Study to integrate CNN inside a WCE to realize a screening tool

Orlando Chuquimia, Thomas Garbay, Weiqin Xu, Andrea Pinna, Xavier Dray, Hichem Sahbi, Bertrand Granado

Study to integrate CNN inside a WCE to realize a screening tool

Orlando Chuquimia[1], Thomas Garbay[1], Weiqin Xu[1], Andrea Pinna[1], Xavier Dray[2], Hichem Sabhi[1], Bertrand Granado[1]

***Abstract*— Screening is a method to improve the early detection of colorectal cancer. Now, screening is based on an immunochemical test that look for blood in faecal samples, but image is the best modality to detect the marker of colorectal cancer : polyps. In 2003 Wireless Capsule Endoscopy was introduced and opened a way to integrate automatic image processing to realize a screening tool. In parallel Convolutionnal Neural Networks have demonstrated their high capacity to detect polyps in many scientific studies, but fail to be integrable. In this article we present our works to integrate CNN or image processing based on a CNN inside a WCE to realize a powerful screening tool.**

I. INTRODUCTION

Colorectal cancer (CRC) is the second highest cause of death by cancer worldwide with a mortality rate of 47.62% corresponding to 880,792 deaths in 2018 [18], [1]. Consequently it is a major health problem. 95% of CRC begin as a growth on the inner lining of the colon or rectum called as polyp [7] and it is treatable in 90% of the cases if it is detected earlier before polyp become adenocarcinomas [2]. One of the key solutions recommended by the European Code Against Cancer to cure or greatly improve outcomes is early detection through screening of gastrointestinal (GI) tract [21].

Today, image is the modality to analyze the colon and find polyps. The colonoscopy is the tool for screening, diagnosis and therapy in the gastrointestinal tract. However, it is a painful examination, often traumatic and poorly tolerated by patients. The colonoscopy is invasive and need an anaesthesia, a specialist and a controlled environment. Furthermore, the colonoscopy doesn't allow the visibility of all the regions near the colon. There exist other methods as the colorectal tomography (CTC) that are non invasive method, but it cannot detect polyps less than 1cm and they expose the patient to a radiation exposure [12].

In many countries, the screening process starts with a test that Fecal Occult Blood Test (FOBT) or a Fecal Immunochemical Test (FIT). These tests are used to determine if it is necessary to realize a colonoscopy. FOBT has a low sensitivity of only 38% [3]. For the FIT test sensitivity varies, a study showed a variation from 89% for a FIT calibrated to detect less than 20 g/g of blood to 70% if it is calibrated to detect 20 to 50 g/g of blood [14]. Decreasing the sensitivity of FIT test increase its specificity and reduce the number of useless colonoscopies. Thus, it is prefered to increase the specificity and decrease the sensitivity of the FIT test and realize a periodic screening.

As we show there is a need for a screening tool with a high sensitivity and a high specificity. In 2003 Paul Swain and al. [18] introduced Wireless Capsule Endoscopy (WCE), a simple pill that patient swallows and that transmits images of the gastrointestinal tract via a Radio Frequency communication through the body. More than 1.6 million patients worldwide have used this technology for the small bowel, esophagus and colon (more than 125,000 procedures a year). The available WCEs for the colon [12], like the PillCam Colon 2, has a length of 31 mm and diameter of 11 mm, a battery life of 10 hours, a resolution of 256x256 pixels and an image sampling rate around 2 to 4 frames per second.

TABLE I
COMMERCIALLY AVAILABLE WCE FOR COLON [12]

Manufacturer	PillCam COLON	PillCam COLON2	MedTronics (Given Imaging)
Size[mm](Length x diameter)	31x11	31x11	31x11
Battery [h]	10	10	10
Image Resolution[pixels]	256x256	256x256	256x256
Image Sampling rate [fps]	4	4-35	4 -35

Our idea is to integrate inside a WCE an intelligent image processing that can detect a polyp. The question is what is the best image processing ?

As in many domain Convolutional Neural Network (CNN) demonstrate its capacity to detect polyp lesions.

