

Colonic Polyp Classification with Convolutional Neural Networks

Eduardo Ribeiro^{1,2} and Andreas Uhl¹

¹University of Salzburg - Department of Computer Sciences - Salzburg, Austria

²Federal University of Tocantins - Department of Computer Sciences - Tocantins, Brasil

Michael Häfner

St. Elisabeth Hospital
Vienna, Austria

Abstract—Texture patch classification is an important task in many different computer-aided medical systems. Convolutional Neural Networks (CNN's) have become state-of-the-art for many computer vision tasks in recent years. In this paper, we propose the use of CNN's for the automated classification of colonic mucosa for colon polyp staging in the context of colon cancer screening. This deep learning approach has the property of extracting features and classifying images in the same architecture by exploiting directly the input image pixels being successful in handling distortions such as different light conditions, presence of partial occlusions, etc. For this type of deep learning approach it is common to require that the database contains large amounts of data, which is quite rare in the medical field. The method proposed allows the use of small patches (subimages) to increase the size of the database as well to classify different regions in the same image. We show experimentally that this model is more efficient than some of the commonly used features for colonic polyp classification.

Index Terms—Deep Learning, Colonic Polyp Classification, Convolutional Neural Networks

I. INTRODUCTION

Due to the size and complexity of the gastrointestinal tract, many diseases are associated with it, for example: adenomas, polyps, Crohn's disease, celiac disease, Helicobacter pylori infection, among others. However, the leading cause of death related to intestinal tract is caused by the growth of cancerous cells (polyps) in its various parts. Especially in the final segment of the large intestine (colon) and rectum, the colonic polyps have a rather high prevalence and are known to either develop into cancer or to be precursors of colon cancer.

The diagnosis of cancer in an advanced stage increases the mortality risk among patients with color-rectal cancer and can be detected by a physician through an endoscopy procedure. The use of this endoscopic apparatus integrated with high resolution acquisition devices further expanded the research in clinical decision support system area. Intelligent systems can assist in many aspects of colon polyp diagnosis such as accentuating parts of the colon that can possibly have lesions or polyps while the physician performs the colonoscopy procedure, or generating automatic reports about parts of colonoscopy videos that require more attention when they are being analyzed by the physician. Such systems are used to support medical diagnosis, detecting abnormal lesions and/or classifying them, improving the readability of the information, segmenting areas of interest or even predicting possible diagnosis automatically [1], [2].

In the literature, apart from being based on traditional low-resolution white-light colonoscopy, some studies focus mainly on the use of computer-aided diagnosis (CAD) systems

related to more advanced colonoscopic images and videos. For computer assisted staging of colon polyps, high-magnification colonoscopes have been used, providing images which are up to 150-fold magnified, thus uncovering the fine surface structure of the mucosa as well as small lesions. Depending on the light source used, colon cancer-oriented CAD systems are divided into two categories: High-magnification chromoendoscopy [3], [1] and high-magnification endoscopy combined with narrow band imaging [4], [5]. However, these expensive devices are only used in larger center and require intensive training of the endoscopist to deliver high quality imagery. Recently, High-Definition (HD) colonoscopes represent a significant advance and are on the way to become clinical standard due to the significantly better image quality (and reasonable costs). Example images of colonic polyps, acquired with such an endoscope, are given in Fig. 1 (a).

In this work we used highly detailed images acquired by a HD endoscope without chromoendoscopy (staining the mucosa). Instead, we employ Pentax virtual chromo-endoscopy (i-Scan technology) which is a method consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces [6]. In Fig. 1 (b), an adenomatous polyp acquired using the i-Scan 1 image enhancement technology can be seen [7].

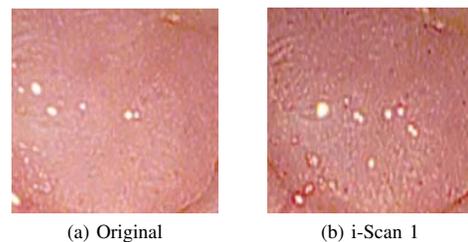


Fig. 1: Images of a polyp without image enhancement (a) and using digital i-Scan 1 technology (b).

