A Deep Learning Approach for Red Lesions Detection in Video Capsule Endoscopies

Paulo Coelho[1(✉)], Ana Pereira[2], Argentina Leite[3], Marta Salgado[4], and António Cunha[3]

[1] Polytechnic Institute of Leiria, Leiria, Portugal
[2] University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
paulo.coelho@ipleiria.pt
[3] INESC TEC (Formerly INESC Porto) and UTAD – University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[4] Centro Hospitalar Porto, Porto, Portugal

Abstract. The wireless capsule endoscopy has revolutionized early diagnosis of small bowel diseases. However, a single examination has up to 10 h of video and requires between 30–120 min to read. Computational methods are needed to increase both efficiency and accuracy of the diagnosis. In this paper, an evaluation of deep learning U-Net architecture is presented, to detect and segment red lesions in the small bowel. Its results were compared with those obtained from the literature review. To make the evaluation closer to those used in clinical environments, the U-Net was also evaluated in an annotated sequence by using the Suspected Blood Indicator tool (SBI). Results found that detection and segmentation using U-Net outperformed both the algorithms used in the literature review and the SBI tool.

Keywords: Lesion detection · Gastrointestinal bleeding
Machine learning · Capsule endoscopy · Deep learning · U-Net

1 Introduction

Approximately 300,000 hospitalizations per year in the United States of America are associated with gastrointestinal bleeding and in 5% of those cases it is not possible to immediately identify the bleeding's source [14]. The small bowel is one of the major organs where bleeding from unknown sources occurs (also named as Obscure Gastrointestinal Bleeding - OGIB). Full and direct visualization of the small bowel is not possible through high endoscopy or colonoscopy, due to the organ's length and its morphological diversity [8]. To overcome this issue, direct visualization of the small bowel through endoscopic methods with emphasis to the Wireless Capsule Endoscopy (WCE), have greatly evolved in the last decades, revolutionizing the knowledge and clinical approach of several pathologies [10].

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 553–561, 2018.
https://doi.org/10.1007/978-3-319-93000-8_63

Recently, several Computer Aided Diagnostic (CAD) designs have been developed allowing for an automatic or semi-automatic lesion detection (e.g. polyps, ulcers, tumors and bleeding). Extensive reviews of these CAD systems can be found in [3,4,9]. Rapid Reader commercial software has been one of the most used diagnostic support tool since it provides, amongst other tools, the Suspected Blood Indicator (SBI), which identifies frames with possible red lesions in the gastrointestinal (GI) tract, based on color. However, several studies have shown that the results obtained using this tool are not completely satisfactory [1,4,6,16].

Computational methods for automatic image processing and analysis, such as smoothing filters, noise removal, contour detection or segmentation, can be used to facilitate the detection of anomalies/pathologies and to homogenize the response between different clinicians. Since 2012, Convolutional Neural Network (CNN), commonly known as "deep learning", started to present significantly better results than previous methods, automatically extracting characteristics from data and thus supporting new developments in CAD systems [9]. In this paper, an evaluation of deep learning U-Net architecture is presented for detecting and segmenting red lesions in the small bowel. Moreover, the comparison between its results and those found in the literature review is also presented.

2 Deep Learning Approach

CNN is a technology for learning generic resources in computational tasks that uses a hierarchy of computational layers and begins by mapping an input (image) to obtain an output (class). The lower layers are composed by convolution, normalization and pooling layers, alternating between each other, while the upper layers are fully connected and correspond to traditional neural networks [9].

The model used in this study is the U-Net architecture, proposed by Ronneberger et al. to segment images [7]. It is a CNN modification and presents a U shape due to symmetrical form presented by the contracting path (the left branch) and the expansive path (the right branch), as illustrated in Fig. 1.

Summarily, a repetitive pattern of convolution operation, followed by a Rectified Linear Unit (ReLU) and a down-sampling process – with a step of 2 – is performed in the contraction path. Regarding the expansive path, it includes an up-sampling operation of the previously obtained feature map, followed by a convolution (has the effect of halving the feature channels) and the concatenation with the characteristics map obtained in the contracting path. A final convolutional layer is added to map feature vectors to the desired number of classes [7]. The contraction path intents to capture context and the expanding path allows an accurate feature location.

