




Colonoscopic Polyp Detection with Deep Learning Assist

Alexandre Neto^{1,2}, Diogo Couto¹, Miguel Coimbra^{2,3} and António Cunha^{1,2}

¹*Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal*

²*Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, 3200-465 Porto, Portugal*

³*Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal*

Keywords: Deep Learning, Colorectal Cancer, Polyps, Computer Vision, Artificial Intelligence, Colonoscopy.

Abstract: Colorectal cancer is the third most common cancer and the second cause of cancer-related deaths in the world. Colonoscopic surveillance is extremely important to find cancer precursors such as adenomas or serrated polyps. Identifying small or flat polyps can be challenging during colonoscopy and highly dependent on the colonoscopist's skills. Deep learning algorithms can enable improvement of polyp detection rate and consequently assist to reduce physician subjectiveness and operation errors. This study aims to compare YOLO object detection architecture with self-attention models. In this study, the Kvasir-SEG polyp dataset, composed of 1000 colonoscopy annotated still images, were used to train (700 images) and validate (300 images) the performance of polyp detection algorithms. Well-defined architectures such as YOLOv4 and different YOLOv5 models were compared with more recent algorithms that rely on self-attention mechanisms, namely the DETR model, to understand which technique can be more helpful and reliable in clinical practice. In the end, the YOLOv5 proved to be the model achieving better results for polyp detection with 0.81 mAP, however, the DETR had 0.80 mAP proving to have the potential of reaching similar performances when compared to more well-established architectures.

1 INTRODUCTION


Colorectal Cancer (CRC) is the third most commonly diagnosed type of cancer (10.0% of total cancer cases) and the second deadliest type of cancer (9.4% of the total cancer deaths), estimating in more than 1.9 million colorectal cancer cases and 935,000 deaths worldwide following the report of GLOBOCAN 2020 (Sung et al., 2021). This type of cancer has higher rates in men than in women and has more incidence in Europe, North America and Eastern Asia (Sung et al., 2021).


Usually, CRC has precursors, namely polyps growing on the surface of the colon or rectum mucosal tissue. These polyps can change into cancer over many years, depending on their type and other associated risk factors. The main types of polyps are inflammatory, adenomatous and serrated (Figure 1).


Other risk factors associated with polyps can as well indicate CRC risks, such as their size and their

number (Huck & Bohl, 2016; Shaukat et al., 2020). Over time polyps can accumulate mutations and consequently develop high-grade dysplasia that can lead to the invasion into the submucosa and metastasis (Shaukat et al., 2020).

Thus, for these reasons, CRC screening along with polyp detection and removal are fundamental and allow for CRC prevention. Colonoscopy is the gold standard screening method which involves an endoscope that examines the entire length of the colon and detects and removes polyps. Using this screening tool allows us to detect polyps more often and remove them before developing mutations that can lead to CRC, leading to a higher survival rate (Montminy et al., 2020). However, due to lack of attention and tiredness, mistakes could be made by experts, leading to misdiagnosis. Indeed, polyp detection may be difficult to detect since some of them are hidden behind folds and only appear on the screen for a few moments; additionally, some lesions are flat and with subtle colour changes and may not

