Effective Deep Learning for Semantic Segmentation Based Bleeding Zone Detection in Capsule Endoscopy Images

Tonmoy Ghosh*, Linfeng Li, and Jacob Chakareski
Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35401, USA
*Email: tghosh@crimson.ua.edu

***Abstract*—Capsule endoscopy (CE) is a non-invasive way to detect small intestinal abnormalities such as bleeding. It provides a direct vision of the patients entire gastrointestinal (GI) tract. However, a manual inspection of the huge number of images produced thereby is tedious and lengthy, and thus prone to human errors. This makes automated computer assisted decision-making appealing in this context. This paper introduces a novel deep-learning based semantic segmentation approach for bleeding zone detection in CE images. A bleeding image features three regions labeled as bleeding, non-bleeding, and background. Thus, a convolutional neural network (CNN) is trained using SegNet layers with three classes. A given CE image is segmented using our training network and the detected bleeding zones are marked. The proposed network architecture is tested on different color planes and best performance is achieved using the hue saturation and value (HSV) color space. Experimental performance evaluation is carried out on a publicly available clinical dataset, on which our framework achieves 94.42% global accuracy and 90.69% weighted intersection over union (IoU), two state-of-the-art classification metrics. Performance gains are demonstrated over several recent state-of-art competiting methods in terms of all performance measures we examined, including mean accuracy, mean IoU, global accuracy and weighted IoU.**

***Index Terms*—bleeding detection, capsule endoscopy, deep learning, SegNet, convolutional neural network.**

I. Introduction

Bleeding is a very common symptom of many GI tract diseases such as vascular lesions (angiodysplasia), small bowel tumors, coeliac disease, and Crohns disease [1]. The best way to visualize the entire GI tract is capsule endoschopy, as it is non-invasive and capable of capturing images of the small intestine, where traditional endoscopies (e.g., colonoscopy and gastroscopy) cannot reach. However, a major challenge of CE is its reviewing process. The capsule is developed like a shape of vitamin pill that contains a camera, light source, wireless transmitter, and battery. Generally, after swallowing the capsule, it travels 8 hours in the GI tract during that time, capturing 50,000+ images. Reviewing all these images is a tedious error-prone task for a physician. Moreover, some bleeding regions are too small for naked-eye detection. All these difficulties motivate computer-aided programs for automated bleeding detection to assist the physicians.

In order to detect bleeding frame and bleeding regions of a capsule endoscopy images, the producer (Given Imaging that developed capsule endoscopy technologies in 2000) supplies a software kit to detect some diseases. The sensitivity and specificity of this software are very poor, which are reported as 37% and 59% respectively in [2]. Also, a word-based histogram method is presented in [3]. The k-means clustering is used to find the word dictionary to detect bleeding frame and the two-stage salient map is proposed to detect bleeding regions. Another method introduces super-pixel based bleeding segmentation [4], which is computationally complex. A region-of-interest based bleeding detection on YIQ color space is proposed in [5] and [6], while the performance of bleeding region detection is not satisfactory . Fully convolutional networks (FCNs) based bleeding segmentation method is developed in [7]. This method provides satisfactory performance. A cluster based bleeding zone detection method is proposed in [8], where authors used k-means clustering to segment bleeding zones. In order to detect small colon bleeding, HSV (hue intensity value) domain and support vector classifier are used in [9]. Green and red intensity ratio based bleeding detection is proposed in [10]. To detect a bleeding zone, statistical color features and support machine vector classifier are implemented in [11]. All these methods can detect bleeding images or bleeding areas, however, high accuracy detection is still missing. Since traditional machine learning approaches have their own limitations, deep learning could be potentially a more efficient alternative solution.