In [15] a two section CNN is proposed, a downsampling path that contain convolutional and max-pooling layers and a upsampling path with convolutional and upsampling layers. Auxiliary classifiers are injected to train the network to reduce the problem of vanishing gradients and accelerate the back-propagation of gradient flow. Multilevel contextual features are classified in the last fully connected layer. They have performed their experiments in a dataset with three classes: Normal, Blood, Polyp and Ulcer-Erosion with 400, 50, 50 and 100 frames respectively extracted from five WCE videos. The CNN has trained with the 50% of frames of each class and the rest to evaluate the classification. Their CNN obtains an average accuracy over 85%.

In [11], AlexNet model [13] pre-trained on the ILSVRC 2012 dataset [17] was used and modified to input patches of size 96x96. Furthermore, the kernel size, of the two

[1] LIP6, CNRS UMR 7606, Sorbonne Université, Paris, France.
[2] APHP, Hôpital Saint-Antoine, Sorbonne Université, Paris, France.

first pooling layers is decreased from 3 to 2 and the last pooling layer is removed to modify the output layer for two outputs (polyp or non-polyp). They increased the number of examples applying random mirroring, rotation, up- and down-scaling, cropping, and brightness adjustment in the original database. They use a sliding-window strategy to determine the polyp presence or absence in a video sequence. They evaluate their CNN performance in a dataset of 120 frames (60 with a polyp), the use 80 images (40 with a polyp) to train and the rest to test, their experimentation has shown an accuracy of 60%.

In [25], three CNNs trained at different image scales (x1, x0.5, x0.25), the last fully connected layer is removed, the outputs of the convolutional layer of each CNN are fed as input to a single Multi Layer Perceptron (MLP) network that is trained separately. The training was performed exclusively on the CVC-CLINIC(Computer Vision Center/Universitat Autonoma de Barcelona and Hospital Clinic from Barcelona, Spain) database composed by 1200 WCE images achieving an accurate classification until 90%.

In [22] a 3D CNN is used for polyp detection by leveraging spatial-temporal information from colonoscopy videos using a single output probability map. The 3D-CCN is capable of learning more representative spatio-temporal features from colonoscopy videos. They have evaluated their method on the ASUMayo Clinic Polyp Database [4] of MICCAI 2015 Challenge on Polyp Detection. Their 3D CNN has shown a recall until 71% and a precision of 88%.

In [24] a novel regression-based CNN pipeline is presented for polyp detection during colonoscopy. Their pipeline is composed of 2 parts: the first part uses the ResYOLO model pre-trained, with a large non-medical image database and further fine-tuned with colonoscopy images, in the second part, temporal information was incorporated via a tracker named Efficient Convolution Operators (ECO) for refining the detection results given by ResYOLO. They have evaluated their CNN on 17,574 frames extracted from 18 endoscopic videos of the AsuMayoDB. their experimentation has shown a precision of 88.6%, recall of 71.6%.

In [23] a rotation invariant and image similarity constrained Densely Connected Convolutional Network (RIIS-DenseNet) model is proposed. Their rotation-invariant regularization constraint is then introduced to explicitly enforce learned features. Their method achieves an accuracy 95.62% for polyp detection on a WCE dataset.

Despite CNN achieve very good performance on polyp detection, all these methods are running on an external computer and contribute to help the physician in his diagnosis, but they are not useable for our purpose, because they do not consider WCE constraint as real time execution, form factor of the pill, energy consumption. In particular CNN methods use a high number of synapses and neurons, more than 4 millions, that is not implantable in a 5x5 mm^2 chip inside a pill.

How it is possible to use CNN such a powerful tool to embed an adapted processing inside iWCE ?

The purpose of this article is to describe our work to integrate CNN inside a WCE. We present our study based on a Deep Learning algorithm, a Convolutional Neural Network (CNN) to identify a more powerful image processing method, a convolution kernel dedicated to polyp localization. We propose a method to extract the convolution kernels from a trained CNN. Then we evaluate the performance of this method of the entire chain and finally us make a conclusion.

II. The CNN

We first fine tuned a trained CNN, the GoogleNet [19], to classify images in two classes : with polyps and without polyp. The architecture of this CNN is shown in Figure 1. The model we have chosen is the InceptionV3 proposed by Google [19], this CNN was the winner of the 2014 ILSVRC Contest with less number of synapses and neurons in comparison with Alexnet, VGGnet, Lenet or Resnet.