^a <https://orcid.org/0000-0002-4132-3186>

^b <https://orcid.org/0000-0001-7501-6523>

^c <https://orcid.org/0000-0002-3458-7693>

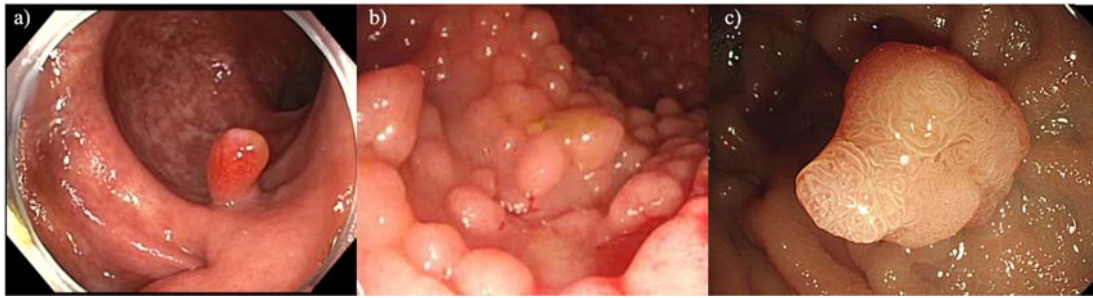


Figure 1: *a)* inflammatory polyps, *b)* adenomatous polyps and *c)* serrated polyps.

be recognized easily by human eyes. Besides, endoscopists with more experience have a higher detection rate when compared with inexperienced ones, leading sometimes to inconsistent diagnostics (Murakami et al., 2021).

To aid endoscopists and maintain the consistency between different exams and examiners, computer-aided systems appear to help minimize these issues. Currently, with the modern-day computational power, computer-aided systems rely mostly on machine learning and Deep Learning (DL) algorithms which can help during colonoscopy procedures. Computer-aided systems can be divided into Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx), where the first is designed to help in the detection of polyps during colonoscopy while the second aims to classify the polyp as adenomatous or hyperplastic/serrated or invasive cancer (Mori et al., 2017; van der Sommen et al., 2020).

Object detection DL algorithms are the gold standard for the development of CADe systems. These types of models deal with object instance detection, according to a certain class, given an input image, to find the precise location of the object and surround it within bounding boxes (Sharma & Mir, 2020; Srivastava et al., 2021). These algorithms can be mainly divided into two types: one-stage and two-stage detectors. One stage detector considers all positions on the image as possible candidates for object targeting and tries to categorize each of these Regions Of Interest (ROI) as an object or background. On the other hand, two-stage detectors propose ROI in the first stage, which is then used for the second stage, where features are extracted from these proposed ROIs for class prediction (Sharma & Mir, 2020).

Polyp detection is a problem already well solved in the machine learning community with clinical available solutions in the market. However, the solutions available use well-defined classic object detection architectures and more recent algorithms

are now available. In this study we will address the challenge of Computer-Aided Detection, by implementing recent object detection architectures that will be evaluated for the task of polyp detection and localization in colonoscopy images. Well-defined models will be compared, namely YOLOv4, and more recent versions of this architecture, YOLOv5, with new methods which rely on attention mechanisms for detection, to understand if similar performances can be reached.

This work is organised as follows: section 1 gives the clinical context about polyp detection and DL applied to this field; section 2 describes some studies using object detection algorithms applied to polyp detection, explains how object detection DL models work, followed by the contributions of this work, section 3 shows how the methodology of this work was organized, section 4 describes and discusses the results achieved; and for last, in section 5, are taken the respective conclusions and point out the future directions of this work.

2 LITERATURE REVIEW

Object detection models seek to classify each identified target in the image that is surrounded by bounding boxes. Thus, at the same time, beyond the localization, the identified object is classified accordingly to its class. As said before, mainly exists two types of object detection architectures, namely, one-stage and two-stage detectors.

Follows some examples of polyp detection studies using both types of detectors, a brief explanation of the object detection architectures used for this work and, in the end, the contributions of this study.

2.1 State-of-the-Art

In recent years, several studies have been published presenting new CAD methods which can improve the polyp detection rate of colonoscopies. This field of

study has already huge contributions with clinical solutions already available in the market to be used in the hospital environment. For this work, only studies which used one-stage detectors were selected, especially with the focus on studies which used YOLO architecture versions.

