Transfer Learning for Colonic Polyp Classification Using Off-the-Shelf CNN Features

Eduardo Ribeiro[1,2(✉)], Andreas Uhl[1], Georg Wimmer[1], and Michael Häfner[3]

[1] Department of Computer Sciences, University of Salzburg, Salzburg, Austria
uft.eduardo@uft.edu.br,
uhl@cosy.sbg.ac.at,
gwimmer@cosy.sbg.ac.at
[2] Department of Computer Sciences, Federal University of Tocantins, Palmas, Tocantins, Brazil
[3] St. Elisabeth Hospital, Vienna, Austria
michael.haefner@elisabethinen-wien.at
http://www.wavelab.at

Abstract. Recently, a great development in image recognition has been achieved, especially by the availability of large and annotated databases and the application of Deep Learning on these data. Convolutional Neural Networks (CNN's) can be used to enable the extraction of highly representative features among the network layers filtering, selecting and using these features in the last fully connected layers for pattern classification. However, CNN training for automatic medical image classification still provides a challenge due to the lack of large and publicly available annotated databases. In this work, we evaluate and analyze the use of CNN's as a general feature descriptor doing transfer learning to generate "off-the-shelf" CNN's features for the colonic polyp classification task. The good results obtained by off-the-shelf CNN's features in many different databases suggest that features learned from CNN with natural images can be highly relevant for colonic polyp classification.

Keywords: Deep learning · Convolutional Neural Networks · Colonic polyp classification

1 Introduction

The leading cause of deaths related to intestinal tract is the development of cancer cells (polyps) in its many parts. An early detection (when the cancer is still at an early stage) can reduce the risk of mortality among these patients. More specifically, colonic polyps (benign tumors or growths which arise on the inner colon surface) have a high occurrence and are known to be precursors of colon cancer development. As a consequence, it is recommended that everyone over an

E. Ribeiro—This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

T. Peters et al. (Eds.): CARE 2016, LNCS 10170, pp. 1–13, 2017.
DOI: 10.1007/978-3-319-54057-3_1

age of 50 years be examined regularly [32]. This exam can be done through an endoscopy procedure that is a minimally invasive and relatively painless diagnostic medical procedure that enables specialists to obtain images of internal human body cavities.

Several studies have shown that automatic detection of image regions which may contain polyps within the colon can be used to assist specialists in order to decrease the polyp miss rate [3,28,31]. Such detection can be performed by analyzing the polyp appearance that is generally based on color, shape, texture or spatial features applied to the video frames denoted as polyp detection [1,21,30].

Subsequently, the polyps can be automatically classified using different aspects of shape, color or texture into hyperplastic, adenomatous and malignant. The so-called "pit-pattern" scheme proposed by Kudo et al. [18] can help in diagnosing tumorous lesions once suspicious areas have been detected. In this scheme, the mucosal surface of the colon can be classified into 5 different types designating the size, shape and distribution of the pit structure [6,9,12]. These five pit-pattern types can allow to group the lesions into two main classes: normal mucosa or hyperplastic polyps (healthy class) and neoplastic, adenomatous or carcinomatous structures (abnormal class) as can be seen in Fig. 1(a–d). This approach is quite relevant in clinical practice as shown in a study by Kato et al. [17].

In this work we focus on the polyp classification into these two classes. The different types of pit patterns [18] of these two classes can be observed in Fig. 1(e–f) [14]. 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 different appearances of polyps [32]. Also, to find a robust and a global feature extractor that summarizes and represents all these pit-patterns structures in a single vector is very difficult and Deep Learning can be a good alternative to surpass these problems.

Deep learning Neural Networks have been of great interest in recent years, mainly due to the new variations of so-called Convolutional Neural Networks

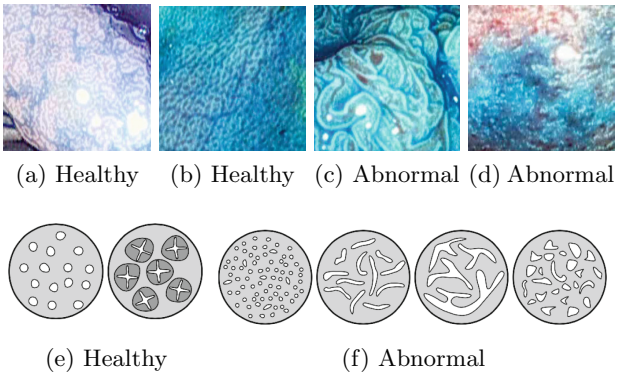

Fig. 1. Example images of the two classes (a–d) and the pit-pattern types of these two classes (e–f).