We adapt its last layer, the fully-connected layer, with a structure with two outputs dedicated to determine the presence of a polyp.

To train this network, we use a 11952 image database divided into 10023 images with polyp and 1929 images without polyp. These images are issued from 18 videos of endoscopic examinations of Hospital Clinic, Barcelona, Spain. This dataset was used as a train dataset for EndoVisSub2017-GIANA contest [4]. Each image is associated with a ground-truth : a binary-image that indicates the position of the polyp in the image.

We use 70% of the data to train our CNN and the remaining 30% to test this performance. The result shows an accuracy of more than 98%. Although we have achieved very good accuracy, it is impossible to embed this convolutional neural network with four million parameters into a capsule.

Two ways can be taken to find how it is possible to use such a powerful tool to embed an adapted processing inside the iWCE:

- Reduce the size of the network
- Use the CNN to find the best image processing before the classification in our processing chain.

In our work we take the second direction : we analyze the CNN to find the best image processing to localize polyps before the classification. We propose a method to extract the convolution kernels from a trained CNN.

III. Proposed method

A CNN is a structure in which multiple layers of image filters are connected in series. Each layer of an image filter consists of several convolution kernels connecting two adjacent layers of neurons. This structure is inspired by the connection of human brain neurons. The output of each neuron is a feature map. By analyzing these feature maps we want to find valuable convolution kernels which help to locate potential polyp areas.

The goal is to construct an image processing inspired from CNN, that means we will train a CNN to classify the images at first, then make a visual analysis of the result of each kernel from the first layers to the still interpretable layers.

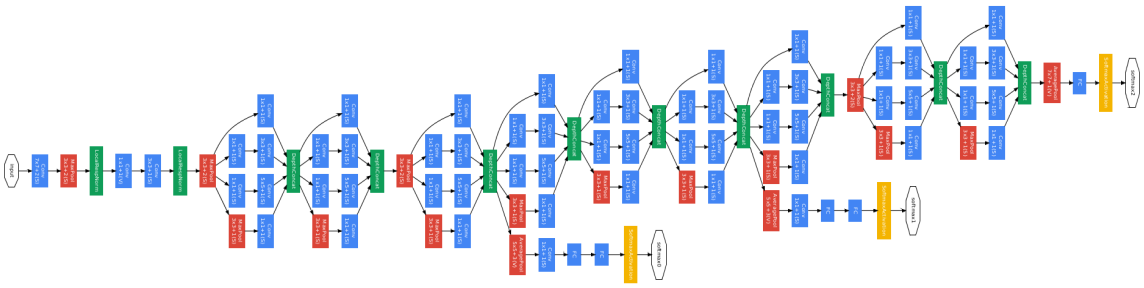

Fig. 1. The Googlenet CNN

Still interpretable means that we can visualize in the feature maps the result of the processing.

A. *Analyzing the feature maps*

To analyze the kernels we look at the output of each layer. The Figure 2 shows an example of the input and the first layer outputs of our network. We concentrate our analysis for the 5 first layers that are still interpretable. In addition the outputs of the first two layers will be more favorable to their integration inside the iWCE because they require less calculations. If we look the Figure 2, we see the original image in its three components, red green blue, and the first convolution layer of our network. Because there are too many outputs for the first layer, we just show a part of them.

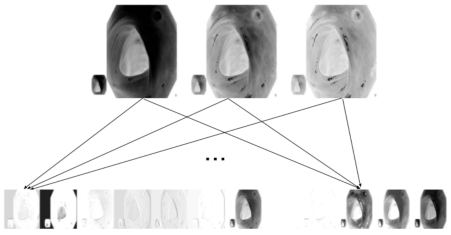

Fig. 2. Firsts Layer of the Deep Learning

By looking feature map by feature map we have identified a kernel that realizes an interesting processing, visible in the Figure 3. After the processing by this kernel the area where the polyp is located has been darkened. This means that this kernel could be effective to locate polyps.