In the work of Pacal & Karaboga (2021) a CAD system was implemented to detect polyps and consequently helping to prevent CRC. For these different versions were implemented of a scaled YOLOv4 where the backbone or the entire architecture is replaced by a Cross Stage Partial Network (CSP). The first version is a YOLOv4-CSP, using as backbone a CSPDarknet53, replacing the first CSP layer with a residual DarkNet layer. In the Neck, was used the PAN with CSP. In the end, on the SPP module, was added a CSP. The second version removed the last CSP block on the backbone and place it with a transformer block with CSP. The remaining blocks were employed with CSPNet. To train these models was used the CVC-ClinicDB dataset and to test and evaluate them used the ETIS-LARIB and CVC-ColonDB. The first version of the model achieved a precision of 92%, recall of 83% and F1-Score of 87%. The second version with transformer blocks achieved a precision of 89%, recall of 81% and F1-Score of 85%.

The study of Chen et al. (2021) proposed an automatic polyp detection algorithm, using Single Shot Multibox Detecto (SSD) based on a VGG-16 model, changing the fully connected layer to a convolutional layer and four convolutional layers with decreased scales added successively. This model was then compared to the real annotations available from the datasets and to the results from a Mask R-CNN. A total of 4900 images, 2000 for training, 1500 for validation and 1400 for testing the model were collected. The SSD reach an mAP of 96%, higher than the manual detection and the Mask R-CNN.

Shen et al. (2021) proposed a convolutional transformer for polyp detection. The Convolutional Transformer network (COTR) is composed of a CNN responsible for feature extraction, transformer encoder layers with convolutional layers for feature encoding, a transformer decoder layer for object querying and a feed-forward network for object detection. To train this model was used the CVC-ClinicDB dataset and ETIS-LARIB and CVC-ColonDB for testing. COTR reached 92% precision, 83% sensitivity and 87% F1-Score to the ETIS-LARIB and 92% precision, 94% sensitivity and 93% F1-Score on the CVC-ColonDB.

The work done by Wan et al. (2021) combined a YOLOv5 model with self-attention mechanisms to

detect polyps. For the feature extraction module, an attention block is added for the enhancement of the feature channels. To train this model was used a Kvasir-SEG dataset which contains a total of 1000 images and 1000 images were collected from a local hospital to construct the WCY dataset. To increase the number of data available was used mosaic data augmentation. This model achieved 92% of precision, 90% of recall and 91% of F1-Score for the Kvasir dataset and 92% of precision, 92% of recall and 92% of F1-Score for the WCY dataset.

Quan et al. (2022) developed a CAD system, called EndoVigilant, based on single shot detection architecture for polyp detection. To train this system 83,000 colonoscopic images were used, which included polyps of various sizes, morphology and difficulty detection, annotated and reviewed by a specialized team. To validate the Endo Vigilant system, 21,454 colon images from an external dataset were used. The system achieved a sensitivity of 0.90, specificity of 0.97 and AUC of 0.94 per image.

In the selected studies are proposed modify DL models with addition of self-attention mechanisms in standard networks like YOLO architecture. These proposed algorithms are then compared to the results available in the state-of-the-art of polyp detection. In our study, the comparisons between standard algorithms like YOLO and specific design self-attention object detection architectures are made under the same study and circumstances

2.2 Contributions

Regarding DL object detection models, more precisely one-stage detectors, the proposal ROI is made simultaneously with the classification of the target object, which makes this type of architecture much quicker compared to two-stage detectors. You Only Look Once (YOLO) detectors target the object in a single regression problem, by simultaneously predicting multiple bounding boxes and the respective class probabilities for each of those boxes (Figure 2), turning this algorithm extremely fast by looking at the image globally and generalizing the representation of the object (Redmon et al., 2016).