For classic white-light endoscopies, several studies have shown that automatic image analysis can be successfully employed to *detect* colorectal polyps in order to assist physicians to decrease the polyp miss rate by detecting image regions that may contain polyps within the colon [8], [9]. Such detection can be performed by analyzing the polyp appearance generally based on color, shape, texture or spatial features applied to the video frames [10], [11], [12]. Colonic polyps may present different aspects of color, shape and texture depending on the

way they are captured by the camera, being influenced, for example, by the viewing angle, the distance from the capturing camera or even by the colon insufflation as well as the degree of colon muscular contraction [11].

Besides that, automatic polyp *classification*, e.g. based on the so-called pit pattern scheme [13], can help in diagnosing tumorous lesions once suspicious areas have been detected [2], [14], [3]. In this paper we also focus on classification and aim to differentiate polyps into two classes: normal mucosa or hyperplastic polyps (class healthy) and neoplastic, adenomatous or carcinomatous structures (class abnormal) as can be seen in Fig 2 (a-d). The different types of pit patterns [13] of these two classes can be observed in Fig. 2 (e-f) [7]. However, the classification can be a difficult task due to several factors such as the lack or excess of illumination, the blurring due to movement or water injection and the appearance of polyps [14], [11].

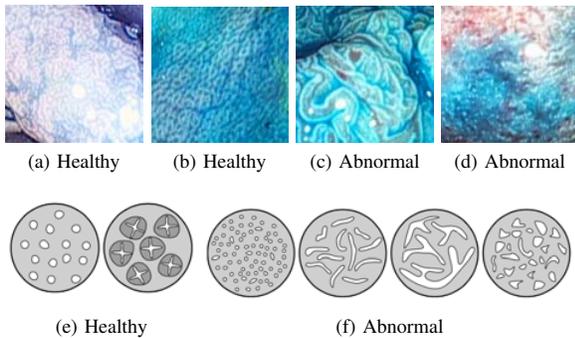


Fig. 2: Example images of the two classes (a-d) and the pit-pattern types of these two classes (e-f).

In the literature, existing computer-aided diagnosis techniques generally make use of feature extraction methods of color, shape and texture in combination with machine learning classifiers to perform the classification of colon polyps [15], [16]. Convolution Neural Networks are a promising methodology to help to improve these tasks.

Convolution Neural Networks (CNN's) have been demonstrated to be effective for discriminative pattern recognition in big data and in real-world problems mainly to learn both the global and local structures of images [17]. More recently, CNN were also tested for Computer-aided diagnosis systems such as the analysis, segmentation and prediction of knee cartilage as well as feature extraction from lung CT images [18]. The main advantage of this approach is that the same method can be used for the extraction of strong features that are invariant to distortion and position at the same time of the image classification. The intrinsic feature extractor is formed during the CNN training adapting to the context of the database. Finally, the neural network classifier can make use of these inputs to delineate more accurate hyperplanes helping the generalization of the network. However, one of the problems in the application of this approach is that the deep

layers of the CNN work best with structures based on edges, lines and curves, originating from object detection, however most medical databases have more texture-like images having no distinct structures of exactly these types. Another concern is the limitation of the availability of annotated images from medical image databases, since to avoid overfitting a large number of images is necessary to be available during the network training. In this work, we use smaller subimages and some strategies such as Dropout and ReLU activation functions to minimize this problem.

II. METHODOLOGY

We use an architecture of Convolutional Neural Network based on [17] to show that is possible to use this approach to also classify colonic polyp images. The network will need some modification to allow texture pattern recognition. Fig. 3 shows an illustration of the Convolutional Neural Network used in one of the experiments of this work.

A CNN is very similar to traditional Neural Networks in the sense of being constructed by neurons with their respective weights, biases and activation functions. As in Neural Networks, each neuron receives a series of inputs (representing dendrites) which are weighted and summed by the output neurons (representing a neuron's axon). In the case of CNN's, convolutional layers form the first levels (usually with a subsampling step) followed by one or more fully-connected neural networks similar to the multilayer neural networks [19].

