Abnormality Detection in Wireless Capsule Endoscopy Images Using Deep Features

Daniel G. P. de Sá(✉), Giulia de A. Freulonx, Marcio P. Ferreira, Alexandre C. P. Pessoa, Darlan B. P. Quintanilha, and Aristófanes C. Silva

Applied Computing Group (NCA - UFMA), Federal University of Maranhão, São Luís, MA, Brazil
{daniel.piorsky,giulia.freulon}@discente.ufma.br,
{marcio.ferreira,alexandre.pessoa,dquintanilha,ari}@nca.ufma.br

Abstract. The capsule endoscopy examination is a common medical procedure used to diagnose and treat gastrointestinal tract diseases without the need for invasive procedures. Images captured during the examination can reveal a wide range of abnormalities, including lesions, inflammation, ulcers, bleeding, and tumors. However, interpreting these images can be a challenge for physicians since the videos contain a large number of frames (images) to be analyzed. To attempt to achieve an early diagnosis and reduce the lethality of gastrointestinal system pathologies, the use of artificial intelligence has been extensively studied to alleviate the workload of healthcare professionals, as the large number of images resulting from an examination makes manual categorization of each image challenging. This work studied the use of machine learning methods such as OneClassSVM and XGBoost based on features extracted from deep neural networks and compared them to traditional convolutional neural network methods, such as the ResNet152 network. The Kvasir-Capsule and ERS datasets were used to evaluate the proposed methods, focusing on classifying images as normal or abnormal. Among the evaluated methods, XGBoost showed the best results among others, with a weighted F1-score of 0.71 on the ERS dataset and 0.87 on the Kvasir-Capsule dataset. The class imbalance in both datasets proved to be a continuous challenge, adding to the challenge of the low quantity images in the ERS dataset.

Keywords: Endoscopy · Deep Features · One-Class · XGBoost

1 Introduction

A endoscopy represents one of the procedures used for the diagnosis of gastrointestinal diseases. Usually conducted through the ingestion of a tube equipped with a camera (endoscope), this method produces a sequence of images that experts in the field have the ability to classify as indicative of health or adverse conditions. Various forms of diseases can be identified through this technique [7].

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 173–184, 2024.
https://doi.org/10.1007/978-3-031-60665-6_13

In contrast, Wireless Capsule Endoscopy (WCE), a more recent innovation, offers a non-invasive, patient-friendly approach to meticulously visualizing the gastrointestinal tract. Compared to the traditional approach of endoscopy, which involves the insertion of a flexible tube through the mouth, capsule endoscopy offers several significant advantages. This includes reduced invasiveness, increased patient comfort, more comprehensive imaging, a reduced risk of complications, and the ability to access hard-to-reach areas [2,28].

The capsule endoscope is a small capsule that contains a camera and is swallowed by the patient, allowing the doctor to visualize the gastrointestinal tract. During the examination, the capsule endoscope captures hundreds of images that are later reviewed by the physician to detect abnormalities. The result of this endoscopic examination is a long video of the entire gastrointestinal system of the patient. Such a video results in a relatively large number of images that can depict either normal tissue or tissue with some pathology [12].

The lethality of various diseases in the human gastrointestinal system can be significantly reduced with early diagnosis. That is, the earlier diseases such as colorectal cancer are detected, the lower the risk of mortality and permanent sequelae for patients [8].

Therefore, the use of machine learning techniques and deep neural networks to address the binary classification problem of pathologies becomes essential in assisting healthcare professionals in their work. Due to the large quantity of images generated by exams such as capsule endoscopy, manually verifying each image by a trained physician becomes slow and impractical. Thus, models created by these techniques serve to automate anomaly detection and achieve early diagnosis.

One approach to address this challenge is the application of artificial intelligence algorithms to assist in the identification of anomalies in capsule endoscopy videos, as manual analysis of these images by doctors can be exhaustive due to their large quantity. The use of these techniques to support healthcare professionals in early disease diagnosis is already widely explored by the scientific community, as highlighted in recent studies [15,22]. As a result, several approaches and datasets have been proposed to achieve this goal.