and the use of efficient parallel solvers improved by GPU's [2]. Deep learning is closely related to the high-level representation obtained by raw data such as images and is very effective when applied to large and annotated databases. However, the lack of available annotated medical image databases big enough to properly train a CNN is still a problem [2]. The use of transfer learning by pre-trained CNN's can help avoid this problem, however the existing available pre-trained CNN's are trained with natural images with very different features from the texture-like mucosa patterns in the colonic polyp images.

In this paper, we explore the use of Convolutional Neural Networks (CNN's) pre-trained with natural images to use them as medical imaging feature extractors, specifically of rectal colon images for colonic polyps classification. Rather than directly train a CNN with medical images, we apply a simple transfer method using pre-trained Convolutional Neural Networks. The assumption is that the patterns learned in the original database can be used in colonoscopy images for colonic polyp classification. In particular, we explore 11 different architectures (from 5000 to 160 million parameters) and depths (different numbers of layers), describing and analyzing the effects of pre-trained CNN's in different acquisition modes of colonoscopy images (8 different databases). This study was motivated by recent studies in computer vision addressing the emerging technique of transfer learning using pre-trained CNN's presented in the next section.

2 CNN's in Medical Image Classification

In recent years there has been an increased interest in machine learning techniques that is based not on hand-engineered feature extractors but using raw data to learn the representations.

This type of model has been very successful in large annotated databases, such as ImageNet [16] dataset that contains around 1.2 million images divided into 1000 categories. For these tasks, it is common to have a large number of parameters (in order of millions), requiring a significant amount of processing power to train the Neural Network. The CNN's can learn through their numerous layers and millions of connections if they are trained with sufficient examples, which becomes a significant difficulty in the medical area [8]. This problem occurs because of the lack of large, annotated and publicly available medical image databases such as the existing natural image databases, so that is a difficult and costly task to acquire and annotate such images and due to the specific nature of different medical imaging modalities which seems to have different properties according to each modality [15].

Some current pattern recognition techniques set aside handcrafted feature extraction algorithms to feed a Deep Learning Neural Network directly with raw data simultaneously acting as features extractor and image classifier at the same time [8,23]. These networks use many consecutive convolutional layers followed by pooling layers that reduce the data dimensionality making it, concomitantly, invariant to geometric transformations. Such convolution filters (kernels) are built to act as feature extractors during the training process and recent research

indicates that a satisfactorily trained CNN with a large database can perform properly when it is applied to other databases, which can mean that the kernels can turn into a universal feature extractor [23].

The works of Raza et al. [23] and Oquab et al. [20] suggest that the use of CNN's intermediate layer outputs can be used as input features to train other classifiers (such as support vector machines) for a number of other applications different from the original CNN obtaining a good performance. In fact, despite the difference between natural and medical images, some feature descriptors designed especially for natural images are used successfully in medical image detection and classification, for example: texture-based polyp detection [1], Fourier and Wavelet filters for colon classification [32], shape descriptors [14], local fractal dimension [13] for colonic polyp classification etc. In light of this, transfer learning that is a method used to harness the knowledge obtained by another task can be a good option to represent these kind of features.

Recently, works addressing the use of deep learning techniques in endoscopic images and videos are explored in many different ways, for example, to classify digestive organs in wireless capsule endoscopy images [34], detect lesions of endoscopy images [33] and automatically detect polyps in colonoscopy videos [22,27]. Also, pre-trained CNN's have been successfully used in the identification and pathology of X-ray and computer tomography modalities [8]. However, the application of transfer learning in endoscopic and colonoscopic images has not yet been exploited.

3 Materials and Methods

Using the inductive transfer learning, there are basically three types of strategies exploiting CNN's for medical image classification. Such strategies are described in the following and can be employed according to the intrinsic characteristics of each database [15].

When the available training database is large enough, diverse and very different from the database used in all the available pre-trained CNN's (in a case of transfer learning), the most appropriate approach would be to initialize the CNN weights randomly (training the **CNN from scratch**), and train it according to the medical image database for the kernels domain adaptation, that is, to find the best way to extract the features of the data in order to classify the images properly. This strategy, although ideal, is not widely used due to the lack of large and annotated medical image database publicly available for training the CNN.

