Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Deep Neural Networks approaches for detecting and classifying colorectal polyps

Alba Nogueira-Rodríguez [a,b,c], Rubén Domínguez-Carbajales [d], Hugo López-Fernández [a,b,c], Águeda Iglesias [d], Joaquín Cubiella [d], Florentino Fdez-Riverola [a,b,c], Miguel Reboiro-Jato [a,b,c], Daniel Glez-Peña [a,b,c,*]

[a] Department of Computer Science, University of Vigo, ESEI, Campus As Lagoas, 32004 Ourense, Spain
[b] The Biomedical Research Centre (CINBIO), Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain
[c] SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Spain
[d] Department of Gastroenterology, Complexo Hospitalario Universitario de Ourense, Instituto de Investigación Sanitaria Galicia Sur, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Ourense, Spain

ARTICLE INFO

Article history:
Received 5 November 2019
Revised 20 January 2020
Accepted 6 February 2020
Available online xxxx

Keywords:
Deep learning
Convolutional Neural Network (CNN)
Polyp detection
Polyp localization
Polyp classification
Computer-aided diagnosis

ABSTRACT

Deep Learning (DL) has attracted a lot of attention in the field of medical image analysis because of its higher performance in image classification when compared to previous state-of-the-art techniques. In addition, a recent meta-analysis found that the diagnostic performance of DL models is equivalent to that of health-care professionals. In this scenario, a lot of research using DL for polyp detection and classification have been published showing promising results in the last five years. Our work aims to review the most relevant studies from a technical point of view, focusing on the low-level details for the implementation of the DL models. To do so, this review analyzes the published research covering aspects like DL architectures, training strategies, data augmentation, transfer learning, or the features of the datasets used and their impact on the performance of the models. Additionally, comparative tables summarizing the main aspects analyzed in this review are publicly available at https://github.com/sing-group/deep-learning-colonoscopy.

1. Introduction

Colorectal carcinoma (CRC) is one of the most common malignancies in western countries. In Spain, it is the second most common cancer after prostate cancer in men, and breast cancer in women. An estimated 40,000 new cases of CRC are diagnosed every year in Spain and 39% of those affected die from this disease. Two different strategies have been proposed to reduce the burden of the disease: CRC screening and early diagnosis in symptomatic patients [1].

Colonoscopy plays a key role in any of the strategies aimed at reducing the incidence and mortality due to CRC. In this sense, the detection and endoscopic resection of the precursor lesions, mainly colorectal adenomas, produces a significant reduction of the CRC incidence [2]. However, diagnostic performance is limited and up to 3.7% of CRC are post colonoscopy or interval CRC: CRCs diagnosed in the three years after a normal colonoscopy [3]. One of the reasons beyond this problem is the limited capacity of the endoscopists to detect subtle lesions. In fact, the miss rate for adenomas is 26%, 9% for advanced adenomas and 27% for serrated polyps [4]. In order to improve the quality of colonoscopy, several indicators have been proposed. The adenoma detection rate (ADR) has become the most important quality indicator of screening colonoscopy because it is directly related with key outcome measures, such as interval cancer incidence and mortality.

In the last few years, a variety of optical technologies has been developed to assist endoscopists in the determination of the histological diagnosis. However, these technologies require: (*i*) the usage of magnification endoscopy equipment, (*ii*) the application of stains or pigments, and (*iii*) highly experienced endoscopists. To overcome these limitations, the international NICE (Narrow band imaging Colorectal Endoscopic classification) classification was released to be used together with Narrow Band Imaging (NBI) endoscopes, with or without magnification. As these techniques still require highly experienced endoscopists, there has been a growing interest in the development of Computer-Aided Diagnosis (CAD) systems for automatic polyp detection and prediction of the histology without polyp resection and biopsies.

* Corresponding author at: Despacho 306, Edificio Politécnico, Campus Universitario As Lagoas S/N, 32005 Ourense, Spain.
E-mail address: dgpena@uvigo.es (D. Glez-Peña).

https://doi.org/10.1016/j.neucom.2020.02.123
0925-2312/

The classification of images of colorectal tumors have been accomplished by CAD systems based on Machine Learning (ML). As an example of this, Tamaki et al. proposed a CAD system colorectal tumor classification in NBI endoscopy using local features [5], achieving an accuracy of 96% on a 10-fold cross validation using a dataset of 908 NBI images and 93% using an independent test dataset. It is important to note that these kinds of ML-based systems require a high expertise and an extensive preprocessing of the image datasets in order to extract the relevant features of the images that are used as input by the ML algorithms. This issue is extensible to any type of medical image analysis.

In this scenario, Deep Learning (DL) has attracted a lot of attention in the field of medical image analysis because of its higher performance in image classification when compared to previous state-of-the-art techniques [6,7]. Simply speaking, these models have the advantage that do not require a previous preprocessing of the image datasets and they can be trained using the raw images thanks to their ability to automatically extract and learn the relevant features. As a matter of fact, a recent review gathers more than three hundred studies on the field of medical image analysis based in DL techniques [8] and a recent meta-analysis found that the diagnostic performance of DL models is equivalent to that of health-care professionals [9]. The analysis of colonoscopic videos and images has not lagged far behind this hype and a significant number of studies on automatic polyp detection and classification have been carried out since 2014. Remarkably, some of the participants in the Automatic Polyp Detection sub-challenge conducted as part of the Endoscopic Vision Challenge at the international conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) in 2015 [10] already used DL approaches. Since then, a growing number of studies using DL to accomplish these tasks are published every year with promising results. Thus, DL is expected to aid endoscopists to improve their diagnosis performance by automatically detecting and classifying colorectal polyps [11]. Probably the most important limitation for the application of DL to medical image analysis and colonoscopic videos and images in particular is the need of large, annotated datasets of videos or images of the particular domain. As the creation of this high-quality datasets is difficult in the biomedical domain due to the high costs in terms of both the money and expertise required, strategies like data augmentation can help to deal with this issue.

Recently, Samy A. Azer reviewed the usage of DL for automatic polyp detection, focused on discussing the current challenges when dealing with CRC screening and training programs, quality measures in colonoscopy, and the role that CNNs play in increasing the detection rate of polyps and early cancerous lesions [12]. Similarly, Min et al. made a broader review about the application of DL in gastrointestinal endoscopy, putting emphasis on the effects of DL on gastroenterology with a special interest on automatic diagnosis based on endoscopic findings [13]. Both reviews agree that DL can help endoscopists in providing more accurate diagnosis and improve the current CRC detection rates. In order to complement these reviews, our work aims to review the most relevant studies using DL for polyp detection and classification from a technical point of view, focusing on the low-level details for the implementation of the DL models. Thus, this review analyzes the published research covering aspects like DL architectures, training strategies, data augmentation, transfer learning, or the features of the datasets used and their impact on the performance of the models, which is also discussed in a specific section. As a complementary contribution, a public repository with comparative tables summarizing the main aspects of the research works covered in this review has been published at https://github.com/sing-group/deep-learning-colonoscopy, an open repository that is expected to grow with information about future studies published in the same topic.

The rest of the paper is structured as follows. Section 2 identifies the main applications of DL in colonoscopies and characterizes the most relevant studies carried out in the field. Section 3 describes both public and private datasets used in the surveyed approaches. Section 4 analyzes the DL architectures while Section 5 compares their performance for polyp detection and classification tasks. Finally, Section 6 summarizes the main conclusions extracted from the state of the art.

2. Deep learning for polyp detection and classification

The research reviewed in this study applied different types of Deep Neural Networks on two well-differentiated tasks in colonoscopy: polyp detection (including localization) and polyp classification. Polyp detection refers to the task of trying to predict whether there are one (or more) polyps in a given video frame. Most of polyp detection studies also include polyp localization, where systems also try to predict the position of the polyp(s) in the image. Table 1 summarizes the studies focused on polyp detection (5/21) and localization (15/21). Most of the studies are based on conventional colonoscopy, although two of them use only wireless capsule endoscopy (WCE) and another two combine both colonoscopy types. For automatic polyp detection or localization systems to be useful in real clinical practice, a key feature is their ability to operate in real-time. In this regard, it is important to note that only 7 out of 21 studies report that they can perform real-time polyp detection. Thus, there is still room for improvement in this important aspect. Studies also vary in the type of localization reported. While most of them (12/21) predict the polyp localization as a bounding box, others predict them as binary masks (4/21). Finally, it is worth mentioning that 13 out of 21 studies can deal with the presence of multiple polyps.

On the other hand, polyp classification studies aim to distinguish between types of polyps and predict their histological classification (e.g. adenoma, hyperplastic, etc.). Table 2 summarizes the studies focused on different polyp classifications. In this case, all studies use conventional colonoscopy and there is heterogeneity regarding the classes or types of polyps that they try to differentiate. In this regard, only two of them are coinciding in distinguishing adenomatous from hyperplastic polyps. Lui et al. [37] goes beyond polyp classification and focus on larger endoscopic lesions with risk of submucosal invasion and lymphovascular permeation with the goal of distinguishing endoscopically curable lesions versus endoscopically incurable lesions. The remaining ones tried to distinguish neoplastic from non-neoplastic, adenoma from non-adenoma, hyperplastic vs. neoplastic and adenoma from hyperplastic from traditional serrated adenoma. Only one study is able to classify polyps in real time.