In this work, the CNN input is a $(m \times m \times d)$ image (or patch) where $(m \times m)$ is the dimension of the patch and d the number of channels (depth) of the image, in the case of this work: the 3 RGB channel, $d = 3$. The convolutional layer consists of k learnable filters (also called kernels) with size $(n \times n \times d)$ where $(n \leq m)$. Such filters are convolved throughout the image by the product between the inputs and the filter resulting in a new output matrix. Convolving all the k filters and stacking these matrices will form the output volume also called activation maps or feature maps.

In addition, in the convolution step a padding in the input volume is used with zeros (zero padding) to control the spatial volume of output maps as it is appropriate to preserve the exact size of the original inputs. Besides, the stride of the convolution along the spatial dimension has to be specified: the larger the stride, the smaller the overlapping, decreasing the output volume dimensions.

After the convolution, a pooling layer is included to subsample the image by average functions (mean) or max-pooling over regions of size $(p \times p)$. These functions are used to reduce the dimensionality of the data in the following layers (upper layers) and to provide a form of invariance to translation thus making over-fitting control.

One of the most used activation functions in the CNN's and also used in this work is the ReLU rectifier function $f(x) = \max(0, x)$ where x is the neuron input that is demonstrably more efficient than other activation functions [20]. This function accelerates the convergence of the stochastic

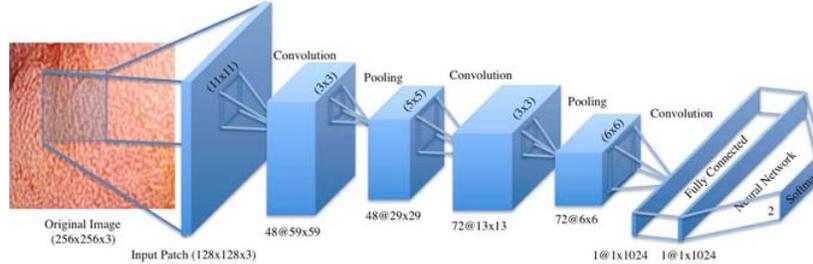


Fig. 3: An illustration of the CNN architecture for colonic polyp classification (CNN-05).

gradient descent algorithm mainly because of its non-linear and unsaturated characteristics.

An alternative to prevent overfitting in large neural networks also used in this work is the Dropout approach [21]. The Dropout disables (drops) feature detector nodes that are weak in the hidden layers of the network during the training forward pass. This is done to reduce interdependence between nodes simulating the training of many large networks with different connections in each iteration [21].

At the end of CNN there is a fully connected layer as a regular Multilayer Neural Network with the activation functions and its offset bias. The activation function used in this part is the Softmax function that generates a well-formed probability distribution on the outputs.

III. EXPERIMENTAL SETUP AND RESULTS

Due to the limitation of colonic polyp images to train a good CAD system, the main elements of the proposed method are: (1) extracting and preprocessing images in order to have a database with a suitable size (2) the use of CNN's for feature learning and good generalization, (3) the use of methods to avoid overfitting in the training phase.

For the evaluation tests we use a colonic polyp image database consisting of 100 images of size 256×256 from 62 patients using a high-definition (HD) endoscope (Pentax HiLINE HD+ 90i Colonoscope) with i-Scan mode 1 without chromoscopy (staining the mucosa) [6], [7], [22]. These images were extracted from HD video frame regions having histological findings, thus polyp detection is covered in this stage of data preparation. Despite the fact the frames being high-definition, the image size was chosen (i) to be large enough to describe a polyp and (ii) small enough to cover just one class of mucosa type (only healthy or only abnormal area). The database consists of two classes containing 25 healthy images from 18 patients and 75 abnormal images from 56 patients. Some patients may appear in both classes considering that different types of lesions or healthy tissues may be established inside the colon of a single patient. The videos were acquired during colonoscopy sessions between the years 2011 and 2013 at the Department for Internal Medicine (St. Elisabeth Hospital, Vienna).

Usually, some simple preprocessing techniques are necessary for the image feature generation. In this work we apply the normalization by subtracting the mean and dividing by the

standard deviation of its elements as in [23] corresponding to local brightness and normalization contrast. We also perform data augmentation by flipping each original image horizontally and vertically, and rotating the original image 90° for the right and left. Besides that, we flipped horizontally the rotated images, then we flipped vertically the horizontally flipped image, totalizing 7 new samples for each original image. After the data augmentation (resulting in 800 images), we randomly extract 75 subimages from each healthy image and 25 subimages from each abnormal image for the training set.