3 Datasets and Experiments

3.1 Datasets

For this study it was necessary to compile 2 custom datasets with frames from the small bowel since, in the best of our knowledge, there is no publicly available

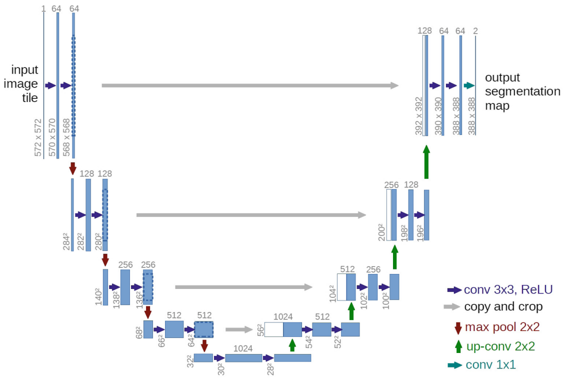

Fig. 1. The architecture of U-Net, reproduced from [7]. Each blue box in this figure represents a multi-channel feature map. The data pass through the horizontal lines simultaneously. The deep blue arrows represent activation functions. (Color figure online)

Table 1. Datasets characterization.

	Set 1	Set 2	Total
Lesion	1,131	439	1,570
No lesion	2,164	161	2,325
Total	3,295	600	

dataset with annotated red lesions in Video Capsule Endoscopy (VCE) images and with adequate size to train the U-Net architecture. The datasets characterization are presented in Table 1.

As criteria for compilation of Set 1, it was decided to have a dataset with images as diverse as possible - from different cameras, such as MiroCam, PillCam SB1, SB2 and SB3 - and with different red lesions, such as angioectasias, angiodysplasias, bleeding and others. It has 3,295 frames from which 1,131 have lesions. This set has a similar size to other aforementioned works [5]. All lesions were annotated manually. The images have 320 × 320 or 512 × 512 resolutions, although they were all resampled to 512 × 512 when applied to the U-Net.

Figure 2 presents a frame and the respective annotation mask example. The annotation process is subjective and very time consuming. For example, it is difficult for a human being to rigorously annotate the smooth border of blood diluted in small bowel fluids, which result in annotations with wide variability and impacts the segmentation results.

For Set 2, it was decided to have a dataset with a sequence of 600 images from a PillCam SB3 video to get an evaluation of the model closer to the clinical reality. The set contains 73% of frames with red lesions, each one labelled manually as Blood/Non-blood based on the human judgment for Ground Truth

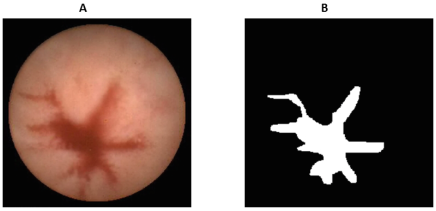

Fig. 2. a. Bleeding example; b. Manually annotated mask.

(GT) and on the result of the Given SBI to get an evaluation of the SBI tool in the set[1].

3.2 Evaluation Criteria

The metrics used to evaluate the U-Net performance in the detection process are derived from the basic cardinalities of the confusion matrix, namely the true positives (TP), the false positives (FP), the true negatives (TN) and the false negatives (FN) [11]. These measures assume that there is an overlap between two partitions, in this particular case the actual existence of red lesions in a given image and the possibility of this being predicted through the proposed method. From the aforementioned measures, one can obtain test validity indicators such as Accuracy - ACC (1), True Positive Rate (Sensitivity) - TPR (2), True Negative Rate (Specificity) - TNR (3):

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$TNR = \frac{TN}{TN + FP} \tag{3}$$

The ACC is defined as the portion of correctly classified elements to the total number of elements. TPR is a quantification of the algorithm capacity to correctly classify an image truly containing red lesions, i.e., it is the portion of frames with lesion that had a positive classifier result. Analogously, the TNR is a quantification of the algorithm capability in correctly classifying images truly without red lesions, i.e., it is the portion that non-lesion frames will be classified as normal by the classifier.

As segmentation metric the Dice Coefficient – DICE (4) was used [11]:

$$DICE = \frac{2 \cdot tp}{2 \cdot tp + fp + fn} \tag{4}$$

[1] At the time of submission these datasets were waiting for publication approval from the Ethical Council. In case of approval it will be available at https://rdm.inesctec.pt/dataset/nis-2018-003.