Recent research has focused on the development of robust binary classification models for capsule endoscopy images. [11] utilized the Kvasir-Capsule database [27] in conjunction with deep learning techniques to reduce false negatives, a crucial aspect due to the severity associated with non-detection of gastrointestinal diseases. [13] utilizes fractal dimension for feature extraction and a random forest classifier to detect abnormal frames.

O WCENet [14], a deep convolutional neural network model, classifies and segments anomalous regions in WCE images into four categories (polyp, vascular, inflammatory, or normal). It achieves an accuracy of 98% and an area under the ROC curve of 99%, outperforming nine conventional machine learning and deep learning models on the KID dataset. This performance highlights its potential clinical use.

The performance of eight deep learning-based models for polyp detection and classification is compared in [16]: Faster RCNN [25], YOLOv3 [24], YOLOv4 [4],

SSD [20], RetinaNet [18], DetNet [17], RefineDet [30], and ATSS [29]. The results indicate that the RefineDet model achieved the best performance in polyp detection, with an F1-score of 88.6.

In the study by [21], an intelligent approach is proposed for classifying alimentary canal diseases such as Barrett's, Esophagitis, Hemorrhoids, Polyps, and Ulcerative colitis. The method involves image preprocessing, the application of Empirical Wavelet Transform (EWT) for extracting distinct patterns, and a two-stage classification using deep Convolutional Neural Networks (CNNs). The results show 96.65% accuracy in abnormal image detection and 94.25% accuracy in classifying these images into specific diseases.

Notably, the state-of-the-art works in this field have primarily focused on binary classification and have not evaluated anomaly detection through deep feature extraction or one-class classification methods. This study aims to develop a method for anomaly detection using video capsule endoscopy images by utilizing features extracted from convolutional neural networks and one-class classification method. The capsule endoscopy, coupled with machine learning techniques, plays a pivotal role in early diagnosis of gastrointestinal diseases, reducing their lethality and improving patient outcomes.

2 Materials and Methods

This section comprises the description of the procedures used in this study, including image acquisition, preprocessing, the machine learning techniques employed, loss function, and evaluation metrics.

2.1 Dataset

In the development of this study, two distinct datasets were used. The first one is the Kvasir-Capsule dataset [27], consisting of images captured by capsules ingested by patients. This dataset includes a total of 47,238 images from 117 videos. Out of the collected images, 34,338 were labeled as "Normal Clean Mucosa", referring to images with little or no fluid and mucosa with healthy villi and no pathological findings, representing the normal class. The remaining 12,900 images were labeled in the following categories: "Foreign Body", "Polyp", "Ulcer", "Erosion", "Blood - Hematin", "Blood - Fresh", "Angiectasia", "Erythema", "Lymphangiectasia", "Reduced Mucosal View", "Ileocecal Valve", "Ampulla of Vater", and 'Pylorus", representing the abnormal class.

The second dataset is the ERS dataset [6], also intended for multi-label classification. This dataset includes 123 labels divided into 5 categories: "Gastro", "Colono", "Healthy", "Blood", and "Quality", and these labels were assigned according to the Minimal Standard Terminology 3.0 (MST 3.0) standard [1]. The dataset consists of approximately 6,000 precisely labeled images and about 115,000 imprecise images, collected from 1,520 VCE videos of 1,135 patients. The precisely labeled images were classified by medical professionals from the Medical University of Gdańsk [6].

This study used only the precise images, with 1,019 labeled as normal and 2,494 labeled as abnormal. The low number of precise images poses a challenge for training, which adds to the challenge of class imbalance between the normal and abnormal classes.

2.2 Preprocessing

The images were resized to a size of 224×224 pixels, using their 3 RGB color channels. Subsequently, image normalization was performed using the Min-Max method to ensure that the pixel values ranged from 0 to 1.

2.3 Extraction of Deep Features

The analysis of medical images is a complex task due to the diversity and complexity of the information contained in these images. Visual features extracted from images play a crucial role in classification, anomaly detection, and disease diagnosis. However, in many cases, low-level features are not sufficient to capture the complexity of medical information [19].