3. Datasets

The successful development of DL-based models for automatic polyp detection and classification requires the availability of big datasets of polyp images or videos along with high-quality, manual annotations provided by experts. These annotations provide the ground truth (i.e. the accurate polyp localizations in the frames or the associated histological classification) necessary to train the DL models. In general purpose object localization and classification, large datasets like ImageNet [43], with more than 14 million images hand-annotated in 20,000 categories, or Microsoft's COCO [44], with more than 2,500,000 images with object highlighting, labeling, and classification into 91 object types, are widely used to develop and improve DL models. However, in

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Table 1
Summary of studies focused on polyp detection and localization.

Study	Date	Endoscopy type	Imaging technology	Localization type	Multiple polyp	Real time
Tajbakhsh et al. 2014, 2015 [14,15]	Sept. 2014/Apr. 2015	Conventional	N/A	Bounding box	No	**Yes**
Zhu R. et al 2015 [16]	Oct. 2015	Conventional	N/A	Bounding box (16x16 patches)	**Yes**	No
Park and Sargent 2016 [17]	March 2016	Conventional	NBI, WL	Bounding box	No	No
Yu et al. 2017 [18]	Jan. 2017	Conventional	NBI, WL	Bounding box	No	No
Zhang R. et al. 2017 [19]	Jan. 2017	Conventional	NBI, WL	No	No	No
Yuan and Meng 2017 [20]	Feb. 2017	WCE	N/A	No	No	No
Brandao et al. 2018 [21]	Feb. 2018	Conventional/ WCE	N/A	Binary mask	**Yes**	No
Zhang R. et al. 2018 [22]	May 2018	Conventional	WL	Bounding box	No	No
Misawa et al. 2018 [23]	June 2018	Conventional	WL	No	**Yes**	No
Zheng Y. et al. 2018 [24]	July 2018	Conventional	NBI, WL	Bounding box	**Yes**	**Yes**
Shin Y. et al. 2018 [25]	July 2018	Conventional	WL	Bounding box	**Yes**	No
Urban et al. 2018 [26]	Sep. 2018	Conventional	NBI, WL	Bounding box	No	**Yes**
Wang et al. 2018 [27,28]	Oct. 2018	Conventional	N/A	Binary mask	**Yes**	**Yes**
Qadir et al. 2019 [29]	Apr. 2019	Conventional	NBI, WL	Bounding box	**Yes**	No
Blanes-Vidal et al. 2019 [30]	March 2019	WCE	N/A	Bounding box	**Yes**	No
Zhang X. et al. 2019 [31]	March 2019	Conventional	N/A	Bounding box	**Yes**	**Yes**
Misawa et al. 2019 [32]	June 2019	Conventional	N/A	No	**Yes**	No
Zhu X. et al. 2019 [33]	June 2019	Conventional	N/A	No	No	**Yes**
Ahmad et al. 2019 [34]	June 2019	Conventional	WL	Bounding box	**Yes**	**Yes**
Sornapudi et al. 2019 [35]	June 2019	Conventional/ WCE	N/A	Binary mask	**Yes**	No
Wittenberg et al. 2019 [36]	Sept. 2019	Conventional	WL	Binary mask	**Yes**	No

Table 2
Summary of studies focused on polyp classification.

Study	Date	Endoscopy type	Imaging technology	Classes	Real time
Ribeiro et al. 2016 [38]	Oct. 2016	Conventional	WL	Neoplastic vs. Non-neoplastic	No
Zhang R. et al. 2017 [19]	Jan. 2017	Conventional	NBI, WL	Adenoma vs. hyperplastic Resectable vs. non-resectable Adenoma vs. hyperplastic vs. serrated	No
Byrne et al. 2017 [39]	Oct. 2017	Conventional	NBI	Adenoma vs. hyperplastic	**Yes**
Komeda et al. 2017 [40]	Dec. 2017	Conventional	NBI, WL, Chromoendoscopy	Adenoma vs. non-adenoma	No
Chen et al. 2018 [41]	Feb. 2018	Conventional	NBI	Neoplastic vs. hyperplastic	No
Lui et al. 2019 [37]	Apr. 2019	Conventional	NBI, WL	Endoscopically curable lesions vs. endoscopically incurable lesion	No
Kandel et al. 2019 [42]	June 2019	Conventional	N/A	Adenoma vs. hyperplastic vs. traditional serrated adenoma	No

biomedical domains, it is usually difficult to create such big and high-quality due to the high costs in terms of both the money and expertise required. In the case of colonoscopy, some publicly available datasets for polyp detection and classification have been released in the last few years. Although these datasets, summarized in Table 3, are not as large as general-purpose datasets, they allowed the development of an important amount of studies. In particular, three datasets annotated for polyp detection and localization, namely CVC-ClinicDB [45], ETIS-Larib [46], and ASU-Mayo Clinic Colonoscopy Video [47], are very popular as they were used in the Automatic Polyp Detection sub-challenge at MICCAI 2015. While CVC-ClinicDB and ETIS-Larib datasets are composed of annotated frames, the ASU-Mayo Clinic dataset is composed of 38 fully annotated videos selected to show maximum variation in colonoscopy procedures. As it can be seen in Fig. 1, these are the publicly available datasets used by most of the reviewed studies. Authors of the original CVC-ClinicDB dataset extended their contribution with three more publicly available datasets: CVC-ColonDB [48,49], CVC-PolypHD [48,49], and CVC-ClinicVideoDB [50]. Regarding polyp classification, Mesejo et al. [51] contributed with a dataset of 76 short videos (both NBI and WL) with hyperplastic, adenoma, and serrated polyps. Although useful, these datasets are still not enough for the development of accurate, reliable DL-based models for polyp detection and classification. For this reason, 22 out of the 27 studies have collected specific and private datasets for conducting their research, which are summarized in Table 4. As Fig. 2 shows, 15 of these 22 studies only use private datasets, while 7 use both public and private datasets.

Regarding the annotation of the publicly available datasets for polyp detection, all of them provide the polyp locations as binary masks, that is, the ground truth is composed by the pixels where the polyp is present. The prediction of these binary masks can be directly addressed with DL models for image segmentation (e.g. Wang et al. 2018 [27]). Nevertheless, as noted in the previous section, several studies opted for predicting the polyp locations using bounding boxes and thus the ground truth of these datasets must be adapted in the first place. Zheng Y et al. converted them this way: "The minimal bounding box (x, y, w, h) that covers each binary mask was generated, where x, y, w and h denotes the coordinates of the centroid, width and height of the box, respectively. Then each bounding box was normalized to the image size." [24].

4. Architectures

In DL, the development of a model for classification or localization of objects in images is usually done by using Convolutional Neural Networks (CNN) architectures. This can be done by either using an off-the-shelf CNN implementation or by designing a new architecture. In the latter case, the new model is usually created by using an off-the-shelf DL model as a basis, although it is also possible to create completely new models by combining several custom convolutional layers and fully connected layers (FCL).

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Table 3
Summary of publicly available colonoscopy datasets.

Dataset	References	Description	Format	Resolution (w × h)	Ground truth	Used in
CVC-ClinicDB	Bernal et al. 2015 [45] https://polyp.grand-challenge.org/CVCClinicDB/	612 sequential WL images with polyps extracted from 31 sequences with 31 different polyps.	Image	388 × 284	Polyp locations (binary mask)	Brandao et al. 2018 [21], Zheng Y. et al. 2018 [24], Shin Y. et al. 2018 [25], Wang et al. 2018 [27], Qadir et al. 2019 [29], Sornapudi et al. 2019 [35], Wittenberg et al. 2019 [36]
CVC-ColonDB	Bernal et al. 2012 [48] Vázquez et al. 2017 [49]	380 sequential WL images with polyps extracted from 15 videos.	Image	574 × 500	Polyp locations (binary mask)	Tajbakhsh et al. 2015 [15], Brandao et al. 2018 [21], Zheng Y. et al. 2018 [24], Sornapudi et al. 2019 [35]
CVC-PolypHD	Bernal et al. 2012 [48] Vázquez et al. 2017 [49]	56 WL images.	Image	1920 × 1080	Polyp locations (binary mask)	Sornapudi et al. 2019 [35]
ETIS-Larib	Silva et al. 2014 [46] https://polyp.grand-challenge.org/EtisLarib/	196 WL images with polyps extracted from 34 sequences with 44 different polyps.	Image	1225 × 966	Polyp locations (binary mask)	Brandao et al. 2018 [21], Zheng Y. et al. 2018 [24], Shin Y. et al. 2018 [25], Ahmad et al. 2019 [34], Sornapudi et al. 2019 [35], Wittenberg et al. 2019 [36]
Kvasir-SEG	Pogorelov et al. 2017 [52] https://datasets.simula.no/kvasir-seg/	1 000 polyp images	Image	Various resolutions	Polyp locations (binary mask)	–
ASU-Mayo Clinic Colonoscopy Video	Tajbakhsh et al. 2016 [47] https://polyp.grand-challenge.org/AsuMayo/	38 small SD and HD video sequences: 20 training videos annotated with ground truth and 18 testing videos without ground truth annotations. WL and NBI.	Video	N/A	Polyp locations (binary mask)	Yu et al. 2017 [18], Brandao et al. 2018 [21], Zhang R. et al. 2018 [22], Ahmad et al. 2019 [34], Sornapudi et al. 2019 [35], Wittenberg et al. 2019 [36]
CVC-ClinicVideoDB	Angermann et al. 2017 [50]	18 SD videos.	Video	768 × 576	Polyp locations (binary mask)	Shin Y. et al. 2018 [25], Qadir et al. 2019 [29]
Colonoscopic Dataset	Mesejo et al. 2016 [51] http://www.depeca.uah.es/colonoscopy_dataset/	76 short videos (both NBI and WL).	Video	768 × 576	Polyp classification (Hyperplastic vs. adenoma vs. serrated)	Zhang R. et al. 2017 [19]

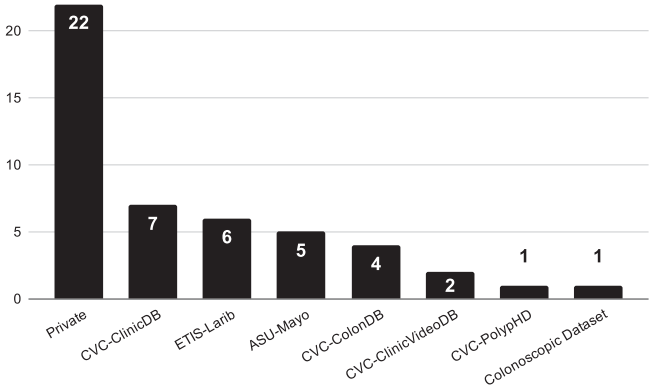

Fig. 1. Histogram showing the number of studies by dataset type.