To overcome the above challenges, this paper introduces a novel deep-learning based framework for automated bleeding zone detection that advances the state-of-the-art. Our major contributions are the design of an efficient deep-learning architecture for accurate bleeding zone detection and its comprehensive assessment and comparison to the state-of-the-art, demonstrating consistent gains over multiple performance metrics. Concretely, we formulate an automatic bleeding zone detection method via SegNet based semantic segmentation. SegNet is a pixel-wise segmentation method, which works on a deep convolutional encoder-decoder architecture. First, a convolutional neural network is trained using SegNet layers. The training CE images are labeled with 3 classes: bleeding, non-bleeding, and background. To delineate bleeding zones, a given CE image is then segmented using this network. Bleeding zone detection accuracy is assessed on publicly available datasets and performance is compared to other established methods.

The rest of the paper is organized as follows. Our framework is described in Section II, the experimental assessment and analysis are carried out in Section III, and concluding remarks are provided in Section IV.

II. Proposed Method

The shape and size of the bleeding zones in the capsule endoscopy images are arbitrary in nature. Also, the location

This work has been supported in part by NSF award CCF-1528030.

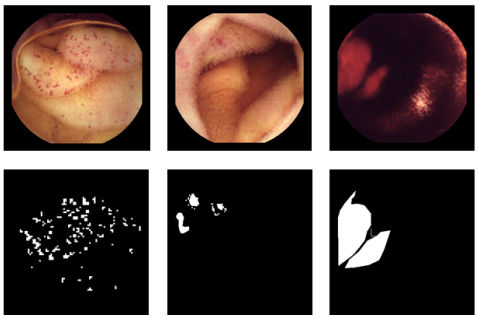


Fig. 1: Sample bleeding images with their corresponding ground truth

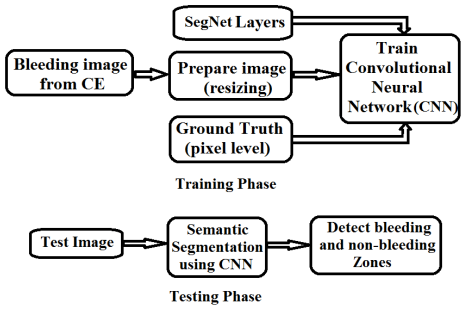


Fig. 2: Block diagram of the proposed method

of bleeding pixels is random. Moreover, bleeding zones are not purely red colored but in different shades of red and it also suffers from illumination changes over time. Three bleeding image samples from CE are presented in the Fig. 1. The first row illustrates actual bleeding images followed by their corresponding ground truth in the second row. In the first image, small bleeding zones are randomly distributed all over the images. In the second image, bleeding zones which exhibit on the top-left corner are hard to detect by naked eyes. The third image suffers illumination problem caused by poor contrast. All these difficulties make bleeding detection task more challenging. To overcome these difficulties, SegNet based semantic segmentation method using deep learning algorithm is proposed. The diagram of the proposed method is illustrated in Fig. 2. In the training phase, the prepared image, SegNet layers and ground truth (pixel level) are given as inputs to train convolutional neural network (CNN). Similarly, in the testing phase, a test image is segmented using that trained CNN and finally bleeding zones are detected.

Now-a-days deep learning is widely used in computed vision and image processing application. Deep learning considers as a computational models that are consisted of several processing layers to study representations of data with multiple levels of abstraction. Convolutional neural network is one of the deep learning tools. In order to detect bleeding zones, in this paper, a CNN based Semantic segmentation method is proposed. This method can understand the CE images at pixel level, which means the trained network can assign each pixel in the image to an object class. In 2012, a patch classification method was proposed by [12], which use patches to define the fully connected layers. However, it can only analyze the fixed size images. In 2014, a new CNN named Fully

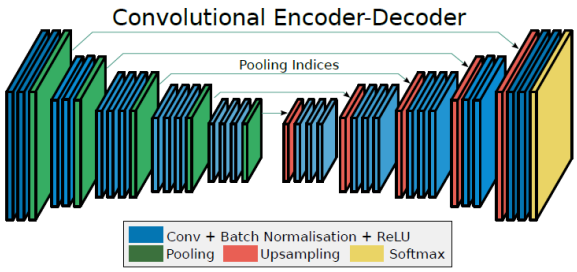


Fig. 3: The SegNet architecture used in bleeding detection [14]