Another alternative for large databases, but in this case, similar to a pre-trained CNN training database is the **CNN fine-tuning**. In fine-tuning the pre-trained network training continues with new entries (with a new database) for the weights to adjust properly to the new scenario reinforcing the more generic features with a lower probability of overfitting. This approach is also not widely applicable in case of medical image classification, again because of the limitation in the number of annotated medical images available for the appropriate network fine-tuning.

When the database is small, the best alternative is to use an **off-the-shelf CNN** [15]. In this case, using a pre-trained CNN, the last or next-to-last linear fully connected layer is removed and the remaining pre-trained CNN is used as a feature extractor to generate a feature vector for each input image from a different database. These feature vectors can be used to train a new classifier (such as an SVM) to classify the images correctly. If the original database is similar to the target database, the probability of the high-level features to describe the image correctly is high and relevant to this new database. If the target database is not so similar to the original, it can be more appropriate to use higher-level features, IE features from previous layers of CNN.

In this paper, we consider the knowledge transfer between natural images and medical images using off-the-shelf pre-trained CNN's. The CNN will project the target database samples into a vector space where the classes are more likely to be separable. This strategy was inspired by the work of Oquab et al. [20], which uses a pre-trained CNN in a large database (ImageNet) to classify images in a smaller database (Pascal VOC dataset) with improved results. Unlike that work, instead copy the weights of the original pre-trained CNN to the target CNN with additional layers, we use the pre-trained CNN to project data into a new feature space. This is done through the propagation of images from the colonic polyp database in the CNN, getting the resultant vector from the last CNN's layer and obtaining a new representation for each input sample. Subsequently, we use the feature vector set to train a linear classifier (for example support vector machines) in this representation to evaluate the results as used in [2,8].

To explore the use of different off-the-shelf CNN architectures for the computer-aided classification problem, we will describe below the elements to make the evaluation possible.

3.1 Data

The use of integrated endoscopic apparatus with high-resolution acquisition devices has been an important object of research in clinical decision support system area. With high-magnification colonoscopies is possible to acquire images up to 150-fold magnified, revealing the fine surface structure of the mucosa as well as small lesions. Recent work related to classification of colonic polyps used highly-detailed endoscopic images in combination with different technologies divided into three categories: high-definition endoscope (with or without staining the mucosa) combined with the i-Scan technology (1, 2, 3), high-magnification chromoendoscopy [9] and high-magnification endoscopy combined with narrow band imaging [7].

Specifically, the i-Scan technology (Pentax) used in this work is an image processing technology consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces [14].

There are three i-Scan modes available: i-Scan1, which includes surface enhancement and contrast enhancement, i-Scan2, that includes surface enhancement, contrast enhancement and tone enhancement and i-Scan3 that, besides

including surface, contrast and tone enhancement, also increases lighting emphasizing the features of vascular visualization [32]. In this work we use an endoscopic image database (CC-i-Scan Database) with 8 different imaging modalities acquired by an HD endoscope (Pentax HiLINE HD+ 90i Colonoscope) with images of size 256×256 from video frames either using the i-Scan technology or without any computer virtual chromoendoscopy (¬CVC). Table 1 shows the number of images and patient per class in the different i-Scan modes. The mucosa is either stained or not stained. Despite the fact the frames being high-definition originally, 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). Also, the image labels (ground truth) were provided according to their histological diagnosis.

Table 1. Number of images and patients per class of the CC-i-Scan databases gathered with and without CC (staining) and computed virtual chromoendoscopy (CVC).

i-Scan mode	No staining				Staining			
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3
Non-neoplastic								
Number of images	39	25	20	31	42	53	32	31
Number of patients	21	18	15	15	26	31	23	19
Neoplastic								
Number of images	73	75	69	71	68	73	62	54
Number of patients	55	56	55	55	52	55	52	47
Total nr. of images	112	100	89	102	110	126	94	85

3.2 Pre-trained Convolutional Neural Networks Architectures

We mainly explore six different CNN architectures trained to perform classification in the ImageNet ILSVRC challenge data. The input of all tested pre-trained CNN's has size $224 \times 224 \times 3$ and the descriptions as well as the details of each CNN are given as follows:

- The **CNN VGG-VD** [25] uses a large number of layers with very small filters (3×3) divided into two architectures according to the number of their layers. The CNN **VGG-VD16** has 16 convolution layers and five pooling layers while the CNN **VGG-VD19** has 19 convolution layers, adding one more convolutional layer in three last sequences of convolutional layers. The fully connected layers have 4096 neurons followed by a softmax classifier with 1000 neurons corresponding to the number of classes in the ILSVRC classification. All the layers are followed by a rectifier linear unit (ReLU) layer to induce the sparsity in the hidden units and reduce the gradient vanishing problem.
- The **CNN-F** (also called Fast CNN) [4] is similar the CNN used by Krizhevsky et al. [16] with 5 convolutional layers. The input image size is 224×224 and

the fast processing is granted by the stride of 4 pixels in the first convolutional layer. The fully connected layers also have 4096 neurons as the CNN VGG-VD. Besides the original implementation, in this work we also used the MatConvnet implementation (beta17, [29]) of this architecture trained with batch normalization and minor differences in its default hyperparameters and called here **CNN-F MCN**.

- The **CNN-M** architecture (medium CNN) [4] also has 5 convolutional layers and 3 pooling layers. The number of filters is higher than the Fast CNN: 96 instead of 64 filters in the first convolution layer with a smaller size. We also use the MatConvNet implementation called **CNN-M MCN**.
- The **CNN-S** (slow CNN) [4] is related to the "accurate" network from the Overfeat package [24] and also has smaller filters with a stride of 2 pixels in the first convolutional layer. We also use the MatConvNet implementation called **CNN-S MCN**.
- The **AlexNet** CNN [16] has five convolutional layers, three pooling layers (after layer 2 and 5) and two fully connected layers. This architecture is similar to the CNN-F, however, with more filters in the convolutional layers. We also use the MatConvNet implementation called **AlexNet MCN**.
- The **GoogleLeNet** [26] CNN has the deepest and most complex architecture among all the other networks presented here. With two convolutional layers, two pooling layers and nine modules also called "inception" layers, this network was designed to avoid patch-alignment issues introducing more sparsity in the inception modules. Each module consists of six convolution layers and one pooling layer concatenating these filters of different sizes and dimensions into a single new filter.

3.3 Experimental Setup

In order to form the feature vector using the pre-trained CNNs, all images are scaled using bicubic interpolation to the required size for each network, in the case of this work: $224 \times 224 \times 3$. The vectors obtained from the linear layers of the CNN have size: 1024×1 for the GoogleLeNet CNN and 4096×1 for the other networks due to their architecture specificities.

To allow the CNN features comparison and evaluation, we compared them with the results obtained by some state-of-the-art feature extraction methods for the classification of colonic polyps [32] which are: Blob Shape adapted Gradient using Local Fractal Dimension method (**BSAG-LFD** [13]), Blob Shape and Contrast (**Blob SC** [14]), Discrete Shearlet Transform using the Weibull distribution (**Shearlet-Weibull** [5]), Gabor Wavelet Transform (**GWT Weibull** [32]), Local Color Vector Patterns (**LCVP** [11]) and Multi-Scale Block Local Binary Pattern (**MB-LBP** [11]). All these feature extraction methods (with the exception of BSAG-LFD) were applied to the three RGB channels to form the final feature vector space.

For the classical features, the classification accuracy is also computed using a SVM classifier however, with the original images (without resizing) trained using the Leave-One-Patient-out cross validation strategy as in [10] to make

sure the classifier generalizes to unseen patients. This cross-validation is applied to the methods from the literature as well as to off-the-shelf CNN's features. The accuracy measure based on the percentage of images correctly classified in each class is used to allow an easy comparability of the results due to the high number of methods and databases to be compared.

4 Results and Discussion

The accuracy results for the colonic polyp classification in the 8 different databases are reported in Table 2. As can be seen, the results in Table 2 are divided into two groups: off-the-shelf features and concatenating them with state-of-the-art features.

Among the 11 pre-trained CNN investigated, the CNN that presents lower performance were GoogleLeNet, CNN-S and AlexNet MCN. These results may indicate that such networks themselves are not sufficient to be considered off-the-shelf feature extractors for the polyp classification task.

Table 2. Accuracies of the methods for the CC-i-Scan databases in %.