In this work we propose to extract subimages of size 128×128 from the original images. We explored the hypothesis that the colonic polyp classification with the CNN can be done only with a part of the image, and then we trained the network with smaller subimages instead of the entire image. This helps to reduce the size of the network, reducing its complexity and can allow different polyp classifications in the same image using different subimages in different parts of the image. Additionally, choosing smaller regions in a textured image can diminish the degree of intra-image variances in the dataset as the neighborhood is limited.

The CNN proposed by this work to satisfy the requirements cited in the beginning of this section is presented in Fig. 3 and consists of the following layers, parameters and configuration.

- Input Layer: subimages from the original image, of size $128 \times 128 \times 3$.
- Two combinations of convolutional and pooling layers: first convolutional layer consisting of 48 filters of size 11×11 and second convolutional layer consisting of 72 filters of size 5×5 . Both layers have padding 0 and stride set to 2 being followed by a ReLU rectifier function. After each convolutional layer there is a max-pooling layer consisting of windows with size 3×3 and stride set to 2;
- One convolutional layer to map the feature maps to the fully-connected output layer consisting of 1024 filters of size 6×6 .
- Fully-connected output layer: consists of a neural network with a hidden layer (with 1024 neurons) and a Softmax output layer depending on the number of the classes (in this case, two classes). Also, the Dropout method was used to regularize the two last fully-connected layers.

These hyperparameters were selected based on the works

TABLE I: Accuracy results from different CNN configurations for inputs of size $128 \times 128 \times 3$.

Network Index	No. of Convolutional Filters/Size			Connected Layer	Acc
	Layer 1	Layer 2	Layer 3		
CNN-01	48/7x7	72/4x4	512/5x5	512	76%
CNN-02	48/11x11	72/5x5	512/6x6	512	84%
CNN-03	24/11x11	48/5x5	1024/6x6	1024	86%
CNN-04	24/11x11	72/4x4	2048/5x5	2048	80%
CNN-05	48/11x11	72/5x5	1024/6x6	1024	87%

TABLE II: CNN configuration for input subimages of size $227 \times 227 \times 3$ and its respective accuracy.

Size of Inputs	No. of Convolutional Filters/Size				Connected Layer
	Layer 1	Layer 2	Layer 3	Layer 4	
227x227 x3	96/11x11	256/5x5	384/3x3	384/3x3	4096
	Layer 5	Layer 6	Layer 7	Layer 8	
	256/3x3	384/3x3	384/3x3	4096/6x6	
Accuracy: 79%					

[19] and [23] that investigated the impact of filter sizes likewise the number of filters in classification and consider this a satisfactory architecture. Also, empirical adjustment tests in the architecture such as changing the size and number of filters as well as the number of units in the fully connected layer were made and are shown in Table I. In this case, to compare the 5 different architectures in a faster way compared to the final experiments, we used cross validation evaluation with 10 different CNN's for each architecture. In nine of them, we removed 56 patients for training and used 6 for tests and, in one of them, we removed 54 patients for training and used 8 for test. The accuracy result given for each architecture is the average accuracy from each of the 10 CNN's trained. It can be seen that the architecture CNN-05 (described previously) obtained the best results, therefore, chosen to perform the subsequent tests.

We also tested a CNN architecture to be trained with bigger subimages ($227 \times 227 \times 3$) with the same cross-validation as for the results in Table I. The CNN configuration can be seen in Table II and it can be concluded that the accuracy result was not satisfactory (79%). This can be explained by the fact that neural networks involving a large number of inputs require a great amount of computation in training, requiring more data to avoid overfitting (which is not available given the size of our dataset).

For the subsequent experiments, with CNN-05 configuration, we trained one CNN for each patient from the database assuring that there are no images from patients of the validation set in the training set and configuring what we call leave-one-patient-out (LOPO) cross validation as in [24] to make sure the CNN's classifier generalizes to unseen patients. We choose the LOPO instead the classical leave-one-out cross validation (LOOCV) to try avoid overfitting in the training database at the same time that reduce the number of training networks (62 patients instead of 100 images). This cross-validation was also used in the methods used to compare from the literature.