To address this limitation, deep learning features, also known as "deep features", emerge as a solution. These features are intermediate representations learned by deep neural networks during training on extensive datasets. They encode hierarchical and abstract information about the objects present in the images, making them more discriminative and informative when compared to traditional features [10].

In this context, the ResNet152 network [9] was chosen for deep feature extraction. Despite its extensive depth with 152 layers, which can typically result in performance degradation issues, the residual connections embedded in the architecture, allowing for the direct passage of information between convolutional layers, help mitigate this problem. The choice of ResNet152 in this study is based on its superior performance compared to other evaluated architectures, making it the ideal choice for deep feature extraction.

Additionally, to address class imbalance, the "Binary Focal Loss" loss function was used. This function was specially designed for binary classification tasks, offering an effective solution for situations where classes are not equally distributed, i.e., one class occurs more frequently compared to the other. It adjusts the training focus to the most challenging examples, considering the correct class probability (p_t), an adjustment factor (α_t) to balance the classes, and a modulation parameter (γ). These elements together allow the model to prioritize examples that are more difficult to classify correctly, improving its ability to deal with class imbalance and, consequently, enhancing the accuracy of the final classification. The "Binary Focal Loss" equation is given by:

$$\text{Binary Focal Loss}(p_t) = -\alpha_t \cdot (1 - p_t)^{\gamma} \cdot \log(p_t) \tag{1}$$

2.4 Anomaly Detection

One-Class methods are a class of machine learning algorithms specifically designed to address problems where there is a predominant class (normal class) and a minority class of interest (anomalies). They play a crucial role in situations where identifying uncommon patterns is essential, such as in the analysis of medical images for early disease detection and anomaly identification.

The fundamental concept of One-Class methods is to create a model that learns only from examples of the normal class. This approach is based on the assumption that the normal class is well represented in the dataset and that anomalies are rare and different from this class. The main objective is to establish a boundary or threshold that encompasses the normal class, identifying any examples that fall outside this threshold as anomalies [23].

In this work, the One-Class Support Vector Machine (One-Class SVM) method was used [26]. Its operation involves creating a hyperplane that separates the data of the class of interest from regions considered not to belong to that class.

Given a training dataset $X = x_1, x_2, \ldots, x_n$, where x_i represents an example, the goal is to find a separation hyperplane that maximizes the margin around the positive class. Training examples are mapped to a high-dimensional space using a kernel function. The most common kernel used in OC-SVM is the Gaussian kernel (RBF - Radial Basis Function). The One-Class SVM solves the following optimization problem [26]:

$$\min \frac{1}{2}|w|^2 - \nu \sum_{i=1}^{n} \xi_i \tag{2}$$

subject to $\langle w, \phi(x_i) \rangle \geq \rho - \xi_i$ for $i = 1, 2, \ldots, n$, where:

- w is a weight vector defining the separation hyperplane;
- ν is a hyperparameter that controls the amount of data that can fall into the margin region;
- ξ_i are slack variables that allow some examples to be within the margin;
- $\phi(x_i)$ represents the transformation of data into the high-dimensional space;
- ρ is the distance from the hyperplane to the nearest point of the positive class.

Consequently, the One-Class SVM detects any deviation from the "normal" class trained as belonging to an "abnormal" class. This makes it a valuable tool in real-world scenarios where completely new images not present in the training dataset need to be analyzed, enabling effective anomaly identification.

For the purpose of comparing the results of the One-Class method, which exclusively trains with the normal class, with another machine learning approach that uses two classes during the training process, XGBoost algorithm was employed. This choice will allow for the evaluation of performance differences between these two methods regarding the specific task at hand.

XGBoost [5] is a machine learning algorithm based on decision trees that excels in medical image classification. Unlike traditional approaches that use

CNNs, XGBoost employs a set of decision trees to predict an image's class. It can capture complex patterns and nonlinear interactions in the data, making it a powerful option for identifying discriminative features in images. XGBoost is particularly effective in cases where the dataset is unbalanced among its classes and is also known for its interpretability, allowing medical professionals to understand how the model arrives at its decisions.

3 Results

This section presents and discusses the results obtained with the proposed method for abnormality detection using video capsule endoscopy images. For evaluation, the datasets described were divided into three sets: training, validation, and test, ensuring that images from the same video captured from a patient were present in only one of the sets. Tables 1 and 2 present the proportion of images in the three sets for the Kvasir Capsule and ERS datasets, respectively.