Methods	No staining				Staining				
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	$\overline{X}$
1- CNN-F	86.16	89.33	80.65	88.41	86.52	81.40	84.22	80.62	84.66
2- CNN-M	87.45	90.67	81.38	83.58	87.99	89.55	87.40	90.53	87.31
3- CNN-S	88.03	90.00	87.01	77.33	87.25	82.68	87.40	75.54	84.41
4- CNN-F MCN	88.84	82.00	73.15	90.73	85.78	89.55	89.72	83.15	85.36
5- CNN-M MCN	89.53	90.67	88.88	94.66	86.97	89.29	87.40	90.53	**89.74**
6- CNN-S MCN	90.12	91.42	81.38	79.85	89.18	93.49	81.10	84.77	86.41
7- GoogleLeNet	79.65	90.67	72.43	74.51	88.27	80.46	75.60	84.08	80.70
8- VGG-VD16	87.45	85.33	86.38	79.65	92.47	89.80	95.26	92.38	88.59
9- VGG-VD19	83.49	82.67	83.88	87.71	92.47	83.98	94.46	85.59	86.78
10-AlexNet	91.40	87.33	75.65	89.32	87.71	83.03	84.22	79.24	84.73
11-AlexNet MCN	89.42	84.67	78.88	83.78	89.36	83.55	81.10	78.32	83.63
$\overline{X}$	87.41	87.70	80.88	84.50	88.54	86.07	86.17	84.06	85.67
13- Blob SC	77.67	83.33	82.10	75.22	59.28	78.83	66.13	59.83	72.79
14- Shearlet-Weibull	73.72	76.67	79.60	86.80	81.30	69.91	72.38	83.63	78.00
15- GWT-Weibull	79.75	78.67	70.25	84.28	81.30	74.54	77.17	83.39	78.66
16- LCVP	76.60	66.00	47.75	77.12	77.45	79.00	70.01	69.56	70.43
17- MB-LBP	78.26	80.67	81.38	83.37	69.29	70.60	77.22	78.32	77.38
$\overline{X}$	78.71	78.70	74.28	81.61	73.13	75.58	73.61	74.35	76.24
Concatenating 5/8	88.84	85.33	83.88	92.14	93.12	90.49	96.88	94.00	90.58
Concatenating 5/12	92.79	92.67	88.88	96.98	87.71	90.49	88.26	90.53	91.03
Concatenating 5/8/12	95.94	90.00	88.88	92.14	92.30	91.43	97.63	97.46	93.22
Concatenating 5/8/14	91.51	88.67	87.10	93.75	94.68	91.43	98.44	95.85	92.67
Concatenating 5/8/15	90.91	90.00	88.88	92.14	93.94	89.80	96.88	95.61	92.27
Concatenating 5/8/12/14	93.38	88.00	91.38	93.75	93.49	92.12	97.63	94.92	93.08
Concatenating 5/8/12/17	93.38	90.00	91.38	93.75	92.75	92.12	97.63	97.46	**93.55**

As it can be seen, the pre-trained CNN that presents the best result on average for the different imaging modalities ($\overline{X}$) is the CNN-M network trained with the MatConvNet parameters (89.74%) followed by the CNN VGG-VD16 (88.59%). These deep models with smaller filters generalize well with other datasets as it shown in [25], including texture recognition, which can explain the better results in the colonic polyp database. However, there is a high variability in the results and thus it is difficult to draw general conclusions.

Many results obtained by the pre-trained CNN's surpassed the classic feature extractors for colonic polyp classification in the literature. The database that presents the best results using off-the-shelf features is the database staining the mucosa without any i-Scan technology (88.54% on average). In the case of classical features, the database with the best result in the average is the database using the i-Scan3 technology without staining the mucosa (81.61%).

To investigate this difference in the results we asses the significance of them using the McNemar test [19]. By means of this test, we analyze if the images from a database are classified differently or similarly by the other methods. With a high accuracy it is suppose of that the methods will have a very similar response, so the significance level α must be small enough to differentiate between classifying an image as correct or incorrect.