Most of the DL networks are designed to perform classification tasks, where each neuron of the output layer assigns a likelihood to a particular outcome or label. Therefore, when performing localization tasks, where objects should be highlighted with a bounding box or binary mask, the networks should be adapted. The main way to do this is to employ frameworks, which can be divided into object detection frameworks, based on regions of interest (RoI), and segmentation frameworks, based on pixel-wise classification.

The object detection frameworks, such as the R-CNN family [53–55] or YOLO [56], provide a way to split the input image into several RoIs. The RoIs are fed to a DL model that determines if a region contains an object of a certain category. The positively selected RoIs are the output of the network or, in some cases, they are combined in a subsequent step to form the final bounding boxes or binary masks.

Regarding segmentation frameworks, such as Mask R-CNN [57] or SegNet [58], they generate pixel-wise predictions that can be used to localize and highlight objects in an image. In this case, there is not a common strategy among the different frameworks. Mask R-CNN, for example, is an extension of the Faster R-CNN network [55] that, among other changes, adds a CNN to it for mask prediction. On the other hand, SegNet is a deep encoder-decoder architecture, that uses VGG16 [59] as the encoder.

Using off-the-shelf DL models considerably simplifies the model development, as they can be found already implemented in the main DL libraries. These models can be adapted to new tasks by just replacing the original output layer with a new one with as many neurons as required. In addition, by using these models it is also possible to take advantage of the transfer learning (TL) technique, where a model developed for a task is reused for another task. This technique is very common in DL, where models trained in public image databases, such as ImageNet [43], Microsoft's COCO [44] or PASCAL VOC [60], are available in the main libraries to be reused for a different image classification task. In these cases, the initial weights of the models are taken from the pre-trained model, instead of using random values, as usual. Then, during a process called fine-tuning, images of the new classification task are presented to the network that adjusts the weights of some specific layers as in a regular training. The final FCL are usually fine-tuned, as they are responsible for the final classification, or, in some cases, these layers are replaced with a classifier, such as a Support Vector Machine (SVM), that learns from the features generated by the convolutional layers. Regarding the convolutional

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Table 4
Summary of private colonoscopy datasets used in different studies.

Study	Patients	No. Images	No. Videos	No. Unique Polyps	Purpose	Comments
Tajbakhsh et al. 2015 [15]	N/A	35 000 With polyps: 7 000 Without polyps: 28 000	40 short videos (20 positive and 20 negative)	N/A	Polyp localization	
Zhu R. et al. 2015 [16]	N/A	180	–	N/A	Polyp localization	
Park and Sargent 2016 [17]	N/A	652 With polyps: 92	35 (20′ to 40′)	N/A	Polyp localization	
Ribeiro et al. 2016 [38]	66 to 86	85 to 126	–	N/A	Polyp classification (neoplastic vs non-neoplastic)	8 datasets by combining: (i) with or without staining mucosa, (ii) 4 acquisition modes (without CVC, i-Scan1, i-Scan2, i-Scan3).
Zhang R. et al. 2017 [19], Zheng Y. et al. 2018 [24]	N/A	1930 Without polyps: 1 104 Hyperplastic: 263 Adenomatous: 563	–	215 polyps (65 hyperplastic and 150 adenomatous)	Polyp classification (hyperplastic vs. adenomatous)	PWH Database. Images taken under either WL or NBI endoscopy.
Yuan and Meng 2017 [20]	35	4 000 Normal WCE images: 3 000 (1 000 bubbles, 1 000 turbid, and 1 000 clear) Polyp images: 1 000	–	N/A	Polyp detection	
Byrne et al. 2017 [39]	N/A	N/A	388	N/A	Polyp classification (hyperplastic vs. adenomatous)	
Komeda et al. 2017 [40]	N/A	1 800 Adenomatous: 1200 Non-adenomatous: 600	–	N/A	Polyp classification (adenomatous vs. non-adenomatous)	
Chen et al. 2018 [41]	N/A	2 441 Training: - Neoplastic: 1476 - Hyperplastic: 681 Testing: - Neoplastic: 188 - Hyperplastic: 96	–	N/A	Polyp classification (hyperplastic vs. neoplastic)	
Misawa et al. 2018 [23]	73	N/A	546 (155 positive and 391 negative)	155	Polyp detection	
Urban et al. 2018 [26]	>2000	8 641	–	4 088	Polyp localization	Used as training dataset.
Urban et al. 2018 [26]	N/A	1 330 With polyps: 672 Without polyps: 658	–	672	Polyp localization	Used as independent dataset for testing.
Urban et al. 2018 [26]	9	44 947 With polyps: 13 292 Without polyps: 31 655	9	45	Polyp localization	Used as independent dataset for testing.
Urban et al. 2018 [26]	11	N/A	11	73	Polyp localization	Used as independent dataset for testing with "deliberately more challenging colonoscopy videos.".
Wang et al. 2018 [27]	1 290	5 545 With polyps: 3 634 Without polyps: 1 911	–	N/A	Polyp localization	Used as training dataset.
Wang et al. 2018 [27]	1 138	27 113 With polyps: 5 541 Without polyps: 21 572	–	1 495	Polyp localization	Used as testing dataset.
Wang et al. 2018 [27]	110	–	138	138	Polyp localization	Used as testing dataset.
Wang et al. 2018 [27]	54	–	54	0	Polyp localization	Used as testing dataset.
Lui et al. 2019 [37]	N/A	8 000 Curable lesions: 4 000 Incurable lesions: 4 000	–	Curable lesions: 159 Incurable lesions: 493	Polyp classification (endoscopically curable vs. incurable lesions)	Used as training dataset. This study is focused on larger endoscopic lesions with risk of submucosal invasion and lymphovascular permeation.
Lui et al. 2019 [37]	N/A	567	–	Curable: 56 Incurable: 20	Polyp classification (endoscopically curable vs. incurable lesions)	Used as testing dataset. This study is focused on larger endoscopic lesions with risk of submucosal invasion and lymphovascular permeation.
Blanes-Vidal et al. 2019 [30]	255	11 300 With polyps: 4 800 Without polyps: 6 500	N/A	331 polyps (OC) and 375 (CCE)	Polyp localization	CCE: Colorectal capsule endoscopy. OC: conventional optical colonoscopy.
Zhang X. et al. 2019 [31]	215	404	–	N/A	Polyp localization	

(*continued on next page*)

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Table 4 (*continued*)

Study	Patients	No. Images	No. Videos	No. Unique Polyps	Purpose	Comments
Misawa et al. 2019 [32]	N/A	3 017 088	–	930	Polyp detection	Used as training set.
Misawa et al. 2019 [32]	64 (47 with polyps and 17 without polyps)	N/A	N/A	87	Polyp detection	Used as testing set.
Kandel et al. 2019 [42]	552	N/A	–	963	Polyp classification (hyperplastic, sessile serrated adenomas, adenomas)	
Zhu X. et al. 2019 [33]	283	1 991	–	N/A	Polyp detection	Adenomatous polyps.
Ahmad et al. 2019 [34]	N/A	83 716 With polyps: 14 634 Without polyps: 69 082	17	83	Polyp localization	White Light Images.
Sornapudi et al. 2019 [35]	N/A	55	N/A	67	Polyp localization	Wireless Capsule Endoscopy videos. Used as testing set.
Sornapudi et al. 2019 [35]	N/A	1 800 With polyps: 530 Without polyps: 1 270	18	N/A	Polyp localization	Wireless Capsule Endoscopy videos. Used as training set.
Wittenberg et al. 2019 [36]	N/A	2 484	–	2 513	Polyp localization	

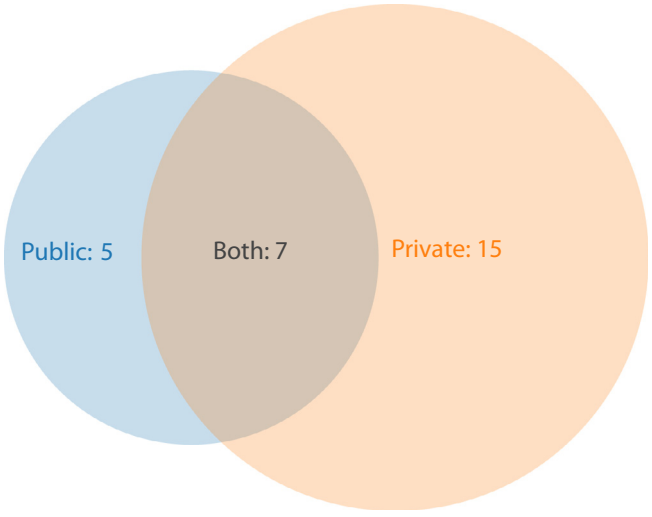

Fig. 2. Euler diagram showing the studies using only public datasets (5), only private datasets (15) and both public and private datasets (7).

layers fine-tuning, as they are responsible of extracting the image features, keeping their weights unchanged implies that the same features learnt in the original training dataset will be extracted from the new images. Therefore, researchers should evaluate whether the original features are compatible with the new task or not, in which case, it is also recommended to fine-tune the convolutional layers. As a conclusion, the use of TL not only speeds up the model development, but also reduces the need for large datasets for model training.