Table 1. Division of the Kvasir Capsule dataset for evaluating the proposed method.

Pathology	Total	Training	Validation	Test
Normal	34338	18130	8315	7893
Abnormal	6659	3259	2461	939

Table 2. Division of the ERS dataset for evaluating the proposed method.

Pathology	Total	Training	Validation	Test
Normal	1019	761	125	133
Abnormal	2494	1590	384	520

The results here shown come from in-dataset scenarios, with no cross-dataset scenarios happening.

3.1 Deep Feature Extraction

The extraction of deep features was conducted using the ResNet152 architecture, which was initially pre-trained with ImageNet dataset weights. This pre-training provides the network with the ability to acquire useful representations of general visual features, making it valuable for computer vision tasks. After this pre-training step with the ImageNet dataset, ResNet152 was fine-tuned using the Kvasir-Capsule and ERS datasets with the goal of classifying images as normal or abnormal.

During the training process, Data Augmentation was employed to regularize the model, preventing overfitting. Various transformations such as rotation,

translation, flipping, shearing, and scaling were applied to enrich the variety and robustness of the training data.

To optimize the model's hyperparameters, the Hyperopt optimizer was used to maximize the F1-score metric. For the Kvasir-Capsule dataset, the best hyperparameters for the Binary Focal Loss were set to $\alpha = 0.35$ and $\gamma = 2.0$. For the ERS dataset, the optimal hyperparameters were set to $\alpha = 0.2$ and $\gamma = 2.6$. In both implementations, the Adam optimizer was used, with a learning rate of 1×10^{-6}. This approach played a crucial role in effectively fine-tuning the model according to the specific characteristics of each dataset, resulting in the maximization of its performance.

Tables 3 and 4 display the classification results of images into normal and abnormal categories using the ResNet152 architecture on two distinct image datasets. In both sets of results, remarkable accuracy is observed in the classification of classes that contain a larger volume of images. This means that the model performed well on classes for which it received more examples during training, namely the "normal" class in the Kvasir-Capsule dataset and the "abnormal" class in the ERS dataset.

Table 3. Results of standard training of ResNet152 with the test split of the Kvasir-Capsule dataset.

Class	F1-Score	Precision	Recall
Normal	0.87	0.89	0.85
Abnormal	0.11	0.10	0.14
Average	0.49	0.50	0.49
Weighted Average	0.79	0.81	0.78

Table 4. Results of standard training of ResNet152 with the test split of the ERS dataset.

Class	F1-Score	Precision	Recall
Normal	0.25	0.25	0.25
Abnormal	0.81	0.81	0.81
Average	0.53	0.53	0.53
Weighted Average	0.69	0.69	0.69

Once the best models were defined for each dataset, the classification layer of the ResNet152 architecture was removed, allowing for the exclusive extraction of deep features.

3.2 Abnormality Detection

In the Kvasir-Capsule dataset, the hyperparameters used in the OneClassSVM method remained consistent with those used during standard training.

However, when applying the XGBoost algorithm, hyperparameters were tuned to `max_depth` = 6, `learning_rate` = 0.25, and `n_estimators` = 5000.

In the ERS dataset, the hyperparameters used in the OneClassSVM method remained consistent with those used during standard training. However, when applying the XGBoost algorithm, hyperparameters were tuned to `max_depth` = 6, `learning_rate` = 0.2, and `n_estimators` = 3560.

All the hyperparameters discussed in the implementations were obtained through the use of the Hyperopt hyperparameter optimizer [3]. It is worth noting that the hyperparameters used in the OneClassSVM method for both datasets remained unchanged. This behavior is related to the fact that the model does not correctly classify the two classes but instead assumes that all images belong to only one category.

Tables 5 and 6 display the best results obtained with the application of the OneClassSVM method on both datasets. However, it is important to highlight that the result in the ERS dataset suggests that the model is essentially classifying all images as belonging to the abnormal class. One of the reasons for this situation may be the class imbalance, with a larger number of abnormal images compared to normal images, which can affect the model's ability to correctly identify normal images.