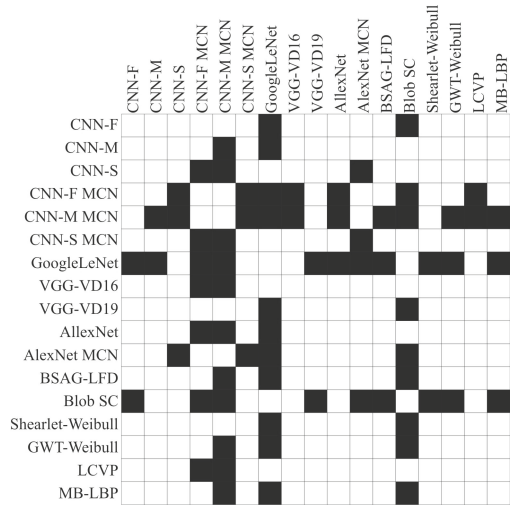

Fig. 2. Results of the McNemar test for the i-Scan3 database without staining. A black square in the matrix means that the methods are significantly different with significance level $\alpha = 0.01$. If the square is white then there is no significant difference between the methods.

The test is carried out on the database that presents the best results with the classic features (i-Scan3 without staining the mucosa) using significance level $\alpha = 0.01$. The results are presented in Fig. 2. It can be observed by the black

squares that, among the pre-trained CNN's, the CNN-M MCN and GoogleLeNet present the most different results comparing to the other CNN's.

Also, in Fig. 2 when comparing the classical feature extraction methods with the CNN's features it can be seen that there is a quite different response among the results, especially for CNN-M MCN that is significantly different from all the classical methods with the exception of the Shearlet-Weilbull method.

The methods with high accuracy are not found to be significantly different which can indicate that, in these methods, almost the same images are classified wrong, independent of the extracted features.

Observing the features that are significantly different in Fig. 2 and with good results in Table 2 we decided to concatenate the feature vectors to see if the features can complement each other. It can be seen also in Table 2 that the two most successful CNN's (CNN-M MCN and VGG-VD16) are significantly different from each other and, at the same time, the CNN-M MCN is significantly different to BSAG-LFD features which, among the classical results, presents the best results.

Based on this difference, the three feature vectors (CNN-M, CNN-M MCN and BSAG-LFD) were concatenated and the results presents a high accuracy on average: 93.22%. When we add to the vector one more classical feature (MB-LBP) that is also significantly different to CNN-M MCN, the result outperforms all the previous approaches: 93.55%.

5 Conclusion

In this paper, we explored and evaluated several different pre-trained CNN's architectures to extract features from colonoscopy images by the knowledge transfer between natural and medical images providing what it is called off-the-shelf CNNs features. We show that the off-the shelf features may be well suited for the automatic classification of colon polyps even with a limited amount of data.

The different used CNNs were pre-trained with an image domain completely different from the proposed task. Apparently the 4096 features extracted from CNN-M MCN and VGG-16 provided a good and generic extractor of colonic polyps features. Some reasons for the success of the classification include the training with a large range of different images, providing a powerful extractor joining the intrinsic features from the images such as color, texture and shape in the same architecture, reducing and abstracting these features in just one vector.

Also, the combination of classical features with off-the-shelf features yields good prediction results complementing each other. We believe that this strategy could be used in other endoscopic databases such as automatic classification of celiac disease. Besides that, this approach will be explored in future work to also detect polyps in video frames and the performance in real time applications will be evaluated. It can be concluded that Deep Learning through Convolutional Neural Networks is becoming essentially the most favorite candidate in almost all pattern recognition tasks.