TABLE III: Accuracy of different strides for overlapping subimages in the CNN-05 evaluation.

Stride	No. of Subimages	Accuracy
1	16384	90.22%
5	676	90.22%
20	49	90.21%
32	25	90.96%
48	9	89.27%
Random	16	90.31%
Random	32	90.65%
Random	64	90.49%

Specifically, the results from the CNNs presented in Tables III and IV are the mean values of the validation set from 62 different CNN's, one for each patient, implemented using the MatConvNet framework [25].

After training the CNN, in the evaluation phase, the final decision for a 256×256 pixel image from the dataset is obtained by majority voting of the decisions of all 128×128 pixel subimages (patches). One of the advantages of this approach is the opportunity to have a set of decisions available to acquire the final decision for one image. Also, the redundancy of overlapping subimages can increase the system accuracy likewise to give the assurance of certainty for the overall decision. As it can be seen in Table III, first we tested with a stride of 1 extracting the maximum number of 128×128 subimages available, totalizing 16384 subimages for each image, resulting in an accuracy of 90.22%. This evaluation is very computationally expensive to perform, so we decided to evaluate with different strides resulting in different number of subimages as it is shown in Table III. We also perform a random patch extraction and it can be concluded that there is not much difference between 16384 subimages or just 32 subimages (accuracy of 90,96%), saving considerable computation time and achieving good results.

In this work, we evaluated the CNN approach comparing with the results obtained by the following state-of-the-art feature extraction methods for the classification of colonic polyps [26]:

- **(BFD)** The blob-adapted Local Fractal Dimension algorithm [22] is based on computing the local fractal dimension with filters adapted to the shapes and sizes of the connected components (blobs).
- **(SSF)** The Blob Shape and Contrast algorithm [7] is a method that analyzes the shape of the blob.
- **(DT-CTW)** The Dual-Tree Complex Wavelet Transform is a multi-scale and multi-orientation wavelet transform. The means and standard deviations are extracted as features from the subband coefficients [3].
- **(MB-LBP)** In the Multi-Scale Block Local Binary Pattern approach [27], the LBP computation is done based on average values of block subregions. This approach is used for several image processing tasks including endoscopic polyp detection and classification [16].
- **(SIFT)** The Dense SIFT Features incorporates the bag-of-visual-words (BoW) method to the SIFT features [5].

The visual words are the cluster centers from the k-means clustering applied to the means of the SIFT descriptors.

- **(VASC-F)** The Vascularization Features represent the shape, contrast, size and underlying color of connected components (blood vessels) [15]. These vessel structures on polyps are segmented by means of the phase symmetry filter.

As the focus of several of the original publications was the feature extraction, all the previously cited feature extraction algorithms were evaluated using a k -NN classifier to allow comparison wrt. discriminativeness of the features [22], [7]. In order to stay consistent to the results published, the results of the feature extraction methods presented in Table IV are the mean values of the 10 results of the k -NN classifier (k -values $k = 1 - 10$) also using the leave-one-patient-out cross (LOPO) validation.

Experiment 1 from Table IV shows our best result using overlapped subimages with stride of 32 resulting in 25 subimages for each image in the evaluation tests compared to the feature extraction methods applied to the original images of size 256×256 . The results demonstrated that our proposed method has a superior performance (90.96%) to the feature extraction methods generally used for colonic polyp image classification. In Experiment 2 from Table IV we also applied the feature extraction methods to overlapped 128×128 pixel subimages with stride of 32 (25 subimages) using majority voting in the final classification as in the CNN evaluation. It can be seen that the results do not exhibit a significant change and our method still outperforms all other feature extraction methods. Some of the reasons for this surpassing result may be the use of three RGB bands from the original image by the CNN instead gray-scale images used by the presented feature extraction methods and the use of k -NN classifier instead of the SVM classifier. Table IV also shows the statistical significance of our results using the McNemar test [28] for the Experiment 1. In this case, number 1 indicates that the CNN is significantly different from the method (with significance level $\alpha = 0.05$). As we can see, the DT-CWT and the SIFT approach are classifying images significantly different to the CNN. However, the McNemar test is highly dependent of the database size [26], which may explain the “no significant differences” between the CNN and the other approaches.