The Dice coefficient is a relative metric that provides a similarity index between predicted and ground truth segmentations. The tp are the total number of pixels belonging to the lesion in both masks: predicted and ground truth. The fp are the total number of pixels predicted as lesion but are not in the ground truth mask. The fn are the total number of pixels predicted as not belonging to lesion but are present in the ground truth mask.

3.3 Implementation Details

The U-Net network was trained from scratch with Set 1, which was split randomly in 80% for training and 20% for validation, to detect and segment red lesions. The training was made using Dice coefficient as cost function, in 3 cycles of 120 epochs with the Adam optimizer. The learning rate was 1E-4, 1E-5 and 1E-6 in each cycle, respectively. The model evaluation was performed by comparing its predictions with the annotated masks, used as ground truth, based on Sect. 3.2 evaluation metrics. The network was implemented in Python 2.7 and all experiments were performed on a machine with an Intel Xeon CPU E5-2650 and 64 GB RAM. The U-Net was implemented using Keras with TensorFlow as backend and was accelerated on an NVIDIA GTX-1080Ti GPU (11 GB on-board memory).

4 Results and Discussion

The SBI tool and most of state of the art methods are used as blood detectors. So, it was decided to evaluate the U-Net model - trained on Set 1 for segmentation - as a detector and the results compared with state of the art bleeding detection with datasets greater than 1,000 images, as showed in Table 2. The authors point out that the relative comparison with other works is carried out under different experimental conditions, since the datasets are different.

Table 2. Comparison between U-Net trained on Set 1 and state of the art bleeding detection.

Author(s)	Images/Patients	ACC(%)	TPR(%)	TNR(%)
Sainju et al. [8]	1,500/3	93.00	96.00	90.00
Figueiredo et al. [2]	4,000/10	92.70	92.90	>90.00
Usman et al. [13]	8,500	92.00	94.00	91.00
Xiong et al. [15]	3,596/5	94.10	91.69	94.59
U-Net	3,295/>5	95.88	99.56	93.93

The U-Net model learned very well to detect red lesions, with only 1 FN frame and 26 FP. Indeed, it has a very good accuracy (ACC = 95.88%), sensitivity (TPR = 99.56%) and specificity (TNR = 93.93%), outperforming Xiong et al.

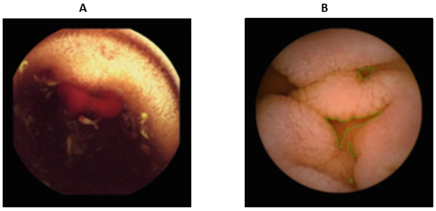

Fig. 3. a. FN frame; b. FP frame example. (Color figure online)

work (the overall most precise present in the literature review) by 1.78% in accuracy and in 7.87% in sensitivity. The specificity is lower 0.66% from the Xiong et al work, but is still higher than other state of the art works. From the analysis of the FN frames it was verified that these occur in lesions in which the background presents similar colors as the one shown in Fig. 3a.

In the case of the 26 FP, in 15 of them the system predicted very small areas that could be ignored, 12 of them appearing in between intestinal folds. Of the remaining 11 cases, 7 are dubious, even after the system's prediction result, since they were manually annotated as not containing red lesions. However, the U-Net considers them as having lesions and, in fact, it can be considered as correct. Finally, the last 4 cases present one or several considerable areas, also located in intestinal folds as can be seen in Fig. 3b.

The segmentation metric for red lesions was obtained from the evaluation of the Set 1 TP frames by averaging the Dice coefficient (DICE = 87.08%). This rate value is biased by the human manual ground truth segmentation in the smooth border of diluted blood in small bowel fluids. In the literature, this result can be compared with the study presented by Tuba et al. [12], that presents an average value for DICE of 84%. In this case, the U-Net outperforms it in 3.08%.

According to Yung et al. [16], SBI showed high sensitivity (TPR = 98.8%) but with low specificity (TNR = 64.0%) even for clinical scenarios for active bleeding. For small bowel pathology with bleeding potential, it shows moderate sensitivity (TPR = 55.3%) and specificity (TNR = 57.8%). To get a fair comparison, the model was evaluated with 600 consecutive new frames belonging to the small bowel and compared with SBI, Set 2, as can be seen in Table 3.

With 4 FN frames and 15 FP, the U-Net obtained a very good accuracy (ACC = 96.83%) and an excellent sensitivity (TPR = 99.09%), much better than SBI, with 185 FN frames. U-Net outperforms SBI by 27.66% in accuracy and 41.23% in sensitivity, but in specificity, it underperforms SBI (without FP frames), by 9.32%. The 15 FP that have been wrongly marked contain small areas in intestinal folds.