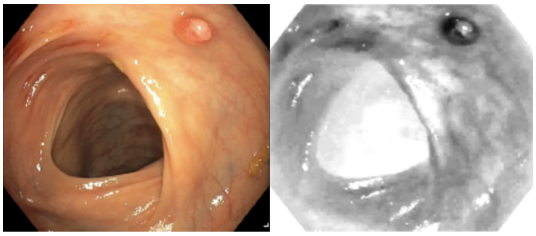

Fig. 3. Input image and the output of convolutional kernel number 13

In addition, to remove the background and the noise, we added a sigmoid-like function $f(x) = \frac{1}{1+e^{-\frac{x-a}{b}}}$ after the output. We use this function only to demonstrate the efficiency of the kernel, it will not be used in the final system. We can see the result in the Figure 4, we take $a = 60$ and $b = 12.8$.

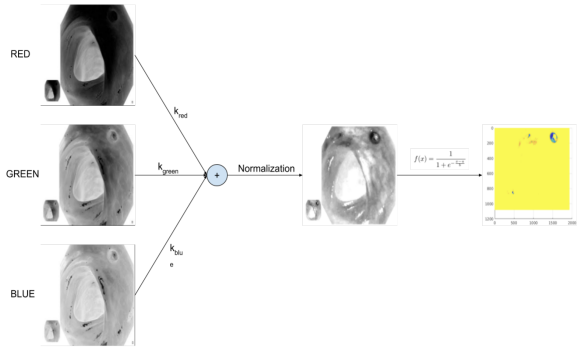

Fig. 4. Input and first layer outputs

Now we have a complete filter that clearly shows the area of the polyp. We used this filter to test several other pictures. The results are visible in Figure 5 and validate the chosen kernel.

B. *Extracting Convolution Kernels and Building a Complete Detection System*

We have extracted the contents of this convolution kernel, there is one kernel by color component and the matrix are resumed in equations 1, 2 and 3.

$$k_{red} = \begin{bmatrix} 0.05635 & 0.16901 & 0.08185 \\ 0.14792 & 0.28381 & 0.14082 \\ 0.10782 & 0.17279 & 0.12224 \end{bmatrix} \quad (1)$$

$$k_{green} = \begin{bmatrix} -0.11226 & -0.17383 & -0.08947 \\ -0.20939 & -0.30939 & -0.20798 \\ -0.11434 & -0.20847 & -0.10023 \end{bmatrix} \quad (2)$$

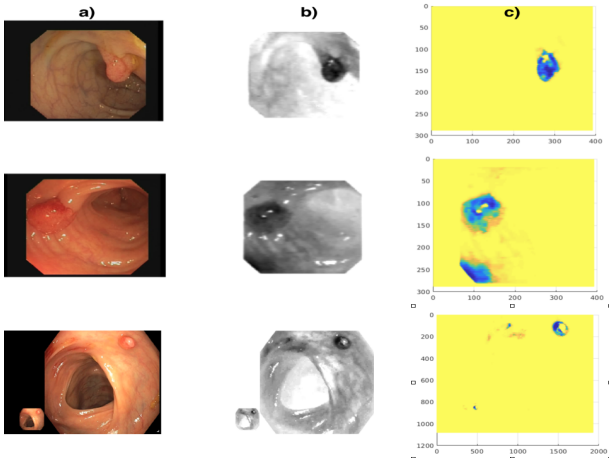

Fig. 5. a) original input image, b) feature map after the convolution kernel processing and c) result after the sigmoid function processing

$$k_{blue} = \begin{bmatrix} 0.01974 & 0.06515 & -0.00436 \\ 0.04877 & 0.05371 & -0.00284 \\ 0.01671 & 0.01576 & -0.01776 \end{bmatrix} \quad (3)$$

In the next section, we include this Kernel to a processing chain as a hybrid method to detect polyps. We consider the output of this kernel as a color model.