Table 5 lists the works where off-the-shelf DL models were used for polyp detection or classification. As can be seen, the main strategy for polyp classification and detection is to use off-the-shelf CNNs pre-trained with ImageNet or other dataset, replacing the last layers with a FCL or an SVM classifier. Ribeiro et al. [38] take to the extreme this strategy, applying it to seven different CNNs. Zhang R. et al. [22] also test several CNNs but, in this case, they use the same architecture (i.e. CaffeNet [61] with an output SVM), experimenting with different SVM kernels and with the convolutional layer connected to this classifier. Specifically, the SVM is evaluated connected to each convolutional layer. The final classifier comprises two CNNs: a first CNN detects polyps and, when a polyp is detected, a second CNN classifies it into adenoma or hyperplasia.

Regarding the polyp localization, each work reviewed use a different framework, including two object detection frameworks and two segmentation frameworks. As expected, the main datasets used for transfer learning are Microsoft's COCO and PASCAL VOC, as they include object segmentation metadata. A noteworthy achievement of the models proposed by Zheng Y. et al. [24] and Wang et al. [27] is that they can localize multiple polyps in real time, which makes them suitable for use in a clinical environment.

Developing a custom DL model has the main advantage that its design can be adapted to the target task, which should be reflected in the model performance. However, creating a new DL model from scratch may be a challenging task as it requires advanced knowledge of the DL techniques, architectures and libraries.

One way to reduce the difficulty of creating a custom DL model is to build it on the basis of an existing model. This is the strategy followed, for example, by Yuan and Meng [20] for polyp detection, where they propose a modification of an Stacked Sparse AutoEncoder (SSAE) to include image manifold information on each layer of the classifier, or by Byrne et al. [39] for polyp classification, where an Inception v3 [62] network was complemented with a credibility function used to generate a credibility score for each frame based on the current frame prediction and the credibility score of previous frames. Despite being a more complex task, this strategy can also be found in polyp location works, such as the work from Urban et al. [26], where authors replace the FCL of VGG16, VGG19 [59] and ResNet-50 [63] networks with convolutional layers following the ideas behind the YOLO framework, or the work from Blanes-Vidal et al. [30], where a R-CNN with a modified AlexNet [64] backbone is proposed. Also in this context, it is worth highlighting the work from Qadir et al. [29], as they propose a new framework to reduce false positives in object detection frameworks by adding a false positive reduction unit that exploits backward and forward temporal dependencies between frames to correct the output. This new framework was tested in combination with Faster R-CNN and SSD [65] frameworks.

Although the changes done to the DL models in the papers cited above are minor, they have proven to be enough for boosting the models performance or increasing the confidence in the generated

Table 5
Summary of off-the-shelf deep learning architectures used in the studies.

Study	Task	Model/s	Framework	TL	Layers fine-tuned	Layers replaced	Output layer
Ribeiro et al. 2016 [38]	Classification	AlexNet GoogLeNet Fast CNN Medium CNN Slow CNN VGG16 VGG19	–	ImageNet	N/A	Layers after last CNN layer	SVM
Zhang R. et al. 2017 [19]	Detection and classification	CaffeNet	–	ImageNet and Places205	N/A	Tested connecting classifier to each convolutional layer (5 convolutional layers)	SVM (Poly, Linear, RBF, and Tahn)
Chen et al. 2018 [41]	Classification	Inception v3	–	ImageNet	N/A	Last layer	FCL
Misawa et al. 2018 [23], Misawa et al. 2019 [32]	Detection	C3D	–	N/A	N/A	N/A	N/A
Zheng Y. et al. 2018 [24]	Localization		YOLOv1	PASCAL VOC 2007 and 2012	All	–	–
Shin Y. et al. 2018 [25]	Localization	Inception ResNet-v2	Faster R-CNN with post-learning schemes	COCO	All	–	RPN and detector layers
Urban et al. 2018 [26]	Localization	ResNet-50 VGG16 VGG19	–	ImageNet Also without TL	All	Last layer	FCL
Wang et al. 2018 [27]	Localization	VGG16	SegNet	N/A	N/A	N/A	N/A
Wittenberg et al. 2019 [36]	Localization	ResNet101	Mask R-CNN	COCO	All (incrementally)	Last layer	FCL

outputs. However, examples of other works with more complex modifications can also be found for localization tasks. Zhu R. et al. [16] propose a framework where a CNN based on LeNet-5 [66] with a SVM output classifier is fed with image patches generated via a sliding window and the positively selected patches are combined to generate a final binary mask. Zhang et al. [31] propose a modification of an SSD, which is extended by adding three new pooling layers and a new deconvolution layer whose features are concatenated to those from the Max-Pooling layer and fed into the detection layer. Brandao et al. [21] and Ahmad et al. [34] propose a general strategy to perform localization with regular CNNs, that is based on replacing the FCLs and the output layer with a convolutional layer and adding a regularization operator between each convolutional and activation layer. This strategy was tested with 6 known CNNs and the best three were also tested with an additional modification consisting on using also shape-form-shading features that provide 3D structure information. Finally, Sornapudi et al. [35] propose a new framework that takes many ideas from the Mask R-CNN, adding some features and replacing the elements after the RoIAlign layer with a custom probability mask network.

Regarding the DL models designed from scratch, several works follow a simple strategy of building them with different combinations of known layers. These models are usually composed of several initial convolutional layers followed by pooling layers and finalized with FCLs. This is the case of Komeda et al. [40], Urban et al. [26] or Lui et al. [37]. In some cases, these layers are combined with other less common layers, such as in the Urban et al. [26] work, where inception and batch normalization layers are used, or in the work of Kandel et al. [42], where convolutional capsule layers [67] are the main components of the proposed model.

When analyzing more complex custom DL models, the works of Tajbakhsh et al. [14,15] stand out because of the particular combination of classic computer vision techniques with DL. In fact, the detection and location of polyps is done with the classic computer vision techniques and the responsibility of the CNN is limited to correct the prediction. Similarly, Park and Sargent [17] present a modification of a previous work where a three step (preprocessing, feature extraction and classification) workflow for polyp location is proposed with classic computer vision techniques. The previous workflow is modified by just replacing the original feature extractor with a CNN.

Experimenting with more complex models, Yu et al. [18] present a custom pure DL localization model based on the use of two 3D Fully Convolutional Network (3D-FCN) whose probabilities maps are combined to generate the final output. The singularity of this model is that one of the 3D-FCN (offline 3D-FCN) is trained, as usual, with a training dataset, while the other 3D-FCN (online 3D-FCN) is initialized with the offline 3D-FCN weights and retrained each 60 frames during video processing. This strategy adapts the network to the particular video characteristics. Apart from that, the 3D characteristic of the convolutional layers used allows the use of the 16 previous frames to perform localization on each frame.

Another example of a custom pure DL localization model is the one presented by Zhang R. et al. [22]. In this case, a custom architecture named RYCO is presented. It consists of two networks: a regression-based model with residual learning detection model (ResYOLO) that locates polyps in a frame, and an Efficient Convolution Operators (ECO) [68] model to track the detected polyps. While the ResYOLO network is responsible for detecting and starting the polyp tracking, the ECO model helps the ResYOLO to track the polyp. Polyp tracking stops when a confidence score is under a certain threshold.

The main characteristics and highlights of the custom models developed in the works reviewed can be found in Supplementary Table 1.

As stated previously, medical image analysis suffers a limitation regarding the availability of big datasets of curated images or videos. This limitation usually leads to the problem of overfitting: the DL models are not able to generalize and fail when they are asked to make predictions on new, unseen images. To overcome this limitation, data augmentation strategies are used in order to

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Table 6
Summary of the different data augmentation strategies used in different studies.