The detailed classification results for the CNN evaluation result with stride of 32 (25 subimages) can be consulted in the confusion matrix displayed in Table V. It is also presented its respective Sensitivity (SE) and Specificity (SP) to delineate the CNN’s ability to correctly identify the polyps. The confusion matrix represents the mean of the normalized 62 confusion matrices obtained by the LOPO evaluation with 62 patients.

From the confusion matrix presented in Table V it can be concluded that, the classification accuracy was 90.96% while the sensitivity was 95.16% which represents a quite positive result since it meant that most of the abnormal polyp images were genuinely classified as such. Besides that, there is a reduced score for false negatives which is relevant for

TABLE IV: The classification results comparing our proposed method with feature extraction algorithms used for colonic polyp classification.

Methods	Acc. Exp. 1	Acc. Exp. 2	Sig.
BFD [22]	87.80%	87.00%	0
SSF [7]	84.70%	85.00%	0
DT-CWT [3]	83.90%	81.00%	1
MB-LBP [16]	82.90%	86.00%	0
SIFT [5]	82.00%	89.00%	0
VASC-F [15]	73.00%	62.00%	1
CNN	90.96%	90.96%	0

TABLE V: Confusion Matrix associated with CNN Colonic Polyp Classification.

		Prediction Outcome		total
		p	n	
Actual Value	p'	True Positive 47.2	False Negative 2.4	P' = 49.6
	n'	False Positive 3.2	True Negative 9.2	N' = 12.4
total		P = 50.35	N = 11.64	
		SE = 95.16%	SP = 74.19%	

this type of application concerning to be cautious with non-detected disorders. In contrast, the specificity score (SP) was lower than the sensibility with 74.19% meaning that the false positive rate was high. It can be explained by the fact that the number of negative samples was quite low comparing to the positive images for the CNN training. In future work, we intend to decrease this false positive percentage by increasing the training database. Even so, in general, the results were very effective.



Fig. 4: Filters from the first convolutional layer visualized as small image patches.

The weight matrices in the convolutional layer represent sets of features learned by the network (filters). These features from the first convolution layer of our trained network are presented in Fig. 4. It can be seen that the network has learned a collection of frequency and orientation-selective kernels, as well as many colored blobs intrinsic to the colonic polyp patterns. Some of them are like Laplacian/Gaussian filters, some are like edge detectors at different directions and others like texture extractors. Based on this observation, it can be inferred that the shape, color and texture information has been learned by the network as good discriminative features to

distinguish the mucosal texture of the colonic polyp image patches. Significant visual features should be captured by these filters for being directly connected to the input image source. Too small filters or too few filters may not capture all the features and generate poor feature maps for the subsequent layers, however, too big or too much filters require a large number of data to improve the accuracy of classification.

IV. CONCLUSION

In this paper, we propose the use of Convolutional Neural Networks (CNN's) to improve the accuracy of colonic polyp classification. This method has the advantage of combining image patches to enlarge the training database, increasing the data volume and consequently the information to perform the deep learning, by the fact that databases containing large amounts of annotated data are often limited for this type of research. The CNN's also use all the intrinsic features of the images such as color, shape and texture, by sharing the filter weights generating strong and representative features that are invariant to local distortions and translations. Different architectures were tested to evaluate the impact of the size and number of filters in the classification as well as the number of output units in the fully connected layer. Our method achieves superior performance compared to the state-of-the-art feature extraction techniques for colonic polyp classification. In future work, to enable even fairer comparison, we will use the outputs of the one-but last CNN layer as inputs into an SVM classifier, and apply an SVM classifier to the classically generated features as well.