The anchor boxes have been introduced in more recent versions of YOLO to help predict multiple objects in the same grid cell and objects with different alignments. YOLOv2, YOLOv3, YOLOv4 and YOLOv5 use anchor boxes with the ability to predict boxes at 3 different scales for detecting objects of different sizes. However, YOLO architectures have some disadvantages such as comparatively low recall and more localization error when compared to Faster

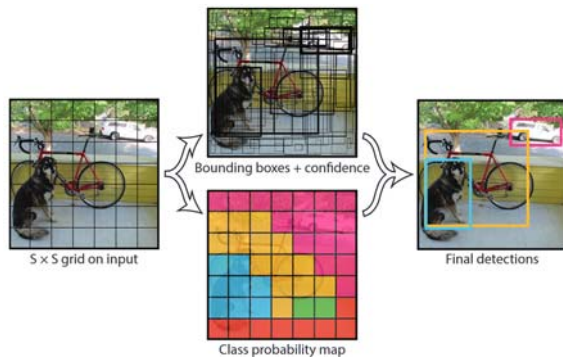


Figure 2: YOLO system detection, dividing the image into $S \times S$ grid and for each cell predicts bounding boxes, confidence for the boxes and class probabilities, from (Redmon et al., 2016).

R-CNN, struggle to detect close objects because each grid can propose only two bounding boxes and struggles to detect small objects (Bochkovskiy et al., 2020; Redmon & Farhadi, 2017, 2018; *Ultralytics/Yolov5*, 2020/2022).

The use of self-attention mechanisms provides the advantage in terms of speed regarding object detection problems, due to the fact of its parallel processing capability and for not using restrictive techniques such as anchor boxes and non-maximum suppression. The End-to-End Object Detection with Transformers (DETR) uses an encoder-decoder architecture, based on transformers, to detect and localize objects in images (Figure 3). Transformers are a self-attention mechanism that model the interactions between pairwise elements in a sequence (Carion et al., 2020).

DETR uses a Convolutional Neural Network (CNN) as a backbone to collect features from the input images that are then flattened and supplemented with a positional encoding before passing it into a transformer encoder. Additionally, DETR is more powerful in cases the object is important in the image, ie the objects to be detected are generally related to each other and the surrounding environment.

For these reasons, this study aims to compare different versions of YOLO architectures with DETR, to understand if the differences in these architectures can enhance or jeopardize the performance of polyp detection.

For this, the following research question was formulated: *Can self-attention mechanisms such as transformers applied to object detection architectures enhance polyp detection?*

Thus, the proposed study will contribute to finding if DETR can have similar performance when compared to YOLO architectures for polyp detection.

3 METHODOLOGY

This work compares recent object detection architectures for polyp detection, namely, YOLOv4 (Bochkovskiy et al., 2020), YOLOv5 [19] (different versions of this model) and the DETR (Carion et al., 2020), as described in section 2.2. The pipeline representing the workflow of this study is presented in figure 4.

First, colonoscopic images were collected from a public database, pre-processed and annotated into datasets. These datasets were then used to train the mentioned models. For last, these trained models were evaluated for polyp detection.

3.1 Data Preparation

To train all the different architectures were collected colonoscopic images with the presence of polyps from the Kvasir-SEG dataset with a total of 1000 images with 640x640 resolution (Borgli et al., 2020). The 1000 images were labelled and manually segmented the polyp outlines by a multidiscipline team composed of engineers and medical doctors from Vestre Viken Health Trust in Norway. In the end, the annotations were reviewed by an experienced gastroenterologist (Borgli et al., 2020; Jha et al., 2019).

From the 1000 polyp images, 700 were used for training and 300 to evaluate the models.

All these images had the respective annotation in txt files, with the respective class, bounding box coordinates (x and y centre), width and height. The box coordinates must be normalized by the dimensions of the image (values must be between 0 and 1).