References

1. Ameling, S., Wirth, S., Paulus, D., Lacey, G., Vilarino, F.: Texture-based polyp detection in colonoscopy. In: Meinzer, H.-P., Deserno, T.M., Handels, H., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2009. Informatik aktuell, pp. 346–350. Springer, Heidelberg (2009)
2. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H.: Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 294–297, April 2015
3. Bernal, J., Schez, J., Vilario, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognit. **45**(9), 3166–3182 (2012). Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2011)
4. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: British Machine Vision Conference, BMVC 2014, Nottingham, 1–5 September 2014
5. Dong, Y., Tao, D., Li, X., Ma, J., Pu, J.: Texture classification and retrieval using shearlets and linear regression. IEEE Trans. Cybern. **45**(3), 358–369 (2015)
6. Ribeiro E., Uhl, A., Häfner, M.: Colonic polyp classification with convolutional neural networks. In: 2016 29th International Symposium on Computer-Based Medical Systems (CBMS), June 2016
7. Ganz, M., Yang, X., Slabaugh, G.: Automatic segmentation of polyps in colonoscopic narrow-band imaging data. IEEE Trans. Biomed. Eng. **59**(8), 2144–2151 (2012)
8. Ginneken, B., Setio, A., Jacobs, C., Ciompi, F.: Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: 12th IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, 16–19 April 2015, pp. 286–289 (2015)
9. Häfner, M., Kwitt, R., Uhl, A., Gangl, A., Wrba, F., Vécsei, A.: Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. Pattern Anal. Appl. **12**(4), 407–413 (2009)
10. Häfner, M., Liedlgruber, M., Maimone, S., Uhl, A., Vécsei, A., Wrba, F.: Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps. In: 2012 25th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6, June 2012
11. Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Color treatment in endoscopic image classification using multi-scale local color vector patterns. Med. Image Anal. **16**(1), 75–86 (2012)
12. Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy. Comput. Methods Programs Biomed. **107**(3), 565–581 (2012)
13. Häfner, M., Tamaki, T., Tanaka, S., Uhl, A., Wimmer, G., Yoshida, S.: Local fractal dimension based approaches for colonic polyp classification. Med. Image Anal. **26**(1), 92–107 (2015)
14. Häfner, M., Uhl, A., Wimmer, G.: A novel shape feature descriptor for the classification of polyps in HD colonoscopy. In: Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Tu, Z. (eds.) MCV 2013. LNCS, vol. 8331, pp. 205–213. Springer, Heidelberg (2014). doi:10.1007/978-3-319-05530-5_20

15. Shin, H., Roth, H., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. CoRR, abs/1602.03409 (2016)
16. Alex K., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105. Curran Associates Inc. (2012)
17. Kato, S., Fu, K.I., Sano, Y., Fujii, T., Saito, Y., Matsuda, T., Koba, I., Yoshida, S., Fujimori, T.: Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions. World J. Gastroenterol. **12**(9), 1416–1420 (2006)
18. Kudo, S., Hirota, S., Nakajima, T.: Colorectal tumours and pit pattern. J. Clin. Pathol. **10**, 880–885 (1994)
19. McNemar, Q.: Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika **12**(2), 153–157 (1947)
20. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, 23–28 June 2014, pp. 1717–1724 (2014)
21. Sun, Y.P., Sargent, D., Spofford, I., Vosburgh, K.G., A-Rahim, Y.: A colon video analysis framework for polyp detection. IEEE Trans. Biomed. Eng. **59**(5), 1408–1418 (2012)
22. Park, S.Y., Sargent, D.: Colonoscopic polyp detection using convolutional neural networks. In: Proceedings of SPIE, vol. 9785, p. 978528 (2016)
23. Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, 23–28 June 2014, pp. 512–519 (2014)
24. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: integrated recognition, localization and detection using convolutional networks. CoRR, abs/1312.6229 (2013)
25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
26. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition (CVPR) (2015)
27. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 79–83, April 2015
28. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans. Med. Imaging **35**(2), 630–644 (2016)
29. Vedaldi, A., Lenc, K.: Matconvnet - convolutional neural networks for MATLAB. CoRR, abs/1412.4564 (2014)
30. Yi, W., Tavanapong, W., Wong, J., Oh, J., de Groen, P.C.: Part-based multi-derivative edge cross-sectional profiles for polyp detection in colonoscopy. IEEE J. Biomed. Health Inform. **18**(4), 1379–1389 (2014)
31. Wang, Y., Tavanapong, W., Wong, J., Oh, J.H., de Groen, P.C.: Polyp-alert: near real-time feedback during colonoscopy. Comput. Methods Programs Biomed. **120**(3), 164–179 (2015)

32. Wimmer, G., Tamaki, T., Tischendorf, J.J.W., Häfner, M., Yoshida, S., Tanaka, S., Uhl, A.: Directional wavelet based features for colonic polyp classification. Med. Image Anal. **31**, 16–36 (2016)
33. Zhu, R., Zhang, R., Xue, D.: Lesion detection of endoscopy images based on convolutional neural network features. In: 2015 8th International Congress on Image and Signal Processing (CISP), pp. 372–376, October 2015
34. Zou, Y., Li, L., Wang, Y., Yu, J., Li, Y., Deng, W.J.: Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1274–1278, July 2015