ACKNOWLEDGMENT

This work has been partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

REFERENCES

- [1] M. Liedlgruber and A. Uhl, "Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review," *Biomedical Engineering, IEEE Reviews in*, vol. 4, pp. 73–88, 2011.
- [2] M. Häfner, L. Brunauer, H. Payer, R. Resch, A. Gangl, A. Uhl, F. Wrba, and A. Vécsei, "Computer-aided classification of zoom-endoscopic images using fourier filters," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 4, pp. 958–970, July 2010.
- [3] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei, "Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *Pattern Analysis and Applications*, vol. 12, no. 4, pp. 407–413, 2009.
- [4] M. Ganz, Xiaoyun Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 8, pp. 2144–2151, Aug 2012.
- [5] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-aided colorectal tumor classification in {NBI} endoscopy using local features," *Medical Image Analysis*, vol. 17, no. 1, pp. 78 – 100, 2013.
- [6] M. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in hd colonoscopy," in *Medical Computer Vision. Large Data in Medical Imaging*, vol. 8331 of *Lecture Notes in Computer Science*, pp. 205–213. Springer International Publishing, 2014.
- [7] M. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in hd colonoscopy," in *MICCAI*, vol. 8331, pp. 205–213. Springer International Publishing, 2014.
- [8] 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, Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).
- [9] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: Near real-time feedback during colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 3, pp. 164 – 179, 2015.
- [10] Sun Young Park, D. Sargent, I. Spofford, K.G. Vosburgh, and Y. A-Rahim, "A colon video analysis framework for polyp detection," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 5, pp. 1408–1418, May 2012.
- [11] Yi W., W. Tavanapong, J. Wong, J. Oh, and P.C. de Groen, "Part-based multidirectional edge cross-sectional profiles for polyp detection in colonoscopy," *Biomedical and Health Informatics, IEEE Journal of*, vol. 18, no. 4, pp. 1379–1389, July 2014.
- [12] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin 2009*, Informatik aktuell, pp. 346–350. Springer Berlin Heidelberg, 2009.
- [13] Kudo S, Hirota S, and Nakajima T, "Colorectal tumours and pit pattern," *Journal of Clinical Pathology*, vol. 10, pp. 880–885, Oct 1994.
- [14] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 565–581, 2012.
- [15] S. Gross, S. Palm, J. Tischendorf, A. Behrens, C. Trautwein, and T. Aach, "Automated classification of colon polyps in endoscopic image data," *SPIE*, vol. 8315, pp. 83150W–83150W–8, 2012.
- [16] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Color treatment in endoscopic image classification using multi-scale local color vector patterns," *Medical Image Analysis*, vol. 16, no. 1, pp. 75 – 86, 2012.
- [17] Alex K., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc., 2012.
- [18] Q. Li, W. Cai, X. Wang, Y. Zhou, D.D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *ICARCV, 2014*, Dec 2014, pp. 844–848.
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*, 2001, pp. 306–351, IEEE Press.
- [20] T. N. Sainath, B. Kingsbury, A. Mohamed, G. E. Dahl, G.e Saon, H. Soltau, T. Beran, A. Y. Aravkin, and B. Ramabhadran, "Improvements to deep convolutional neural networks for lvcsr," in *ASRU, 2013 IEEE Workshop on*, 2013, pp. 315–320.
- [21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [22] A. Uhl, G. Wimmer, and M. Häfner, "Shape and size adapted local fractal dimension for the classification of polyps in hd colonoscopy," in *ICIP, 2014*, Oct 2014, pp. 2299–2303.
- [23] A. Coates, H. Lee, and A.Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," in *AISTATS*, 2011, vol. 15 of *JMLR*, pp. 215–223, JMLR W&CP.
- [24] M. Häfner, M. Liedlgruber, S. Maimone, A. Uhl, A. Vécsei, and F. Wrba, "Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, June 2012, pp. 1–6.
- [25] A. Vedaldi and K. Lenc, "Matconvnet - convolutional neural networks for MATLAB," *CoRR*, vol. abs/1412.4564, 2014.
- [26] G. Wimmer, T. Tamaki, J.J.W. Tischendorf, M. Häfner, S. Yoshida, S. Tanaka, and A. Uhl, "Directional wavelet based features for colonic polyp classification," *Medical Image Analysis*, vol. 31, pp. 16 – 36, 2016.
- [27] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 828–837. Springer Berlin Heidelberg, 2007.
- [28] Quinn McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157.