Figure 4 presents an interesting view of U-Net's performance for clinical application. The chart shows the Set 2 sequence with the area (total of pixels) of the segmented bleeding lesions plotted in black and in background the frames with bleeding are represented in red - for the ground truth - and in blue for the SBI tool.

Table 3. Comparison between U-Net (trained on Set 1) and SBI, when applied to Set 2.

	ACC (%)	TPR (%)	TNR (%)
SBI	69.17	57.86	100.00
U-Net	96.83	99.09	90.68

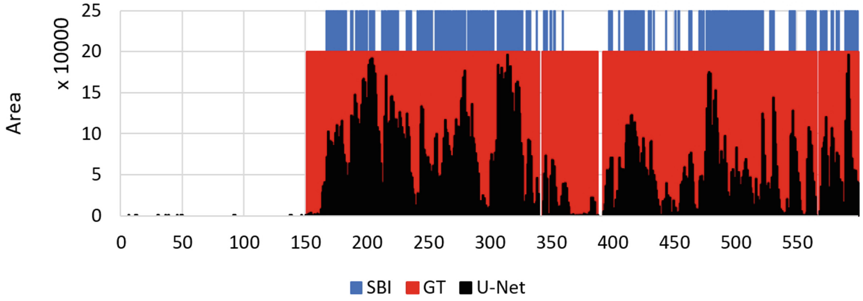

Fig. 4. Set 2: chart of U-Net segmented area versus blood detection in ground truth and SBI. (Color figure online)

In the first 150 sequence frames there are no red lesions. It can be seen that there are some frames with very small areas (black dots) that are detected by the U-Net but unrecognized as blood by the human. The SBI does not detect any red lesion until approximately frame 160, were U-Net are about 1E4 pixels.

After frame 150, the human marked almost all the frames with red lesion, as it can be seen by the continuous red color. The U-Net area has a very good match with GT, as expected (ACC = 96.83%) and also translates the amount of blood in each frame. It can be seen a coherent continuity of blood amount along the sequence and that in the few frames without blood in the interval, around frames 340 and frames 390. The SBI show lots of false negative red lesions, noticed in the frequent discontinuation of blue color, in accordance with SBI ACC = 69.17%. From the chart it seems like the SBI is tuned to identify red lesions without false positives, as the TNR = 100% indicates.

Thus, it can be stated that the U-Net model did very well in the detection and segmentation of bleeding in videos of endoscopy and presents high potential to be useful in clinical environment.

5 Conclusion

In this paper, the U-Net model was evaluated in detecting and segmenting red lesions in endoscopy videos. The U-Net model learned very well to detect red lesions, outperforming the works that showed the best results in the state of the art by 1.78% in accuracy and in 7.87% in sensitivity and having a specificity lower by 0.66%. It was evaluated in a sequence of images and compared with SBI achieving much better accuracy and excellent sensibility, much better than SBI (more 27.66% and 41.23%, respectively). The SBI got a specificity of 100%, 9.32% better than the U-Net model. Thus, the U-Net model had an excellent performance in the detection and segmentation of red lesions in endoscopy videos, presenting an high potential to be useful in clinical environments.

Acknowledgments. This work is financed by the ERDF – European Regional Development Fund through the Operational Program for Competitiveness and Internationalization - COMPETE 2020 Program within project «POCI-01-0145-FEDER-006961», and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia as part of project «UID/EEA/50014/2013».

