Experiment 3			
Threshold percent	5%	56.2%	95.4%
	7%	**96.3%**	**84.3%**
	9%	91.1%	87.4%
	10%	90.4%	86.8%

Table 3: Recall and precision scores in percent by setting the parameters according to the best results in each experiment compared to other paradigms.

Method	[17]	[18]	[21]	[20]	[19]	[22]	Ours
Recall	47.15	60	70.6	67.6	80	88	**96**
Precision	71.6	88	70.7	-	**93**	-	92.7

4. CONCLUSION

In this work, we have introduced a deep learning solution for detecting polyps from colonoscopy. The novel deployment of the AlexNet, a pre-trained architecture used as a feature extractor, along with a classical SVM classifier was proposed. By adopting this approach, the system circumvents the high computational complexity and high resource demand of CNN required in training from scratch and fine-tuning. The series of experiments conducted with the CVC colonDB database, confirmed the rationale behind our hypothesis, which implies that the features derived from a CNN architecture (pre-trained by means of colossal datasets), embed sufficient discriminatory information that could be tailored to our specific CVC-ColonDB dataset. The comparison with state of the art methods clearly confirmed the boost of performance brought by our method. For future work, we plan to deploy our method on other polyp datasets including ASU-Mayo clinic database, as well as other standard trained CNN architectures such as VGGNet.

5. REFERENCES

[1] F. Haggar and R. Boushey, "Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors," *Clinics in Colon and Rectal Surgery*, vol. 22, no. 04, pp. 191–197, nov 2009.

[2] A. Castells, "Choosing the optimal method in programmatic colorectal cancer screening: current evidence and controversies," *Therapeutic Advances in Gastroenterology*, vol. 8, no. 4, pp. 221–233, mar 2015.

[3] B. Taha, N. Werghi, and J. Dias, "Automatic polyp detection in endoscopy videos: A survey," in *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, Feb 2017, pp. 233–240.

[4] G.D. Magoulas, V.P. Plagianakos, and M.N. Vrahatis, "Neural network-based colonoscopic diagnosis using on-line learning and differential evolution," *Applied Soft Computing*, vol. 4, no. 4, pp. 369–379, sep 2004.

[5] D.E. Maroulis et al. "CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames," *Computer Methods and Programs in Biomedicine*, vol. 70, no. 2, pp. 151–166, feb 2003.

[6] Y. Wang et al. "Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1379–1389, July 2014.

[7] C. van Wijk et al. "Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow," *IEEE Transactions on Medical Imaging*, vol. 29, no. 3, pp. 688–698, March 2010.

[8] A. El Khatib, N. Werghi, and H. Al-Ahmad, "Automatic polyp detection: A comparative study," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 2669–2672.

[9] A. E. Khatib, N. Werghi, and H. Al-Ahmad, "Enhancing automatic polyp detection accuracy using fusion techniques," in *Int. Midwest Symposium on Circuits and Systems (MWSCAS)*, Oct 2016, pp. 1–4.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, may 2015.

[11] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 79–83.

[12] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[13] J. Kovacevic P. Vandewalle and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37–47, v 2009.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[16] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, P. Bartlett, F.c.n. Pereira, C.j.c. Burges, L. Bottou, and K.q. Weinberger, Eds., pp. 1106–1114. 2012.

[17] J. Bernal, J. Snchez, and F. Vilario, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166 – 3182, 2012.

[18] N. Tajbakhsh et al. *A Classification-Enhanced Vote Accumulation Scheme for Detecting Colonic Polyps*, pp. 53–62, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[19] N. Tajbakhsh, C. Chi, S. R. Gurudu, and J. Liang, "Automatic polyp detection from learned boundaries," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, April 2014, pp. 97–100.

[20] J. Bernal, J. Snchez, and F. Vilario, "Impact of image preprocessing methods on polyp localization in colonoscopy frames," in *Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013, pp. 7350–7354.

[21] S. H. Bae and K. J. Yoon, "Polyp detection via imbalanced learning and discriminative feature learning," *IEEE Transactions on Medical Imaging*, vol. 34, no. 11, pp. 2379–2393, Nov 2015.

[22] N. Tajbakhsh and S. R. Gurudu and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, Feb 2016.