	Rotation	Flipping	Shearing	Translation	Gaussian smoothing	Crop	Scale	Resize	Random brightness	Zooming	Saturation adjustment	Random contrast	Exposure adjustment	Histogram equalization
# Studies	14	10	4	3	3	3	2	2	2	2	1	1	1	1
Tajbakhsh et al. 2015 [15]	X			X		X	X	X						
Park and Sargent 2016 [17]	X			X										
Ribeiro et al. 2016 [38]	X	X												
Yu et al. 2017 [18]	X			X										
Byrne et al. 2017 [39]		X				X		X						
Brandao et al. 2018 [21]		X												
Zhang R. et al. 2018 [22]	X	X			X				X			X		
Zheng Y. et al. 2018 [24]	X													
Shin Y. et al. 2018 [25]	X	X	X		X				X	X				
Urban et al. 2018 [26]	X	X	X											
Qadir et al. 2019 [29]	X	X	X							X				
Blanes-Vidal et al. 2019 [30]	X	X				X								
Zhang X. et al. 2019 [31]	X													
Zhu X. et al. 2019 [33]	X										X		X	
Sornapudi et al. 2019 [35]	X	X	X		X		X							X
Wittenberg et al. 2019 [36]	X	X												

improve the size and quality of training datasets and mitigate the problem of overfitting [69]. Broadly speaking, data augmentation allows obtaining more images by applying different transformations (e.g. rotation or flipping) to the original set of images. Table 6 summarizes the types of data augmentation used by the different studies reviewed, where it can be seen that 16 out of the 27 studies reportedly apply data augmentation strategies. Studies not appearing in this table either do not use data augmentation (6/27), do not specify whether they use it or not (4/27) or do use it but do not specify the specific types used (1/27). At first, it can be seen that the two most used strategies are rotation (14/16) and flipping (10/16). After these two, there is a group of eight strategies used in two, three or four different studies. Finally, there are four rare strategies used only in one study.

Interestingly, Sornapudi et al. [35], evaluated the same model with and without data augmentation on the ETIS-Larib and ASU-Mayo Clinic Colonoscopy Video datasets. They report a significant performance improvement when the model is trained using data augmentation, with an important reduction of the number of false positives and false negatives and an increase of the true positives (results can be found in Tables 6 and 8 of Sornapudi et al. [35]).

Finally, Table 7 shows a summary of the DL frameworks and libraries used in different studies. The studies not included in the table did not report this information. This data is consistent with the usage statistics reported in [70], where Tensorflow, Caffe and Keras are the most used libraries. It is important to note that, although C3D [71] appears in two works, the network is the same in both cases.

Table 7
Summary of the DL frameworks and libraries used in different studies.

Framework/ Library	# Studies	Used by
Caffe	5	Zhu X. et al. 2019 [33], Yu et al. 2017 [18], Brandao et al. 2018 [21], Wang et al. 2018 [27], Zhang X. et al. 2019 [31]
Tensorflow	2	Chen et al. 2018 [41], Shin Y. et al. 2018 [25]
C3D	2	Misawa et al. 2018 [23], Misawa et al. 2019 [32]
Keras	3	Urban et al. 2018 [26], Sornapudi et al. 2019 [35], Wittenberg et al. 2019 [36]
MatConvNet (MATLAB)	1	Ribeiro et al. 2016 [38]

5. State of the art performance over recent years

5.1. Polyp detection and localization

Polyp detection and localization is generally performed at frame-level, i.e. each frame is predicted independently as containing a polyp and/or where the polyp is located. After each frame is predicted, some studies do a polyp-based analysis by defining a threshold of positive frame predictions where the polyp can be considered as being correctly detected by the model. For example, Misawa et al. defined a positive polyp detection "as the system output over the cutoff value for >75% of the duration of each short video" [23].

For detection, the most widely used evaluation metrics are recall, precision or PPV (Predictive Positive Value), specificity, F1 and F2 [10]:

- Specificity: the proportion of true negatives that were predicted as such. The specificity is given by:

$$specificity = \frac{TN}{TN + FP}$$

- Sensitivity (or recall): the proportion of true positives that were predicted as such. The sensitivity is given by:

$$sensitivity = recall = \frac{TP}{TP + FN}$$

- Precision (or PPV): the proportion of predicted positives that are real positives. The positive predictive value is given by:

$$ppv = precision = \frac{TP}{TP + FP}$$

- F1 score: a measure combining recall and precision. The F1 score is given by:

$$F_1 = 2\frac{recall * precision}{recall + precision}$$

- F_2 score: a measure combining recall and precision. The F_2 score is given by:

$$F_2 = 4\frac{recall * precision}{recall + 4 * precision}$$

Table 8
Summary of detection performance on several public datasets (f = per-frame). Results are for "best model" (i.e.: a single model that is recommended by authors or that with the best balance among recall/precision).

Dataset (i = image, v = video)	Studies	Recall (sensitivity)	Precision (PPV)	F1	F2
ETIS-Larib (i)	Ahmad et al. 2019 [34]	**91.6%**	75.3%	**0.88**	**0.88**
		~90%	~73%	~0.86	~0.86
	Brandao et al. 2018 [21]	83%	74%	0.81	0.81
	Wittenberg et al. 2019 [36]	80.3%	**86.5%**	0.82	0.82
	Shin Y. et al. 2018 [25]	80.29%	72.93%	0.78	0.78
	Sornapudi et al. 2019 [35]	74%	77.4%	0.74	0.74
	Zheng Y. et al. 2018 [24]				
CVC-ColonDB (i)	Sornapudi et al. 2019 [35]	**91.64%**	**89.94%**	**0.97**	**0.91**
	Brandao et al. 2018 [21]	~90%	~80%	~0.85	0.88
CVC-PolypHD (i)	Sornapudi et al. 2019 [35]	78.12%	83.33%	0.80	0.79
CVC-ClinicDB (i)	Wang et al., 2018 [27]	**88.2%**	**93.1%**	**0.91**	**0.89**
	Wittenberg et al. 2019 [36]	86%	80%	0.82	0.85
ASU-Mayo (v)	Shin Y. et al. 2018 [25]	**84.2%**	82.7%	**0.83**	**0.84**
	Zhang R. et al. 2018 [22]	71.6%	**88.6%**	0.79	0.74
	Yu et al. 2017 [18]	71%	88.1%	0.78	0.73
CVC-ClinicVideoDB (v)	Shin Y. et al. 2018 [25]	**84.3%**	**89.7%**	**0.87**	**0.85**
	Qadir et al. 2019 [29]	81.51%	87.51%	0.84	0.83

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Performance shown in different studies is difficult to compare. Although the same measures are used as performance estimators, less strict validation schemes (e.g.: reporting only the training performance, lack of independent test set, etc.) or the procedure to do the model selection can produce bias (e.g. testing many different hyperparameters or architectures over the same test set can be affected by regression to the mean). Moreover, in the case of polyp detection, two particular problems can be present regarding dataset used. On one hand, real scenarios contain proportionally less polyp-containing frames than development datasets, which especially affects the precision measure. On the other hand, image datasets usually contain manually selected images, which are different to those found in a real endoscopy video stream (appearance of instrumental, water, illumination and/or focus variations, etc.), making a priori promising models less reliable on the clinical setting. In this sense, testing on human-unaltered video datasets is challenging but recommended. Interestingly, Urban et al. [26] created a "challenging" video dataset that contain polyp videos where the endoscopist was instructed to perform a "flyby", in order to simulate a real scenario where the endoscopist is missing a polyp.

In order to compare detection studies as fair as possible, Table 8 shows a summary of detection performance reported by different authors using the same public datasets. Supplementary Table 2 provides the complete performance results of these studies in both public and private datasets. Regarding the public image datasets, the best overall performance was achieved by Sornapudi et al. [35] in the CVC-ColonDB dataset (380 WL images), with a recall of 91.64%, a PPV of 89.94%, a F1 of 0.97 and a F2 of 0.91. The highest precision (93.1%) was achieved by Wang et al. [27] using the CVC-ClinicDB (612 WL images). The public dataset used by more studies is the ETIS-Larib (196 WL images), in which Ahmad et al. [34] obtained one of the best performances, with a recall of 91.6% and a F1/F2 of 0.88, with a lower precision. The best precision in this dataset (86.5%) was obtained by Shin Y. et al. [25]. On the other hand, the recall on the two video datasets (ASU-Mayo and CVC-ClinicVideoDB) does not exceed 85%, whilst the highest precision is almost 90% (Shin Y. et al. [25]).

Table 9 shows the average performances and standard deviations of all the polyp detection and localization studies. The first row reports all the average performances calculated by using a frame-based approach. These include those where a manual set of images was used (i.e. the images were manually selected and curated by experts or the researchers) as well as those that use all the available images. To get insight about the possibility that this issue can bias the results, the second and third rows report the average frame-based performances separated based on this condition. Although the differences are not very important, the usage of manual sets of images results in slightly higher average performances with slower standard deviations. Finally, the fourth row reports the average polyp-based performances, where any of the studies use manually selected images. Polyp-based analysis show better recall or sensitivity, but with a significant lower precision and specificity.

5.2. Polyp classification

As mentioned in Section 2, we have found less studies performing polyp classification with heterogeneity regarding the classes or types of polyps that they try to differentiate. Two of the seven studies identified in Section 2 have been excluded (Ribeiro et al. [38], Komeda et al. [40]) from this analysis since they only report accuracies using different configurations. The studies of these results are summarized in Table 10. Supplementary Table 3 provides the complete performance results of these studies in both public and private datasets.