References

1. Buscaglia, J.M., Giday, S.A., Kantsevoy, S.V., Clarke, J.O., Magno, P., Yong, E., Mullin, G.E.: Performance characteristics of the suspected blood indicator feature in capsule endoscopy according to indication for study. Clin. Gastroenterol. Hepatol. **6**(3), 298–301 (2008). http://linkinghub.elsevier.com/retrieve/pii/S1542356507012062
2. Figueiredo, I.N., Kumar, S., Leal, C., Figueiredo, P.N.: Computer-assisted bleeding detection in wireless capsule endoscopy images. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **1**(4), 198–210 (2013). http://www.tandfonline.com/doi/abs/10.1080/21681163.2013.796164
3. Iakovidis, D.K., Koulaouzidis, A.: Software for enhanced video capsule endoscopy: challenges for essential progress. Nat. Rev. Gastroenterol. Hepatol. **12**(3), 172–186 (2015). http://dx.doi.org/10.1038/nrgastro.2015.13%5Cn10.1038/nrgastro.2015.13
4. Koulaouzidis, A., Iakovidis, D.K., Karargyris, A., Plevris, J.N.: Optimizing lesion detection in small-bowel capsule endoscopy: from present problems to future solutions. Expert Rev. Gastroenterol. Hepatol. **9**(2), 217–235 (2015)
5. Koulaouzidis, A., Iakovidis, D.K., Yung, D.E., Rondonotti, E., Kopylov, U., Plevris, J.N., Toth, E., Eliakim, A., Wurm Johansson, G., Marlicz, W., Mavrogenis, G., Nemeth, A., Thorlacius, H., Tontini, G.E.: KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes. Endosc. Int. Open **5**(6), E477–E483 (2017). http://www.thieme-connect.de/DOI/DOI?10.1055/s-0043-105488
6. Park, S.C., Chun, H.J., Kim, E.S., Keum, B., Seo, Y.S., Kim, Y.S., Jeen, Y.T., Lee, H.S., Um, S.H., Kim, C.D., Ryu, H.S.: Sensitivity of the suspected blood indicator: an experimental study. World J. Gastroenterol (WJG) **18**(31), 4169–4174 (2012)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. Med. Image Comput. Comput. Assisted Interv. (MICCAI) **15**(1), 348–356 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Sainju, S., Bui, F.M., Wahid, K.A.: Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. J. Med. Syst. **38**(4), 25 (2014). http://link.springer.com/10.1007/s10916-014-0025-1
9. Seguí, S., Drozdzal, M., Pascual, G., Radeva, P., Malagelada, C., Azpiroz, F., Vitrià, J.: Generic feature learning for wireless capsule endoscopy analysis. Comput. Biol. Med. **79**, 163–172 (2016). http://linkinghub.elsevier.com/retrieve/pii/S0010482516302712

10. Spada, C., Hassan, C., Munoz-Navas, M., Neuhaus, H., Deviere, J., Fockens, P., Coron, E., Gay, G., Toth, E., Riccioni, M.E., Carretero, C., Charton, J.P., Van Gossum, A., Wientjes, C.A., Sacher-Huvelin, S., Delvaux, M., Nemeth, A., Petruzziello, L., de Frias, C.P., Mayershofer, R., Aminejab, L., Dekker, E., Galmiche, J.P., Frederic, M., Johansson, G.W., Cesaro, P., Costamagna, G.: Second-generation colon capsule endoscopy compared with colonoscopy. Gastrointest. Endosc. **74**(3), 581–589 (2011). http://dx.doi.org/10.1016/j.gie.2011.03.1125
11. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging **15**(1) (2015). https://doi.org/10.1186/s12880-015-0068-x
12. Tuba, E., Tuba, M., Jovanovic, R.: An algorithm for automated segmentation for bleeding detection in endoscopic images. In: International Joint Conference on Neural Networks (IJCNN), pp. 4579–4586. IEEE, May 2017. http://ieeexplore.ieee.org/document/7966437/
13. Usman, M.A., Satrya, G., Usman, M.R., Shin, S.Y.: Detection of small colon bleeding in wireless capsule endoscopy videos. Comput. Med. Imaging Graph. **54**, 16–26 (2016). https://doi.org/10.1016/j.compmedimag.2016.09.005
14. Wilcox, C.M., Cryer, B.L., Henk, H.J., Zarotsky, V., Zlateva, G.: Mortality associated with gastrointestinal bleeding events: comparing short-term clinical outcomes of patients hospitalized for upper GI bleeding and acute myocardial infarction in a US managed care setting. Clin. Exp. Gastroenterol. **2**, 21–30 (2009). http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108636/
15. Xiong, Y., Zhu, Y., Pang, Z., Ma, Y., Chen, D., Wang, X.: Bleeding detection in wireless capsule endoscopy based on MST clustering and SVM. In: IEEE Workshop on Signal Processing Systems (SiPS), vol. 35, pp. 1–4. IEEE, October 2015. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7345001
16. Yung, D.E., Sykes, C., Koulaouzidis, A.: The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. Expert Rev. Gastroenterol. Hepatol. **11**(1), 43–51 (2017). https://www.tandfonline.com/doi/full/10.1080/17474124.2017.1257384