3.2 Training

YOLOv5 was trained with a batch size of 16 during 320 epochs, starting with a learning rate at 0.01, using as an optimizer the Stochastic Gradient Descent (SGD) with a momentum of 0.937 and a weight decay of $5e-4$. As loss function was used a Binary Cross-Entropy. These hyperparameters were the ones selected for all versions of YOLOv5, except for YOLOv5x which had 8 as batch size. YOLOv4 follow the same hyperparameters, only changing the batch size for 4 due to memory consumption. The Intersection Over the Union (IoU) between the ground truth and predicted bounding boxes threshold was equal to or superior to 20%. Random transformations in colour, saturation, and brightness

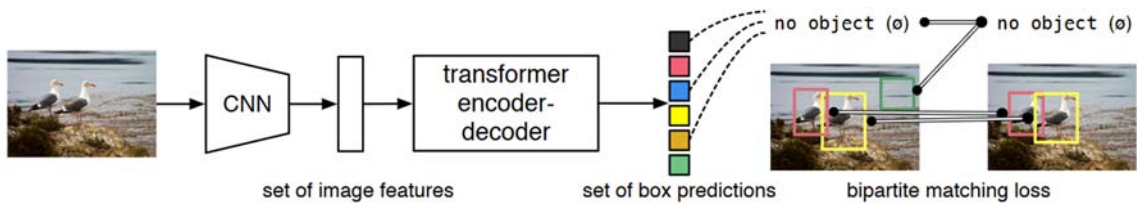


Figure 3: DETR overall architecture from (Carion et al., 2020).

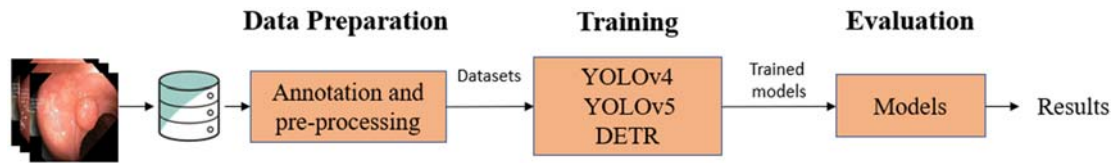


Figure 4: Pipeline of the proposed study.

were applied to the polyp images to increase the dataset. The images also suffered random flipping and were randomly resized. For last, Mosaic and mix-up data augmentation were used.

The DETR model was trained with a batch size of 16 for 320 epochs, starting with a learning rate of 1e-3 for the first 20 epochs and after that passing to 1e-4. The optimizer used was the Adam and the loss function is an optimal bipartite matching function that calculates the best match of predictions given the ground truths, and after calculating the matched pairs for the set, computes the Hungarian loss function. For the backbone, DETR uses a ResNet50.

3.3 Evaluation

The evaluation of polyp detection was made using the detection evaluation metrics used by Common Objects in Context. Average Precision (AP) is the average over multiple IoU values. The AP is the average over all categories, traditionally called mAP.

- AP.50 – AP at IoU=0.50
- AP.75 – AP at IoU=0.75

The AP (1) is to find the area under the precision-recall curve. The AP curve has on the x-axis the recall and on the y-axis the precision. The AP computes the average values of $p(r)$ over the interval from $r=0$ to $r=1$. The mAP (2) is the mean of the AP, that is, the AP for each class (Q) is calculated, and, in the end, it is averaged.

$$AP = \int_0^1 p(r) dr \quad (1)$$

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (2)$$

$$Pre = \frac{TP}{TP+FP} \quad (3)$$

Precision (pre) (2) is the ratio of correct predictions to the number of positive results predicted by the classifier.

Specificity (spe) (4) measures the proportion of the negative cases that were correctly classified. Recall or sensitivity (sen) (5) is the number of correctly predicted results divided by the number of all those that should have been classified as positive.

$$Spe = \frac{TN}{TN+FP} \quad (4)$$

$$Sen = \frac{TP}{TP+FN} \quad (5)$$

F1-Score (F1) (6) outputs a value between zero and one and tries to find the balance between precision and recall, letting know how accurate the model is and how many samples it correctly classifies. The F1 is the harmonic mean of these two.

$$F1 = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (6)$$

4 RESULTS AND DISCUSSION

4.1 Results

The results of the different object detection architectures are presented in Table 1.