In this regard, three studies are focused on the most classical and interesting differentiation: adenoma vs. hyperplastic (Zhang R. et al. [19] and Byrne et al. [39]) and adenoma vs. neoplastic (Chen et al. [41]). Zhang R. et al. [19] also evaluated the possibility of differentiating resectable (serrated adenoma and adenoma) and non-resectable (hyperplasia) polyps. Lui et al. [37] goes beyond polyp classification and focus on larger endoscopic lesions with risk of submucosal invasion and lymphovascular permeation with the goal of distinguishing endoscopically curable lesions versus endoscopically incurable lesions. Interestingly, Kandel et al. [42] evaluated two classifications: hyperplastic vs. adenoma with far focus images and hyperplastic vs. serrated with near focus. This makes sense since with far focus serrated and hyperplastic polyps are almost indistinguishable, thus, the interest consists in differentiating adenomatous and hyperplastic polyps. Then, once the focus is near enough, it is possible to try to differentiate hyperplastic from serrated polyps.

Regarding the way in which performance measures are taken, all studies report frame-based metrics on image datasets, with the only exception of Byrne et al. [39], who reports polyp-based metrics analyzing an unaltered video dataset. Although Byrne

Table 9
Mean performances and standard deviations of all polyp detection and localization studies.

	Recall (sensitivity)	Precision (PPV)	Specificity	F1	F2
Frame-based	86% ± 8%	83% ± 11%	89% ± 12%	0.84 ± 0.08	0.85 ± 0.08
All video frames	84% ± 9%	81% ± 14%	87% ± 13%	0.81 ± 0.09	0.81 ± 0.08
Manually selected images	86.88% ± 7.33%	83.98% ± 8.69%	95.75% ± 4.60%	0.85 ± 0.07	0.86 ± 0.07
Polyp-based (Systematic)	95% ± 6%	48% ± 13%	69% ± 27%	0.63 ± 0.11	0.78 ± 0.06

Table 10
Performance measures of different polyp classification studies. (Classifier is specified in Class1/Class2/. . ./ClassN format. For two-class classifiers the first one is considered the "positive" class. A/H = Adenoma vs. Hyperplastic, R/¬R = Resectable vs. Non-resectable, A/H/S = Adenoma vs. Hyperplastic vs. Serrated, N/H = Neoplastic vs. Hyperplastic, ECL/EIL = Endoscopically curable lesion vs. Endoscopically incurable lesion, H/S = Hyperplastic vs. serrated adenoma, H/A = Hyperplastic vs. Adenoma).

	Classifier	Sensitivity	Specificity	PPV	NPV	Accuracy
Zhang R. et al. 2017 [19]	A/H	87.6%	84.2%	87.30%	87.2%	–
	A/H/S	–	–	–	–	86.7%
	R/¬R	92%	89.9%	95.4%	84.9%	–
Byrne et al. 2019 [39]	A/H	98%	83%	90%	97%	N/A
Chen et al. 2018 [41]	N/H	96.3%	78.1%	89.6%	91.5%	N/A
Lui et al. 2019 [37]	ECL/EIL	88.2%	77.9%	92.1%	69.3%	85.5%
Kandel et al. 2019 [42]	H/S (near focus)	57.14%	68.52%	N/A	N/A	67.21%
	H/A (far focus)	75.63%	63.79%			72.48%

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

et al. [39] reported very good performance results, the fact that they are polyp-based can make them not directly comparable with the other studies. The proposed system analyzes each polyp frame using a credibility score calculated using the previous frames.

The work of Byrne et al. [39] is also the only study that operates in real-time. While in polyp detection and localization it is very important for the systems to produce the outcomes in real-time, in the case of classification it is less important, especially if the input of the automatic polyp classification model is the output of a previous detection or localization system.

6. Conclusions and future directions

This work presents a review of the most relevant studies applying DL to both polyp detection (including localization) and classification from a technical point of view. Most of these studies are focused on polyp localization on conventional colonoscopy (15/27), with a few cases where WCE is used as the colonoscopy method. The lower number of studies dealing with polyp classification found (7/27) may be probably due to the lack of public datasets for this task, as there is only one public dataset suitable for classification [51], but with a limited size for DL.

Research on polyp detection is currently in a more advanced state than on polyp classification. A clear sing of this is that several studies are capable of doing polyp localization of multiple polyps in real-time [24,27,28], two important conditions for they use in clinical environments. In fact, a prospective randomized controlled trial has been already done [28]. On the other hand, polyp classification is still at a less advanced state, with fewer studies available and, moreover, with diverse classification objectives. In addition to this, the fact that there is only a single public polyp classification dataset is also an important problem. Therefore, having a large public dataset for polyp classification would certainly benefit this field, allowing the unification of the objectives and efforts of the studies published in the future.

Regarding the datasets, most of the studies reviewed used private datasets (22/27), in some cases in combination with public datasets (7/22). Although there are 8 public datasets, their size is relatively small for DL tasks, as, in most cases, they only include a few dozen polyps and/or videos. In addition, as commented above, only one dataset [51] is designed for polyp classification. Recently, a public dataset for polyp detection that includes 1 000 polyp images has been published [52], so we expect to see new studies in the future taking advantage of it.

We have found that the architectures proposed and the strategies followed to design DL models are pretty diverse, ranging from studies that use off-the-shelf pretrained models to studies that propose custom designed DL architectures, including studies that modify off-the-shelf models or studies that mix DL with classic computer vision techniques. Besides, performance results are, in general, good, so it does not seem to be a strategy that is better than the others. Despite of the diversity in the architectures, data augmentation techniques have been widely used in the studies reviewed and, for example, Sornaputi et al. [35] reported a better performance of the model when data augmentation is used. Therefore, its use seems to be positive.

Missed rates of colonic polyps are quite high. As we have highlighted previously, the missed rate for adenoma was 26%, for serrated polyps reached 27% and, more importantly, for advanced adenomas reached 9% [4]. The use of auxiliary techniques has shown their utility for reducing the missed rates of colonic polyps. In this respect, CAD devices have demonstrated their ability to improve the ADR. However, this increase in ADR is mainly made at the expense of small polyps and non-advanced adenomas. In an observational study comparing performance of CAD versus original endoscopists and expert reviewing videos, the CAD demonstrated better polyp detection rate (PDR) and ADR, but there were no differences in detection of advanced adenomas [26]. In the randomized controlled trial (RCT) performed in Chinese symptomatic patients, the CAD demonstrated significantly higher PDR and ADR, but only in detection of hyperplastic and non-advanced adenomas, and no improvement was found in detection of advanced colonic lesions [28]. These results highlight a potential problem of CAD: its ability of finding only small non-significant colonic lesions and the lack of demonstration of its potential ability for improving detection of large and significant adenomas or serrated polyps. This fact could be a challenge in the demonstration of the role of these devices in CRC prevention.

Among small polyps (<5 mm) most of them are serrated lesions with no potential risk of harboring an adenocarcinoma and no increased risk of developing a CRC after resection. Hypothetically, the optical diagnosis of histology would prevent histological analysis of tiny lesions (“resect and discard”), would allow to give recommendations for post-polypectomy surveillance immediately after the endoscopic procedure and leave in situ those non-adenomatous lesions in recto-sigma (“leave in situ”). In this sense, the optical diagnosis would reduce the costs associated with histological diagnosis and the risks associated with endoscopic resection. However, minimum requirements for normal use must be met. The American Society for Gastrointestinal Endoscopy (ASGE), through a committee created for the evaluation of new technologies, has established the minimum requirements of the different technologies to implement these strategies. For the “leave in situ” strategy, histology prediction based on optical diagnosis must have a negative predictive value of 90% for adenomas. For the “resect and discard” strategy, there must be a 90% agreement between the recommendations for the follow-up interval based on optical diagnosis and those based on the histopathological evaluation of polyps [72].

DL Networks, and more specifically CNNs, have boosted the performance of computer vision systems, especially regarding detection, localization and classification of objects in images. In the field of colonoscopy, there has also been a lot of research in the last years regarding the usage of these techniques to perform automatic polyp detection and classification. Since the Automatic Polyp Detection sub-challenge conducted at MICCAI 2015, a growing number of studies using DL to accomplish these tasks are published every year. Nevertheless, despite the promising results obtained in some studies, in terms of both performance and computational time, there is still room for improvement and some challenges must be addressed before these systems can be tested and implemented in real clinical settings.

One important aspect of these systems is that they must be able to operate in real-time to be useful in routine clinical practice. The research and improvements coming from the DL and CNN architectures are expected to continue improving the computational performance and allow the development of models that can deal with colonoscopic data in real-time. In addition, these improvements will likely help in improving these systems in terms of classification performance (i.e. accuracy, recall, precision, and so on). However, the improvements in these metrics are also subjected to the availability of better and bigger publicly available datasets annotated for both polyp detection and classification. In this regard, it would be desirable the existence of such public database to complement the existing ones as other authors have also highlighted [11,13,24].

Remarkably, the existence of this database would also help to compare the performance of existing and upcoming models for polyp detection and classification in a better and fairer way. As noted in the previous section, one of the major barriers to compare the performances of the studies reviewed is the heterogeneity

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

regarding the datasets used, as well as the heterogeneity in the training schemes employed. The existence of a large database with a high number of annotated videos and images would allow the creation of reference development and test datasets that the community can use to: (*i*) develop and train DL-based models using the development dataset, which can be further partitioned in training and validation datasets; and (*ii*) test these models or other models developed with private datasets against the reference test dataset. This way, the performance of the models could be straightforwardly compared since all are tested against the same test dataset.