Table 5. Results of applying the OneClassSVM with the test split of the Kvasir-Capsule dataset.

Class	F1-Score	Precision	Recall
Normal	0.61	0.82	0.49
Abnormal	0.38	0.28	0.64
Average	0.50	0.55	0.56
Weighted Average	0.56	0.69	0.52

Table 6. Results of applying the OneClassSVM with the test split of the ERS dataset.

Class	F1-Score	Precision	Recall
Normal	0.00	0.00	0.00
Abnormal	0.89	0.80	1.00
Average	0.44	0.40	0.50
Weighted Average	0.71	0.63	0.80

Tables 7 and 8 present the best results obtained with the application of the XGBoost method on both datasets. It is observed that, contrary to the problem identified in the application of the OneClassSVM method with the ERS dataset, XGBoost does not exhibit the same behavior of labeling all images as belonging to the majority class in the training set. However, there is still a noticeable

tendency for the model to assume that images belong to the majority class during prediction. In the Kvasir-Capsule dataset, the model tends to classify more images as normal, while in the ERS dataset, it tends to classify more images as abnormal.

Table 7. Results of applying XGBoost with the test split of the Kvasir-Capsule dataset.

Class	F1-Score	Precision	Recall
Normal	0.94	0.90	0.99
Abnormal	0.20	0.55	0.12
Average	0.57	0.73	0.56
Weighted Average	0.87	0.87	0.90

Table 8. Results of applying XGBoost with the test split of the ERS dataset.

Class	F1-Score	Precision	Recall
Normal	0.15	0.23	0.11
Abnormal	0.85	0.80	0.90
Average	0.50	0.51	0.51
Weighted Average	0.71	0.68	0.74

The analysis of the results of the three methods - ResNet152, OneClassSVM, and XGBoost - reveals that XGBoost achieved the best performance among them. Therefore, it is evident that, for the task of abnormality detection in video capsule endoscopy images, the combination of XGBoost with deep feature extraction is the most robust and effective approach for addressing this specific problem.

4 Conclusion

In this study, we evaluated the performance of ResNet152 in the binary classification task and explored the use of OneClassSVM and XGBoost methods in an attempt to improve the final results. When employing the Kvasir-Capsule dataset, it became evident that the methods used were impacted by the class imbalance issue, where there was an unequal distribution of images between classes. In the case of the ERS dataset, this class imbalance problem was compounded by the limitation of a low overall quantity of images. As a result, the ResNet152 and XGBoost methods proved to be more effective in correctly classifying images belonging to the class they were predominantly trained on. It is expected that, with the refinement of the techniques used and, most importantly, with better-balanced datasets, more acceptable results can be achieved in the overall context of the classes.

It is also essential to highlight the importance of the patient-based data splitting strategy, which ensures the validity of the obtained results for analysis. One of the main challenges encountered in this work was the use of invalid data splits. This means that, in certain cases, the splits allowed for the sharing of images from the same patient across the training, validation, and test stages. Initially, the results obtained were very promising, but when the infeasibility of these invalid splits was realized, these results had to be discarded. This aspect underscores the critical importance of the patient-based data splitting approach to ensure the reliability of the results in any analysis or experiment.

Although the obtained results did not reach the desired level of performance, the analysis of the challenges encountered here has become essential for understanding them and future improvements.

For future research and the continuation of this work, it would be highly beneficial to investigate ways to improve class balance in the datasets used. Additionally, exploring the possibility of incorporating some or all of the images labeled as "uncertain" in the ERS dataset could be a promising strategy to significantly enhance the final results. These measures have the potential to considerably improve the performance of abnormality detection methods in video capsule endoscopy images.