Overall, all the different architectures reached similar results. All the models achieved an mAP above 0.70, specificity around 0.90, and sensitivity, precision and f1-score around 0.65. The YOLOv5x reached slightly better results and the DETR model,

Table 1: Results for polyp detection for each architecture.

Model	mAP	AP ⁵⁰	AP ⁷⁵	Spe	Sen	Pre	F1
YOLOv4	0.71	0.71	0.57	0.90	0.66	0.65	0.63
YOLOv5l	0.81	0.81	0.68	0.91	0.65	0.65	0.64
YOLOv5m	0.81	0.80	0.68	0.90	0.65	0.65	0.64
YOLOv5n	0.75	0.75	0.60	0.89	0.64	0.64	0.62
YOLOv5s	0.74	0.74	0.63	0.89	0.62	0.63	0.61
YOLOv5x	0.80	0.80	0.69	0.91	0.66	0.67	0.65
DETR	0.80	0.80	0.48	0.90	0.69	0.65	0.66

despite not achieving the best performance, had similar results when compared to the remaining models. In most cases, the models correctly detect the polyp, even though the bounding box does not match the expert annotation, however, this does not mean that the models' are making wrong decisions. Some examples are illustrated in figure 5.

4.2 Discussion

The DETR model achieved similar results to YOLO architectures with an mAP of 0.80. All the models have similar results, with slight differences among the different used metrics. In terms of specificity and precision, the YOLOv5x reached better results and regarding the sensitivity and f1-score DETR had the highest values. Either way, sensitivity and specificity do not fully represent the performance of the model, since the object can be detected despite the detection do not fully match with the ground truth annotation. This can jeopardize the result values but the detection is well-made anyway. Some examples of this will be illustrated in figure 5.

The DETR predicts the polyp localization directly in the input image, without the need for anchor boxes, and thus has more knowledge of polyp localization using pair-wise pixel relations between them, while being able to use the whole image as context. These characteristics of the DETR architecture allow a more complete understanding of the image domain, with more highlights for features which can be associated with polyp presence, such as texture and size regarding the intestinal background context.

Besides analysing the global results of our experiment, some additional insights can be obtained by visually inspecting some individual examples, such as the ones depicted in figure 5. The top row images have a large size polyp present, with the ground truth bounding box in green and the prediction in yellow. All the models can detect the presence of the lesion, with the YOLOv5l as the closer prediction when compared to the manual annotation. Despite the DETR prediction not matching the ground truth annotation, the predicted bounding box surrounds the

local with more polyp characteristics, with a smaller bounding box but with a more precise location of the polyp.

The bottom row images exhibit more disagreement between the models, namely the size of the bounding box which surrounds the polyp and consequently the localization. All the architectures can detect the presence of the polyp but some of them have difficulties in correctly identifying the region of interest, namely the YOLOv4, YOLOv5n and YOLOv5s. The remaining models do a more accurate detection, even if they do not match exactly the ground truth bounding box. These models correctly detect the polyp with bounding boxes more adjusted to the size of the manual annotation.

Figure 6 shows us an example of a misclassification made by all the models, where a bubble was detected instead of the polyp. In this case, all the architectures confused the bubble due to the circular shape, instead of detecting the polyp region, which this time did not have the typical form, even when compared to the examples in figure 5.

In this particular case, learning features such as textures can lead DL models to more accurate decisions. Enhanced texture features can be made to upgrade object recognition algorithms by using pre-processing methods such as the wavelet transform approach, local binary pattern or grey-level co-occurrence matrices. Textures can be a common characteristic in specific lesions, helping DL models to learn the association of certain textures with a determined lesion.