Convolutional Network (FCN) was introduced in [13], which is a popular CNN architecture allowing the segmentation maps to be generated from different size of images. In 2015, a new architecture with encoder and decoder was proposed by many researchers. The one is called U-Net [13], which was introduced for biomedical image segmentation. The other one is called SegNet [14] and proposed algorithm is developed based on this method. Compared to FCN, SegNet is more time and memory efficient. The encoder will reduce the spatial dimension and the decoder will recover the image detail and the dimensions of pictures. For time efficiency, the SegNet method is chosen as the main semantic segmentation method.

A. *SegNet Architecture*

SegNet shares the same property with U-Net, which has an encoder network and a corresponding decoder network [14]. The output will be fed into the pixel-wise classification layer. The architecture for the CE bleeding detection is shown in Fig. 3. For the classification of bleeding, non-bleeding and background objects, VGG16 network [15] is used as the first 13 convolutional layers of the encoder network, each of these layers has a corresponding decoder layer. Thus, we have 13 encoder layers and 13 decoder layers. A huge amount of pictures can be resized to small pictures to decrease the training time. Besides, the fully connected layers will be discarded to keep the high-resolution feature maps at the deepest encoder output. The decoder output will be fed into the classifier to classify each pixel separately. Proposed SegNet architecture is illustrated in Fig. 3. No fully connected layers are used thus it is a CNN. The encoder generates the transferred pool indices then the input is up-sampled by the decoder to produce a sparse feature map. The trainable filter bank then helps to perform the convolution to densify the feature map. Finally, a soft-max classifier is fed by the decoder output feature maps for pixel-wise classification.

One of the main features of SegNet is to obtain the good results with time and memory efficiency, which is achieved in the efficient encoder network. For the encoder network, max-pooling and sub-sampling will be used to achieve translation invariance for the tiny spatial shifts of the input image, then produce a large spatial window for all the pixels. These processes will increase the lossy image representation at the boundaries, which are unacceptable for the segmentation. Thus, the boundary information should be fully stored before the sub-sampling. In practical applications, the encoder feature

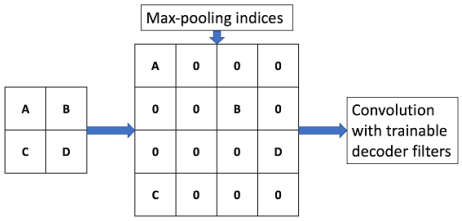


Fig. 4: SegNet Decoder. A, B, C, D are the values in feature map.

maps cannot be fully stored because of the memory limit. Therefore, for each encoder feature map, only the max-pooling indices which including the maximum features in each pooling window will be stored.

Another feature of SegNet is the trainable decoder, which is not available for the fully convolutional network (FCN). After generating the max-pooling indices from the encoders, the related decoder network is used to up-sample the input feature maps using the corresponding indices. The illustration is shown in Fig. 4. A, B, C, D are the values in feature map. The feature maps are up-sampled by using the max pooling indices then convoluted with a trainable decoder filter. The dense feature map then will be obtained from the filter. The decoder of SegNet can produce multiple channels instead of the R, G, B channels used by other networks.

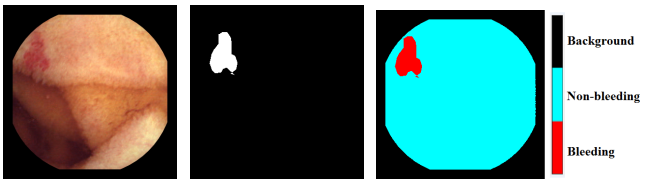
B. Training and Test

In order to precisely highlight the bleeding areas, a convolutional neural network is trained. A SegNet network layer is constructed using $256 \times 256 \times 3$ image size, three input classes and optimal number of encoder-decoder depth. Furthermore, a pre-trained VGG-16 network layer is also considered. A softmax layer is added to minimize the loss function. In the training stage, random image reflection and translation are performed to ensure that the proposed network can work on reflected and translated images. The network model is trained and tested on a single NVIDIA GeForce GTX1070, with momentum 0.9, initial learning rate 10^{-3}, maximum epochs 100, minimum batch size 3.