References

1. Aabakken, L., et al.: Minimal standard terminology for gastrointestinal endoscopy-MST 3.0. Endoscopy **41**(08), 727–728 (2009)
2. Alaskar, H., Hussain, A., Al-Aseem, N., Liatsis, P., Al-Jumeily, D.: Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. Sensors **19**(6), 1265 (2019)
3. Bergstra, J., et al.: Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. In: Proceedings of the 12th Python in Science Conference, vol. 13, p. 20. Citeseer (2013)
4. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: optimal speed and accuracy of object detection. arXiv preprint: arXiv:2004.10934 (2020)
5. Chen, T., et al.: XGBoost: extreme gradient boosting. R Package Version 0.4-2 **1**(4), 1–4 (2015)
6. Cychnerski, J., Dziubich, T., Brzeski, A.: ERS: a novel comprehensive endoscopy image dataset for machine learning, compliant with the MST 3.0 specification (2022)
7. Du, W., et al.: Review on the applications of deep learning in the analysis of gastrointestinal endoscopy images. IEEE Access **7**, 142053–142069 (2019)
8. Hawkes, N.: Cancer survival data emphasise importance of early diagnosis. BMJ **364** (2019). https://doi.org/10.1136/bmj.l408
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR **abs/1512.03385** (2015). http://arxiv.org/abs/1512.03385
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

11. Hollstensson, M.: Detecting gastrointestinal abnormalities with binary classification of the Kvasir-Capsule dataset: a TensorFlow deep learning study (2022)
12. Iddan, G., Meron, G., Glukhovsky, A., Swain, P.: Wireless capsule endoscopy. Nature **405**(6785), 417 (2000). https://doi.org/10.1038/35013140
13. Jain, S., et al.: Detection of abnormality in wireless capsule endoscopy images using fractal features. Comput. Biol. Med. **127**, 104094 (2020) https://doi.org/10.1016/j.compbiomed.2020.104094, https://www.sciencedirect.com/science/article/pii/S001048252030425X
14. Jain, S., et al.: A deep CNN model for anomaly detection and localization in wireless capsule endoscopy images. Comput. Biol. Med. **137**, 104789 (2021) https://doi.org/10.1016/j.compbiomed.2021.104789, https://www.sciencedirect.com/science/article/pii/S0010482521005837
15. Lee, Y., Kang, P.: AnoViT: unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. IEEE Access **10**, 46717–46724 (2022)
16. Li, K., et al.: Colonoscopy polyp detection and classification: dataset creation and comparative evaluations. PLOS ONE **16**(8), 1–26 (2021). https://doi.org/10.1371/journal.pone.0255809
17. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J.: DetNet: a backbone network for object detection. arXiv preprint: arXiv:1804.06215 (2018)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
19. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
20. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision - ECCV 2016. Lecture Notes in Computer Science(), vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
21. Mohapatra, S., Kumar Pati, G., Mishra, M., Swarnkar, T.: Gastrointestinal abnormality detection and classification using empirical wavelet transform and deep convolutional neural network from endoscopic images. Ain Shams Eng. J. **14**(4), 101942 (2023). https://doi.org/10.1016/j.asej.2022.101942, https://www.sciencedirect.com/science/article/pii/S2090447922002532
22. Mukherjee, P., Roy, C.K., Roy, S.K.: OcFormer: one-class transformer network for image classification. arXiv preprint: arXiv:2204.11449 (2022)
23. Perera, P., Oza, P., Patel, V.M.: One-class classification: a survey (2021)
24. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint: arXiv:1804.02767 (2018)
25. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, vol. 28 (2015)
26. Shin, H.J., Eom, D.H., Kim, S.S.: One-class support vector machines-an application in machine fault detection and classification. Comput. Ind. Eng. **48**(2), 395–408 (2005) https://doi.org/10.1016/j.cie.2005.01.009, https://www.sciencedirect.com/science/article/pii/S0360835205000100
27. Smedsrud, P.H., et al.: Kvasir-capsule, a video capsule endoscopy dataset. Sci. Data **8**(1), 142 (2021)
28. Wang, S., Xing, Y., Zhang, L., Gao, H., Zhang, H.: Deep convolutional neural network for ulcer recognition in wireless capsule endoscopy: experimental feasibility and optimization. Comput. Math. Methods Med. **2019** (2019)

29. Zhang, S., Chi, C., Yao, Y., Lei, Z., Li, S.Z.: Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9759–9768 (2020)
30. Zhang, S., Wen, L., Bian, X., Lei, Z., Li, S.Z.: Single-shot refinement neural network for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4203–4212 (2018)