IV. Processing Chain and Results

Following, we give a short description of the whole chain through three steps, see figure 6 :

1) **ROI extraction step:** in this step, we realize a median noise filter on a color model image and an edge detection using standard Canny filter [6], after, we identify the ROI as the image parts containing circular/elliptical edges. We use Hough Transform algorithm [10] to do so. We consider two color models: brightness model and Convolutional Kernel model. The brightness model that preserves the texture information, allows an efficient texture analysis and has a good degree of integration in a system on chip than an image color. The Convolutional Kernel model, as a result of applying one of three convolutional kernels on each RGB color space, the addition of the outputs produce this color model.
2) **ROI description step:** in this step we realize a texture analysis of the ROI and extract 26 texture and luminosity descriptors using co-occurrence matrix algorithm [8].
3) **ROI classification step:** in the classification step we construct a fuzzy-forest using fuzzy-trees to classify each ROI, see [16] for details. An ROI containing a polyp is labeled as $class1$ and an ROI without polyp is classified as $class0$. At image level, an image containing at least one ROI of $class1$ is classified as $class1$ otherwise is classified as $class0$.

The evaluation of these metrics was performed on two annotated polyp datasets and is shown in table II. The first

Fig. 6. Proposed system scheme of polyps detection.

dataset [20] is composed by 10 videos-colonoscopies that display a unique polyp at multiple scales and from different viewing angles, in total 5402 images of which 3857 with polyp. The second dataset [4] is composed by 18 videos of endoscopic examinations of Barcelona Hospital Clinic, Spain for EndoVisSub2017-GIANA challenge, in total 11954 images of which 10025 with a polyp. Each image of both datasets is associated with a ground-truth : a binary-image that indicates the position of the polyp in the image.

We test each color model alone and the aggregation of both. To evaluate this processing chain, we measure the detection rate of the global chain at two levels: ROI level and Image level.

Standard measures are employed:

- True Positive (TP): an ROI classified as $class1$ that contain a polyp or an image of $class1$ that contain a polyp.
- False Positive (FP): an ROI classified as $class1$ that not contain a polyp or an image of $class1$ that not contain a polyp.
- True Negative (TN): an ROI classified as $class0$ that not contain a polyp or an image of $class0$ that not contain a polyp.
- False Negative (FN): an ROI classified as $class0$ or an image of $class0$ that contain a polyp.

Using these measures, we compute the following aggregation metrics:

- $precision = TP/(TP + FP)$.
- $recall = TP/(TP + FN)$.

TABLE II

Color model	brightness model	Convolutional Kernel model	both
ROI-based			
Recall	29.1%	30.4%	29.7%
Specificity	90.6%	88.7%	89.9%
Image-based			
Recall	53.0%	45.3%	66.4%
Specificity	57.7%	59.4%	43.8%

We can notice that we obtain increasing from 53.0% to 66.4% in sensitivity when we aggregate the two color spaces. If we compare to the sensitivity of the FOBT test we obtain a higher sensitivity but with a lower specificity. Comparison with the FIT test show a similar sensitivity, but a lower specificity.

Armed with these positive conclusions that suggest the ability to approach, or even exceed, the performance of current immuno-chemical screening tests, we have initiated

a new path for integrating a CNN into a WCE. The idea is to compress an existing CNN, such as GoogleNet. For this we use a method called distillation. This technique, whose feasibility was first shown by the team of [5], which got on the resolution of eight problems, thanks to a compressed network, on average, a thousand times smaller and a thousand times faster, excellent results, losses due to compression being negligible. In addition, the [9] team tested this methodology on the MNIST database and achieved good results. This work shows that it is possible for a faster compressed network to be able to approximate the function learned previously by a larger and slower network, but intrinsically more efficient. With regard to the problem of the endoscopic video capsule for the detection of polyps, this approach is very promising to be able to integrate a convolutional neural network, smaller and faster than the networks of neurons of the state of the art.

V. Conclusion

Deep Learning is a powerful tool, but cannot be integrated in an iWCE. We have presented here an analysis of a CNN that identify specific kernel to realize the extraction feature of a classification chain. We have identified a kernel in the first layer of the CNN that are adapted to detect a polyp.

The goal of this work is to use a CNN to identify a powerful processing that satisfies embedding requirements of a WCE.

The next step of our work is first identify more complex Kernels in superior layers and analyze their integrability and second use distillation method to compress a powerfull CNN and integrate it inside a WCE.

References

[1] Cancer today. http://gco.iarc.fr/today/home, 2018.

[2] Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., de Haes, J. C. J. M., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., and Takeda, F. The European Organization for Research and Treatment of Cancer QLQ-C30: A Quality-of-Life Instrument for Use in International Clinical Trials in Oncology. *JNCI: Journal of the National Cancer Institute 85*, 5 (Mar. 1993), 365–376.