Finally, when the DL-based systems can operate in real-time achieving acceptable performance rates, their results should be validated in clinical trials to measure their impact in the colonoscopic practice. An important step in this direction, is the study by Wang et al. [28], which is the first prospective randomized controlled trial analyzing an automatic polyp detection during colonoscopy. This study reported increase of ADR by 50%, from 20% to 30%, mainly because of a higher rate of small adenomas found.

CRediT authorship contribution statement

Alba Nogueira-Rodríguez: Conceptualization, Investigation, Writing - original draft. **Rubén Domínguez-Carbajales:** Investigation. **Hugo López-Fernández:** Conceptualization, Writing - original draft. **Águeda Iglesias:** Investigation. **Joaquín Cubiella:** Writing - original draft. **Florentino Fdez-Riverola:** Writing - review & editing. **Miguel Reboiro-Jato:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision. **Daniel Glez-Peña:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

SING group thanks CITI (*Centro de Investigación, Transferencia e Innovación*) from the University of Vigo for hosting its IT infrastructure. This work was partially supported by the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding of ED431C2018/55-GRC Competitive Reference Group and by the Ministerio de Economía, Industria y Competitividad, Gobierno de España under the scope of the PolyDeep project (DPI2017-87494-R). The authors also acknowledge the grants of Alba Nogueira-Rodríguez (predoctoral fellowship ED481A-2019/299) and Hugo López-Fernández (postdoctoral fellowship ED481B 2016/068-0), funded by the Xunta de Galicia. Joaquín Cubiella received grants from Instituto de Salud Carlos III (PI11/00094 and PI17/00837).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neucom.2020.02.123.

References

[1] Cubiella J, Marzo-Castillejo M, Mascort-Roca JJ, Amador-Romero FJ, Bellas-Beceiro B, Clofent-Vilaplana J, et al. Guía de práctica clínica. Diagnóstico y prevención del cáncer colorrectal. Actualización 2018. Gastroenterología y Hepatología. 2018;41:585–96.

[2] S.J. Winawer, A.G. Zauber, M.N. Ho, M.J. O'Brien, L.S. Gottlieb, S.S. Sternberg, et al., Prevention of colorectal cancer by colonoscopic polypectomy, N. Engl. J. Med. 329 (1993) 1977–1981.

[3] S. Singh, P.P. Singh, M.H. Murad, H. Singh, J.N. Samadder, Prevalence, risk factors, and outcomes of interval colorectal cancers: a systematic review and meta-analysis, Am. J. Gastroenterol. 109 (2014) 1375–1389.

[4] S. Zhao, S. Wang, P. Pan, T. Xia, X. Chang, X. Yang, et al., Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis, Gastroenterology 156 (1661–1674) (2019) e11.

[5] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, et al., Computer-aided colorectal tumor classification in NBI endoscopy using local features, Med. Image Anal. 17 (2013) 78–100.

[6] J.H. Chen, S.M. Asch, Machine learning and prediction in medicine — beyond the peak of inflated expectations, N. Engl. J. Med. 376 (2017) 2507–2509.

[7] Nogueira-Rodríguez A, López-Fernández H, Glez-Peña D. Deep Learning Techniques for Real Time Computer-Aided Diagnosis in Colorectal Cancer. In: Herrera-Viedma E, Vale Z, Nielsen P, Martin Del Rey A, Casado Vara R, editors. Distributed Computing and Artificial Intelligence, 16th International Conference, Special Sessions. Cham: Springer International Publishing; 2020. p. 209–12. doi:10.1007/978-3-030-23946-6_27.

[8] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. arXiv:170205747 [cs]. 2017. http://arxiv.org/abs/1702.05747. Accessed 12 May 2017.

[9] X. Liu, L. Faes, A.U. Kale, S.K. Wagner, D.J. Fu, A. Bruynseels, et al., A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis, Lancet Digital Health 1 (2019) e271–e297.

[10] J. Bernal, N. Tajkbaksh, F.J. Sánchez, B.J. Matuszewski, H. Chen, L. Yu, et al., Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge, IEEE Trans. Med. Imaging 36 (2017) 1231–1249.

[11] D.G. Vinsard, Y. Mori, M. Misawa, S. Kudo, A. Rastogi, U. Bagci, et al., Quality assurance of computer-aided detection and diagnosis in colonoscopy, Gastrointest. Endosc. 90 (2019) 55–63.

[12] S.A. Azer, Challenges facing the detection of colonic polyps: what can deep learning do?, Medicina 55 (2019) 473.

[13] J.K. Min, M.S. Kwak, J.M. Cha, Overview of deep learning in gastrointestinal endoscopy, Gut Liver (2019).

[14] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automatic polyp detection using global geometric constraints and local intensity variation patterns, in: P. Golland, N. Hata, C. Barillot, J. Hornegger, R. Howe (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014, Springer International Publishing, 2014, pp. 179–187.

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

[16] R. Zhu, R. Zhang, D. Xue, Lesion detection of endoscopy images based on convolutional neural network features, in: In: 2015 8th International Congress on Image and Signal Processing (CISP), 2015, pp. 372–376.

[17] Park SY, Sargent D. Colonoscopic polyp detection using convolutional neural networks. In: Medical Imaging 2016: Computer-Aided Diagnosis. International Society for Optics and Photonics; 2016. p. 978528. doi:10.1117/12.2217148.

[18] L. Yu, H. Chen, Q. Dou, J. Qin, P.A. Heng, Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos, IEEE J. Biomed. Health. Inf. 21 (2017) 65–75.

[19] R. Zhang, Y. Zheng, T.W.C. Mak, R. Yu, S.H. Wong, J.Y.W. Lau, et al., Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain, IEEE J. Biomed. Health Inform. 21 (2017) 41–47.

[20] Y. Yuan, M.Q.-H. Meng, Deep learning for polyp recognition in wireless capsule endoscopy images, Med. Phys. 44 (2017) 1379–1389.

[21] P. Brandao, O. Zisimopoulos, E. Mazomenos, G. Ciuti, J. Bernal, M. Visentini-Scarzanella, et al., Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks, J. Med. Robot Res. 03 (2018) 1840002.

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

[23] M. Misawa, S.-E. Kudo, Y. Mori, T. Cho, S. Kataoka, A. Yamauchi, et al., Artificial intelligence-assisted polyp detection for colonoscopy: initial experience, Gastroenterology 154 (2027–2029) (2018) e3.

[24] Zheng Y, Zhang R, Yu R, Jiang Y, Mak TWC, Wong SH, et al. Localisation of Colorectal Polyps by Convolutional Neural Network Features Learnt from White Light and Narrow Band Endoscopic Images of Multiple Databases. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI: IEEE; 2018. p. 4142–5. doi:10.1109/EMBC.2018.8513337.

[25] Y. Shin, H.A. Qadir, L. Aabakken, J. Bergsland, I. Balasingham, Automatic colon polyp detection using region based deep CNN and post learning approaches, IEEE Access 6 (2018) 40950–40962.

[26] G. Urban, P. Tripathi, T. Alkayali, M. Mittal, F. Jalali, W. Karnes, et al., Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy, Gastroenterology 155 (1069–1078) (2018) e8.

[27] P. Wang, X. Xiao, J.R. Glissen Brown, T.M. Berzin, M. Tu, F. Xiong, et al., Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy, Nat. Biomed. Eng. 2 (2018) 741–748.

[28] P. Wang, T.M. Berzin, J.R.G. Brown, S. Bharadwaj, A. Becq, X. Xiao, et al., Real-time automatic detection system increases colonoscopic polyp and adenoma