For these reasons, we believe that architectures with self-attention mechanisms show advantages which can be helpful in this specific scenario. These new types of models can perform as well as well-established algorithms such as the YOLOs architectures. As such, our answer to our proposed research question is that self-attention mechanisms such as transformers applied to object detection architectures can achieve similar performances as well-established algorithms in polyp detection in a colonoscopic imaging scenario and deserve more depth research to fully explore this potential.

5 CONCLUSIONS

Early polyp detection has a key role in the prevention and development of CRC. Object detection DL architecture can assist during colonoscopy well specialized and experienced physicians to maintain consistency for each exam.

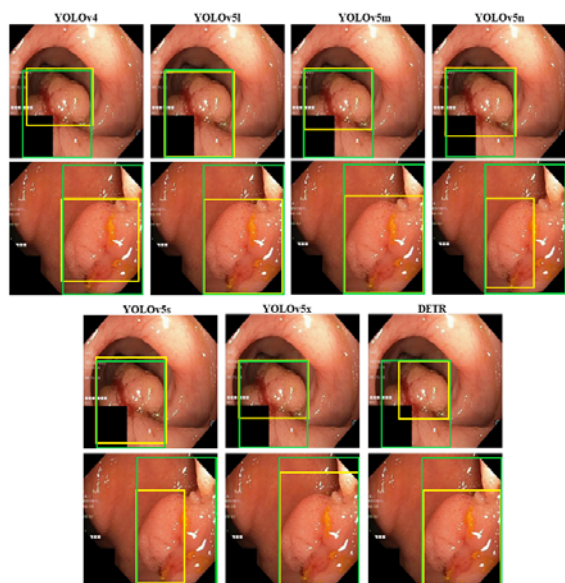


Figure 5: Predictions for each architecture used. **Green line** – ground truth bounding boxes; **Yellow line** – predicted bounding boxes.



Figure 6: Example of a false positive prediction in all architectures.

In this work, the aim was to compare well-established algorithms for this purpose with more recent methods which rely upon attention mechanisms. This was verified since the DETR algorithm had similar or more precise predictions compared to the YOLOs architectures. Several commercial products already exist for polyp detection, achieving satisfying results in clinical practice. Our study showed that object detection algorithms, based on self-attention mechanisms, can have similar performance when compared to well-established architectures such as YOLO, while having additional potential advantages such as less

probability of inductive bias, increasing in speed detection and more contextualization with the surrounding environment of the object, motivating further research in this field with the potential of surpassing current state-of-the-art solutions. In future work, we intend to combine YOLO and SSD architectures with attention blocks from transformers, to understand if this mechanism can further enhance our ability to detect colonic polyps and explore specific texture features associated with each type of polyp to increase the domain knowledge of DL models.

ACKNOWLEDGEMENTS

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project PTDC/EEI-EEE/5557/2020.

REFERENCES

- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection (arXiv:2004.10934). arXiv. <http://arxiv.org/abs/2004.10934>
- Borgli, H., Thambawita, V., Smedsrud, P. H., Hicks, S., Jha, D., Eskeland, S. L., Randel, K. R., Pogorelov, K., Lux, M., Nguyen, D. T. D., Johansen, D., Griwodz, C., Stensland, H. K., Garcia-Ceja, E., Schmidt, P. T., Hammer, H. L., Riegler, M. A., Halvorsen, P., & de Lange, T. (2020). HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data*, 7(1), 283. <https://doi.org/10.1038/s41597-020-00622-y>
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers (arXiv:2005.12872). arXiv. <http://arxiv.org/abs/2005.12872>
- Chen, X., Zhang, K., Lin, S., Dai, K. F., & Yun, Y. (2021). Single Shot Multibox Detector Automatic Polyp Detection Network Based on Gastrointestinal Endoscopic Images. *Computational and Mathematical Methods in Medicine*, 2021, 1–6. <https://doi.org/10.1155/2021/2144472>
- Huck, M., & Bohl, J. (2016). Colonic Polyps: Diagnosis and Surveillance. *Clinics in Colon and Rectal Surgery*, 29(04), 296–305. <https://doi.org/10.1055/s-0036-1584091>
- Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., & Johansen, H. D. (2019). Kvasir-SEG: A Segmented Polyp Dataset (arXiv:1911.07069). arXiv. <http://arxiv.org/abs/1911.07069>