III. Simulation Results and Discussions

A. Dataset Description

The bleeding region detection performance of the proposed method is evaluated using a pixel-wise labeled of CE dataset. Capsule endoscopy images are collected from the publicly available dataset [16], [17], where pixel-wise annotations are offered. For the purpose of experimentation, 303 vascular, 27 angioectasias and 5 other bleeding images are considered. The pixels of these 335 bleeding images are labeled as bleeding, non-bleeding, and background. The black surrounding pixels around the border are annotated as background. The pixels within bleeding regions (dataset provides bleeding pixel ground truth) are labeled as bleeding and the remaining pixels are annotated as non-bleeding. Annotation results are shown in Fig. 5 , where black, cyan and red represent background, non-bleeding and bleeding regions respectively. The size of each image is $360 \times 360 \times 3$ pixels. During the training



(a) (b) (c)

Fig. 5: Illustration of annotation process; (a) actual CE image, (b) bleeding region ground truth, (c) after annotation

step, all the images are resized to $256 \times 256 \times 3$. Randomly selected $201(60\%)$ images are used to train convolution neural network using semantic segmentation layers and the remaining $134(40\%)$ images are selected for testing.

B. Performance Measures

The Performance of semantic segmentation task is measured using four metrics as follows. i) mean intersection over union (mean IoU), ii) weighted intersection over union (weighted IoU), iii) mean accuracy, iv) pixel accuracy. These four metrics are defined as

$$Mean\ IoU = \left(\frac{1}{n_{cl}}\right) \frac{\sum_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{1}$$

$$Weighted\ IoU = \left(\sum_k t_k\right) \frac{\sum_i t_i n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \tag{2}$$

$$Mean\ Accuracy = \left(\frac{1}{n_{cl}}\right) \sum_i \frac{n_{ii}}{t_i} \tag{3}$$

$$Global\ Accuracy = \frac{\sum_i n_{ii}}{\sum_i t_i} \tag{4}$$

where,

- n_{ii} = the number of pixels of class i predicted to belong to class i
- n_{ij} = the number of pixels of class i predicted to belong to class j
- t_i = the total number of pixels of class i
- n_{cl} = the total number of classes

C. Bleeding Zone Detection Performance

During the training phase of the CCN using SegNet layers, training accuracy over the iteration are plotted on the Fig 6. It is observed that accuracy increased quickly and converged, that illustrates the efficiency of the proposed network model. A confusion matrix of the bleeding detection is presented in Table I. A confusion matrix is a particular table arrangement that allows visualization of the performance of an algorithm. The instances in a predicted class are demonstrated by each row of the matrix while the instances in a actual class are presented by each column (or vice versa). Bleeding, non-bleeding and background pixel detection accuracy are found 81.08%, 90.02% and 98.96% respectively, which is very satisfactory. Although around 19% bleeding pixels are detected as non-bleeding pixels due to their very similar intensity characteristic with non-bleeding pixels. On the other hand, around 9.5% non-bleeding pixels are detected as bleeding, but for the clinical perspective, 9.5% false positive is acceptable.

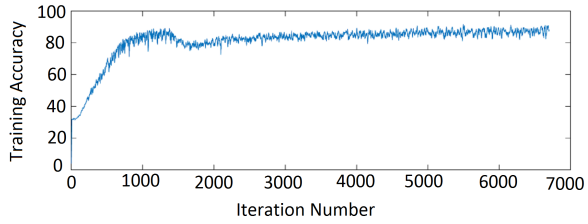


Fig. 6: Training accuracy during the training phase

TABLE I: CONFUSION MATRIX OF BLEEDING DETECTION

	Bleeding	Non-bleeding	Background
Bleeding	537,111 (81.08%)	125,165 (18.89%)	160 (0.02%)
Non-bleeding	1,064,309 (9.42%)	10,175,416 (90.02%)	63,768 (0.56%)
Background	10 (0.00%)	55,993 (1.04%)	5,344,468 (98.96%)