detection rates: a prospective randomised controlled study, Gut 68 (2019) 1813–1819.
[29] H.A. Qadir, I. Balasingham, J. Solhusvik, J. Bergsland, L. Aabakken, Y. Shin, Improving automatic polyp detection using CNN by exploiting temporal dependency in colonoscopy video, IEEE J. Biomed. Health Inform. (2019).
[30] V. Blanes-Vidal, G. Baatrup, E.S. Nadimi, Addressing priority challenges in the detection and assessment of colorectal polyps from capsule endoscopy and colonoscopy in colorectal cancer screening using machine learning, Acta Oncol. 58 (2019) S29–S36.
[31] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, et al., Real-time gastric polyp detection using convolutional neural networks, PLoS ONE 14 (2019) e0214133.
[32] M. Misawa, S. Kudo, Y. Mori, T. Cho, S. Kataoka, Y. Maeda, et al., Tu1990 Artificial intelligence-assisted polyp detection system for colonoscopy, based on the largest available collection of clinical video data for machine learning, Gastrointestinal Endoscopy 89 (2019), AB646–7.
[33] X. Zhu, D. Nemoto, Y. Wang, Q. Li, M. Aizawa, K. Utano, et al., Tu1943 A machine-learning algorithm detect diminutive colorectal polyps: a reading test comparing recognition with human endoscopists, Gastrointest. Endosc. 89 (2019) AB625.
[34] O.F. Ahmad, P. Brandao, S.S. Sami, E. Mazomenos, A. Rau, R. Haidry, Tu1991 artificial intelligence for real-time polyp localisation in colonoscopy withdrawal videos, Gastrointest. Endosc. 89 (2019) AB647.
[35] S. Sornapudi, F. Meng, S. Yi, Region-based automated localization of colonoscopy and wireless capsule endoscopy polyps, Appl. Sci. 9 (2019) 2404.
[36] T. Wittenberg, P. Zobel, M. Rathke, S. Mühldorfer, Computer aided detection of polyps in whitelight-colonoscopy images using deep neural networks, Curr. Direct. Biomed. Eng. 5 (2019) 231–234.
[37] T.K.L. Lui, K.K.Y. Wong, L.L.Y. Mak, M.K.L. Ko, S.K.K. Tsao, W.K. Leung, Endoscopic prediction of deeply submucosal invasive carcinoma with use of artificial intelligence, Endosc Int Open. 07 (2019) E514–E520.
[38] E. Ribeiro, A. Uhl, G. Wimmer, M. Häfner, Exploring deep learning and transfer learning for colonic polyp classification, Comput. Math. Methods Med. (2016), https://doi.org/10.1155/2016/6584725.
[39] M.F. Byrne, N. Chapados, F. Soudan, C. Oertel, M. Linares Pérez, R. Kelly, et al., Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model, Gut 68 (2019) 94–100.
[40] Y. Komeda, H. Handa, T. Watanabe, T. Nomura, M. Kitahashi, T. Sakurai, et al., Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience, Oncology 93 (Suppl 1) (2017) 30–34.
[41] P.-J. Chen, M.-C. Lin, M.-J. Lai, J.-C. Lin, H.H.-S. Lu, V.S. Tseng, Accurate classification of diminutive colorectal polyps using computer-aided analysis, Gastroenterology 154 (2018) 568–575.
[42] P. Kandel, R. LaLonde, V. Ciofoaia, M.B. Wallace, U. Bagci, Su1741 colorectal polyp diagnosis with contemporary artificial intelligence, Gastrointest. Endosc. 89 (2019) AB403.
[43] Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE; 2009. p. 248–55. doi:10.1109/CVPR.2009.5206848.
[44] Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. arXiv:14050312 [cs]. 2015. http://arxiv.org/abs/1405.0312. Accessed 1 Nov 2019.
[45] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, Vilariño F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians, Comput. Med. Imaging Graph. 43 (2015) 99–111.
[46] J. Silva, A. Histace, O. Romain, X. Dray, B. Granado, Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, Int J CARS. 9 (2014) 283–293.
[47] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, IEEE Trans. Med. Imaging 35 (2016) 630–644.
[48] J. Bernal, J. Sánchez, F. Vilariño, Towards automatic polyp detection with a polyp appearance model, Pattern Recogn. 45 (2012) 3166–3182.
[49] D. Vázquez, J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, A.M. López, A. Romero, et al., A benchmark for endoluminal scene segmentation of colonoscopy images, J. Healthcare Eng. 2017 (2017) 1–9.
[50] Angermann Q, Bernal J, Sánchez-Montes C, Hammami M, Fernández-Esparrach G, Dray X, et al. Towards Real-Time Polyp Detection in Colonoscopy Videos: Adapting Still Frame-Based Methodologies for Video Sequences Analysis. In: Cardoso MJ, Arbel T, Luo X, Wesarg S, Reichl T, González Ballester MÁ, et al., editors. Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures. Cham: Springer International Publishing; 2017. pp. 29–41. doi:10.1007/978-3-319-67543-5_3.
[51] P. Mesejo, D. Pizarro, A. Abergel, O. Rouquette, S. Beorchia, L. Poincloux, et al., Computer-aided classification of gastrointestinal lesions in regular colonoscopy, IEEE Trans. Med. Imaging 35 (2016) 2051–2063.
[52] Pogorelov K, Schmidt PT, Riegler M, Halvorsen P, Randel KR, Griwodz C, et al. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference - MMSys'17. Taipei, Taiwan: ACM Press; 2017. p. 164–9. doi:10.1145/3083187.3083212.
[53] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv:13112524 [cs]. 2014. http://arxiv.org/abs/1311.2524. Accessed 4 Nov 2019.
[54] Girshick R. Fast R-CNN. arXiv:150408083 [cs]. 2015. http://arxiv.org/abs/1504.08083. Accessed 4 Nov 2019.
[55] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:150601497 [cs]. 2015. http://arxiv.org/abs/1506.01497. Accessed 18 May 2018.
[56] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. arXiv:150602640 [cs]. 2016. http://arxiv.org/abs/1506.02640. Accessed 4 Nov 2019.
[57] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. arXiv:170306870 [cs]. 2018. http://arxiv.org/abs/1703.06870. Accessed 4 Nov 2019.
[58] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 2481–2495.
[59] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015. http://arxiv.org/abs/1409.1556. Accessed 5 Nov 2019.
[60] M. Everingham, S.M.A. Eslami, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes challenge: a retrospective, Int. J. Comput. Vis. 111 (2015) 98–136.
[61] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional Architecture for Fast Feature Embedding. In: Proceedings of the 22Nd ACM International Conference on Multimedia. New York, NY, USA: ACM; 2014. p. 675–678. doi:10.1145/2647868.2654889.
[62] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 2818–2226.
[63] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 770–778.
[64] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. USA: Curran Associates Inc.; 2012. p. 1097–1105. http://dl.acm.org/citation.cfm?id=2999134.2999257. Accessed 5 Nov 2019.
[65] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, et al., SSD: Single Shot MultiBox Detector, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 21–37.
[66] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, et al., Comparison of learning algorithms for handwritten digit recognition, in: International Conference on Artificial Neural Networks, 1995, pp. 53–60.
[67] Sabour S, Frosst N, Hinton GE. Dynamic Routing Between Capsules. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. USA: Curran Associates Inc.; 2017. p. 3859–3869. http://dl.acm.org/citation.cfm?id=3294996.3295142. Accessed 4 Nov 2019.
[68] Danelljan M, Bhat G, Khan FS, Felsberg M. ECO: Efficient Convolution Operators for Tracking. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. p. 6931–9.
[69] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (2019) 60.
[70] Ranking Popular Deep Learning Libraries for Data Science. KDnuggets. https://www.kdnuggets.com/ranking-popular-deep-learning-libraries-for-data-science.html/. Accessed 5 Nov 2019.
[71] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning Spatiotemporal Features With 3D Convolutional Networks. 2015. p. 4489–97. https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Tran_Learning_Spatiotemporal_Features_ICCV_2015_paper.html. Accessed 5 Nov 2019.
[72] D.K. Rex, C. Kahi, M. O'Brien, T.R. Levin, H. Pohl, A. Rastogi, et al., The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on real-time endoscopic assessment of the histology of diminutive colorectal polyps, Gastrointest. Endosc. 73 (2011) 419–422.

Alba Nogueira-Rodríguez received the BS and MS degrees in computer science from University of Vigo. Currently, she is a Ph.D. student of Computer Science at the SING group, where her research is focused on the application of deep learning for polyp detection and classification.

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123

Rubén Domínguez-Carbajales is information-systems technician at the Servizo Galego de Saúde (SERGAS). He received the BS and MS degrees in computer science from University of Vigo. Currently, He is a Ph.D. student of Computer Science at the SING group.

Hugo López-Fernández received the BS degree in computer science from University of Vigo. He received the MS and PhD degrees from University of Vigo, where he is Postodctoral Researcher in the Next Generation Computer Systems group (http://sing-group.org/). As of November 2019, his 41 scientific contributions indexed in Scopus hold 181 citations from 153 different documents (h-index = 7)

Águeda Iglesias studied BScN degree and MSc degree in Endocrinology and Nutricion, recieving both from University of Vigo. She has experience as a Nurse in Servizo Galego de Saude (SERGAS). Currently, she is working as Nurse and Data Manager of GIODO research group of the Instituto de Investigación Sanitaria Galicia Sur (IISGS).

Joaquín Cubiella (PhD) is a gastroenterologist working in Hospital Universitario de Ourense, Spain. He is the leader of the GIODO research group of the Instituto de Investigación Sanitaria Galicia Sur and his main areas of interest are diagnosis and prevention of colorectal cancer as well as diagnostic biomarkers and endoscopy. He is a member of the governing board of Asociación Española de Gastroenterología. He is involved in the development of clinical practice guidelines and in some of the highest impact research projects in the diagnosis and prevention of colorectal cancer: COLONPREV, COLONPREDICT, EPICOLON and EPoS trials.

Florentino Fdez-Riverola received the BS degree in computer science from University of Oviedo. He received the MS and PhD degrees from University of Vigo, where he is Professor in the Department of Computer Science and Coordinator of the Next Generation Computer System group (http://sing-group.org/). He is author of 210 scientific contributions, of which more than one hundred were published in journals indexed in the SCI. His publications hold 1581 citations from 1227 different documents.

Miguel Reboiro-Jato received the BS degree in computer science from University of Vigo. He received the MS and PhD degrees from University of Vigo, where he is Associate Professor in the Department of Computer Science and Header Researcher at the Next Generation Computer System group (http://sing-group.org/). As of November 2019, his 66 scientific contributions indexed in Scopus hold 583 citations from 478 different documents (h-index = 12).

Daniel Glez-Peña received the BS degree in computer science from University of Vigo. He received the MS and PhD degrees from University of Vigo, where he is Associate Professor in the Department of Computer Science and Header Researcher at the Next Generation Computer System group (http://sing-group.org/). As of November 2019, his 87 scientific contributions indexed in Scopus hold 733 citations from 636 different documents (h-index = 14).

Please cite this article as: A. Nogueira-Rodríguez, R. Domínguez-Carbajales, H. López-Fernández et al., Deep Neural Networks approaches for detecting and classifying colorectal polyps, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.02.123