- Montminy, E. M., Jang, A., Conner, M., & Karlitz, J. J. (2020). Screening for Colorectal Cancer. *Medical Clinics of North America*, 104(6), 1023–1036. <https://doi.org/10.1016/j.mcna.2020.08.004>
- Mori, Y., Kudo, S., Berzin, T., Misawa, M., & Takeda, K. (2017). Computer-aided diagnosis for colonoscopy. *Endoscopy*, 49(08), 813–819. <https://doi.org/10.1055/s-0043-109430>
- Murakami, D., Yamato, M., Amano, Y., & Tada, T. (2021). Challenging detection of hard-to-find gastric cancers with artificial intelligence-assisted endoscopy. *Gut*, 70(6), 1196–1198. <https://doi.org/10.1136/gutjnl-2020-322453>
- Pacal, I., & Karaboga, D. (2021). A robust real-time deep learning based automatic polyp detection system. *Computers in Biology and Medicine*, 134, 104519. <https://doi.org/10.1016/j.combiomed.2021.104519>
- Quan, S. Y., Wei, M. T., Lee, J., Mohi-Ud-Din, R., Mostaghim, R., Sachdev, R., Siegel, D., Friedlander, Y., & Friedland, S. (2022). Clinical evaluation of a real-time artificial intelligence-based polyp detection system: A US multi-center pilot study. *Scientific Reports*, 12(1), 6598. <https://doi.org/10.1038/s41598-022-10597-y>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection (arXiv:1506.02640). arXiv. <http://arxiv.org/abs/1506.02640>
- Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6517–6525. <https://doi.org/10.1109/CVPR.2017.690>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement (arXiv:1804.02767). arXiv. <http://arxiv.org/abs/1804.02767>
- Sharma, V., & Mir, R. N. (2020). A comprehensive and systematic look up into deep learning based object detection techniques: A review. *Computer Science Review*, 38, 100301. <https://doi.org/10.1016/j.cosrev.2020.100301>
- Shaukat, A., Kaltenbach, T., Dominitz, J. A., Robertson, D. J., Anderson, J. C., Cruise, M., Burke, C. A., Gupta, S., Lieberman, D., Syngal, S., & Rex, D. K. (2020). Endoscopic Recognition and Management Strategies for Malignant Colorectal Polyps: Recommendations of the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*, 159(5), 1916–1934.e2. <https://doi.org/10.1053/j.gastro.2020.08.050>
- Shen, Z., Lin, C., & Zheng, S. (2021). COTR: Convolution in Transformer Network for End to End Polyp Detection (arXiv:2105.10925). arXiv. <http://arxiv.org/abs/2105.10925>
- Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. (2021). Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, 8(1), 66. <https://doi.org/10.1186/s40537-021-00434-w>
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- Ultralytics/yolov5. (2022). [Python]. Ultralytics. <https://github.com/ultralytics/yolov5> (Original work published 2020)
- van der Sommen, F., de Groof, J., Struyvenberg, M., van der Putten, J., Boers, T., Fockens, K., Schoon, E. J., Curvers, W., de With, P., Mori, Y., Byrne, M., & Bergman, J. J. G. H. M. (2020). Machine learning in GI endoscopy: Practical guidance in how to interpret a novel field. *Gut*, 69(11), 2035–2045. <https://doi.org/10.1136/gutjnl-2019-320466>
- Wan, J., Chen, B., & Yu, Y. (2021). Polyp Detection from Colorectum Images by Using Attentive YOLOv5. *Diagnostics*, 11(12), 2264. <https://doi.org/10.3390/diagnostics11122264>