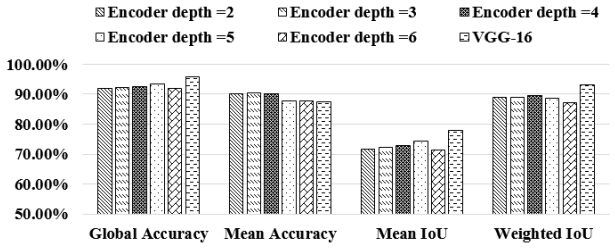


Fig. 7: Bleeding detection performance obtained using different network architecture

TABLE II: PERFORMANCE COMPARISON AMONG COLOR PLANES

	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU
RGB	92.46%	90.02%	72.50%	89.27%
Normalized RGB	93.19%	85.69%	72.31%	89.62%
HSV	94.42%	87.48%	75.63%	90.69%
Lab	94.58%	65.89%	63.20%	90.06%
YIQ	91.77%	89.58%	71.42%	88.18%

While training SegNet network the length of encoder and decoder is varied and the performance of each of the encoders is evaluated, which is presented in Fig. 7. The depth of the network encoder is varied from 2 to 6 and also VGG-16 network architecture is tested. Although slightly better performance is achieved using the VGG-16 network, it is observed that consistent performance is obtained in terms of the length of the encoder depth.

Bleeding detection performance is evaluated in different color space using the proposed method which is presented in table II. Among the color planes, the best global accuracy is obtained using Lab color space and the best mean accuracy is obtained using RGB. However, the best mean IoU and weighted IoU are achieved using HSV color space. While considering the performance of all the four measurements, the most favorable result is attained from HSV.

Moreover, to evaluate the quality of the proposed method, it is quantitatively compared with three state-of-art methods which are Fu et. al. [4], Yuan et al. [3] and Xiao et al. [7]. The method proposed in Fu et. al. incorporates super-pixel based bleeding segmentation. Yuan method uses two-stage saliency

TABLE III: PERFORMANCE COMPARISON AMONG DIFFERENT METHODS

Method	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU
Fu et. al. [4]	83.48%	82.44%	69.73%	70.48%
Yuan et. al. [3]	84.48%	83.65%	72.11%	73.27%
Xiao et. al. [7]	85.77%	85.80%	74.41%	75.32%
Proposed	94.42%	87.48%	75.63%	90.69%

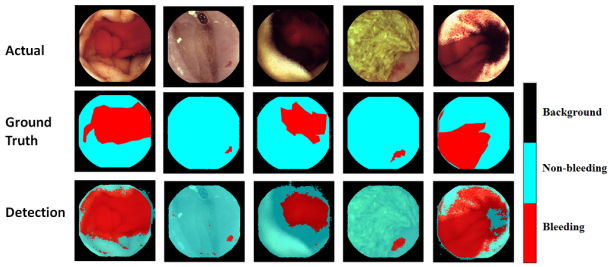


Fig. 8: Bleeding detection output

map extraction to detect bleeding regions. To segment bleeding pixels, fully convolutional networks are implemented by Xiao. The comparative performance results are demonstrated in table III. From the comparison, the proposed method can segment bleeding regions better than other three methods in terms of four performance metrics.

Segmentation results of bleeding detection are shown in Fig. 8, where five sample images are presented. Among the five bleeding images, active bleeding is shown in column 1, 2 and 3 and inactive bleeding is shown in column 2 and 4. In the case of active bleeding, the proposed method detect the bleeding regions with some false positive while on the other case of inactive bleeding, bleeding regions are detected more precisely. It is to be noted that the objective of this work is to assist physician in the reviewing process. The denoted bleeding regions reduce the burden of the physician and they can concentrate only on those regions. Therefore, we find that bleeding region segmentation can be achieved with SegNet network. This network can definitely help the physician to diagnose diseases.