[3] Allison, J. E., Tekawa, I. S., Ransom, L. J., and Adrain, A. L. A comparison of fecal occult-blood tests for colorectal-cancer screening. *New England Journal of Medicine 334*, 3 (1996), 155–160. PMID: 8531970.

[4] Bernal, J. J., Histace, A., Masana, M., Angermann, Q., Sánchez-Montes, C., Rodriguez, C., Hammami, M., Garcia-Rodriguez, A., Córdova, H., Romain, O., Fernández-Esparrach, G., Dray, X., and Sanchez, J. Polyp Detection Benchmark in Colonoscopy Videos using GTCreator: A Novel Fully Configurable Tool for Easy and Fast Annotation of Image Databases. In *Proceedings of 32nd CARS Conference* (Berlin, Germany, June 2018).

[5] Bucilu, C., Caruana, R., and Niculescu-Mizil, A. Model Compression. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2006), KDD '06, ACM, pp. 535–541.

[6] Canny, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 6 (1986), 679–698.

[7] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer 136*, 5 (2015), E359–E386.

[8] Haralick, R. M. Statistical and structural approaches to texture. *Proceedings of the IEEE 67*, 5 (1979), 786–804.

[9] Hinton, G., Vinyals, O., and Dean, J. Distilling the Knowledge in a Neural Network. *arXiv:1503.02531 [cs, stat]* (Mar. 2015). arXiv: 1503.02531.

[10] Hough, P. V. Method and means for recognizing complex patterns, Dec. 18 1962. US Patent 3,069,654.

[11] Hwang, S. Bag Of Visual Words Approach based on SURF Features to Polyp Detection in Wireless Capsule Endoscopy Videos. 4.

[12] Karargyris, A., and Bourbakis, N. Detection of Small Bowel Polyps and Ulcers in Wireless Capsule Endoscopy Videos. *IEEE Transactions on Biomedical Engineering 58*, 10 (Oct. 2011), 2777–2786.

[13] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[14] Lee, J. K., Liles, E. G., Bent, S., Levin, T. R., and Corley, D. A. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. *Annals of internal medicine 160*, 3 (2014), 171–181.

[15] Nawarathna, R. D., Oh, J., Yuan, X., Lee, J., and Tang, S. J. Abnormal Image Detection Using Texton Method in Wireless Capsule Endoscopy Videos. In *Medical Biometrics* (2010), D. Zhang and M. Sonka, Eds., Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 153–162.

[16] Orlando, C., Andrea, P., Xavier, D., and Granado, B. Polyps recognition using fuzzy trees. In *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on* (2017), IEEE, pp. 9–12.

[17] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision 115*, 3 (Dec. 2015), 211–252.

[18] Swain, P. Wireless capsule endoscopy. *Gut 52*, suppl 4 (June 2003), iv48–iv50.

[19] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going Deeper With Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1–9.

[20] Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE Transactions on Medical Imaging 35*, 2 (Feb. 2016), 630–644.

[21] von Karsa, L., Patnick, J., and Segnan, N. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition – Executive summary. *Endoscopy 44*, S 3 (Sept. 2012), SE1–SE8.

[22] Yu, L., Chen, H., Dou, Q., Qin, J., and Heng, P. A. Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos. *IEEE Journal of Biomedical and Health Informatics 21*, 1 (Jan 2017), 65–75.

[23] Yuan, Y., Qin, W., Ibragimov, B., Han, B., and Xing, L. RIIS-DenseNet: Rotation-Invariant and Image Similarity Constrained Densely Connected Convolutional Network for Polyp Detection. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (2018), A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, Eds., Lecture Notes in Computer Science, Springer International Publishing, pp. 620–628.

[24] Zhang, R., Zheng, Y., Poon, C. C., Shen, D., and Lau, J. Y. Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition 83* (2018), 209 – 219.

[25] Zhao, Q., Dassopoulos, T., Mullin, G. E., Meng, M. Q., and Kumar, R. A decision fusion strategy for polyp detection in capsule endoscopy. *Studies in health technology and informatics 173* (2012), 559–565.