IV. CONCLUSION

A novel deep-learning framework is presented for efficient automated bleeding zone detection in CE images. Bleeding zone detection is difficult due to its arbitrary shape and random locations. Moreover, bleeding zones are not purely red colored, but take on different shades of red. All these challenges motivate us to design an effective SegNet deep learning network architecture. We assess the accuracy of multiple color spaces in bleeding zone detection. The efficiency of our framework is validated experimentally on clinical data, observing 94.42% global accuracy, 87.48% mean accuracy, 75.63% mean IoU, and 90.69% weighted IoU, using the HSV color space. Compared to other established methods, performance gains are introduced of 8-10%, 1.5-5%, 1-6% and 15-20%, in global accuracy, mean accuracy, mean IoU and weighted IoU, respectively. We envision that our framework can considerably lessen the burden of clinicians and augment their detection accuracy of bleeding zones.

REFERENCES

[1] I. N. Figueiredo, S. Kumar, C. Leal, and P. N. Figueiredo, "Computer-assisted bleeding detection in wireless capsule endoscopy images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 1, no. 4, pp. 198–210, 2013.

[2] S. C. Park, H. J. Chun, E. S. Kim, B. Keum, Y. S. Seo, Y. S. Kim, Y. T. Jeen, H. S. Lee, S. H. Um, C. D. Kim *et al.*, "Sensitivity of the suspected blood indicator: an experimental study," *World Journal of Gastroenterology: WJG*, vol. 18, no. 31, p. 4169, 2012.

[3] Y. Yuan, B. Li, and M. Q.-H. Meng, "Bleeding frame and region detection in the wireless capsule endoscopy video," *IEEE journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 624–630, 2016.

[4] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," *IEEE journal of Biomedical and Health Informatics*, vol. 18, no. 2, pp. 636–642, 2014.

[5] T. Ghosh, S. Bashar, S. A. Fattah, C. Shahnaz, and K. Wahid, "A feature extraction scheme from region of interest of wireless capsule endoscopy images for automatic bleeding detection," in *2014 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2014, pp. 256–260.

[6] T. Ghosh, S. A. Fattah, S. Bashar, C. Shahnaz, K. Wahid, W.-P. Zhu, and M. O. Ahmad, "An automatic bleeding detection technique in wireless capsule endoscopy from region of interest," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, pp. 1293–1297.

[7] X. Jia and M. Q.-H. Meng, "A study on automated segmentation of blood regions in wireless capsule endoscopy images using fully convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 179–182.

[8] T. Ghosh, S. A. Fattah, K. A. Wahid, W.-P. Zhu, and M. O. Ahmad, "Cluster based statistical feature extraction method for automatic bleeding detection in wireless capsule endoscopy video," *Computers in Biology and Medicine*, vol. 94, pp. 41–54, 2018.

[9] M. A. Usman, G. B. Satrya, M. R. Usman, and S. Y. Shin, "Detection of small colon bleeding in wireless capsule endoscopy videos," *Computerized Medical Imaging and Graphics*, vol. 54, pp. 16–26, 2016.

[10] T. Ghosh, S. A. Fattah, and K. A. Wahid, "Automatic computer aided bleeding detection scheme for wireless capsule endoscopy (WCE) video based on higher and lower order statistical features in a composite color," *Journal of Medical and Biological Engineering*, vol. 38, pp. 482–496, 2018.

[11] S. Suman, A. S. Malik, M. Riegler, S. H. Ho, I. Hilmi, K. L. Goh *et al.*, "Detection and classification of bleeding region in wce images using color feature," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*. ACM, 2017, pp. 1–6.

[12] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, 2012, pp. 2843–2851.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: challenges for essential progress," *Nature Reviews Gastroenterology & Hepatology*, vol. 12, no. 3, pp. 172–186, 2015.

[17] A. Koulaouzidis, "Kid: Koulaouzidis-iakovidis database for capsule endoscopy," 2015.