

# Segmentação de Pólipos em Imagens de Colonoscopia utilizando YOLOv8

Sandro Luis de Araujo Junior<sup>1</sup>, Michel Hanzen Scheeren<sup>1</sup>,  
Rubens Miguel Gomes Aguiar<sup>2</sup>, Eduardo Mendes<sup>1</sup>,  
Ricardo Augusto Pereira Franco<sup>3</sup>, Pedro Luiz de Paula Filho<sup>1</sup>

<sup>1</sup>Universidade Tecnológica Federal Do Paraná (UTFPR)  
Medianeira – PR – Brasil.

<sup>2</sup>Centro de Excelência em Inteligência Artificial  
Goiânia – GO – Brasil.

<sup>3</sup>Universidade Federal de Goiânia  
Goiânia – GO – Brasil.

{sandro.1999, emendes}@alunos.utfpr.edu.br

{michelhanszschereen, rubensgaguiar0}@gmail.com

pedrol@utfpr.edu.br, ricardofranco@ufg.br

**Abstract.** *Polyp segmentation in colonoscopy images is an important computer-aided diagnostic task, as it can assist doctors in identifying and consequently removing polyps, thus contributing to the reduction of cases of colorectal cancer, one of the most common and lethal types of cancer. In this work, the capacity of different variants of the YOLOv8 algorithm was evaluated in the task of polyp segmentation, using three public databases of colonoscopy images. Among the different versions, YOLOv8n proved to be the most effective alternative, despite being the simplest version. The results achieved reached 0.919 dice and 0.877 IoU, thus demonstrating the effectiveness of the model.*

**Resumo.** *A segmentação de pólipos em imagens de colonoscopia é uma importante tarefa de diagnóstico auxiliado por computador, uma vez que pode auxiliar médicos na identificação e consequentemente remoção dos pólipos, contribuindo assim para a redução dos casos de câncer colorretal, um dos tipos de câncer mais comum e letal. Neste trabalho, a capacidade das diferentes variantes do algoritmo YOLOv8 foi avaliada na tarefa de segmentação de pólipos, utilizando para isto três bases públicas de imagens de colonoscopia. Dentre as diferentes versões, o YOLOv8n se mostrou a alternativa mais eficaz, apesar de ser a versão mais simples. Os resultados alcançados chegaram à 0,919 de dice e 0,877 de IoU, evidenciando assim a eficácia do modelo.*

## 1. Introdução

O câncer é uma das principais causas de morte no Brasil e no mundo. Estima-se que em 2020 ocorreram 19,3 milhões de novos casos, e quase 10 milhões de mortes devido à esta doença. Dentre os diversos tipos de câncer, pode-se destacar o câncer colorretal,

que é o terceiro mais diagnosticado e letal, representando cerca de 10% do total de mortes [Sung et al. 2021].

A maior parte dos casos de câncer colorretal se iniciam a partir dos pólipos. Os pólipos são crescimentos celulares não cancerígenos que ocorrem na parede interna do intestino grosso, mas que podem evoluir para um câncer. Dessa forma, uma das principais formas de prevenção do câncer colorretal é a identificação e remoção dos pólipos através de exames de colonoscopia [Keum and Giovannucci 2019].

Neste exame, um tubo flexível é inserido no ânus do paciente até chegar ao reto e ao cólon. Este instrumento, conhecido como colonoscópio, possui uma pequena câmera na ponta, que permite a coleta de imagens nas quais pode ser detectada a presença dos pólipos. Outros instrumentos podem ser passados por dentro do colonoscópio para realizar a remoção dos pólipos [Lee et al. 2014].

As imagens obtidas durante este exame são avaliadas por médicos especializados. No entanto, devido à natureza subjetiva da análise, parte dos pólipos pode não ser identificada, permitindo assim o seu desenvolvimento. Estudos indicam que até 27% dos pólipos com menos de 5mm podem passar despercebidos, mesmo quando avaliados por colonoscopistas experientes [Heresbach et al. 2008, Rex et al. 1997]. Dessa forma, surge a oportunidade da utilização de algoritmos de visão computacional para auxiliar no diagnóstico dessas lesões.

Uma das tarefas de visão computacional mais utilizadas em aplicações biomédicas para auxiliar na identificação de pólipos e tumores é a segmentação. A segmentação é uma operação que divide uma imagem em regiões, possibilitando assim a separação de objetos e texturas em uma imagem [Pacal et al. 2020]. Por exemplo, em uma imagem contendo um ou mais pólipos, um algoritmo de segmentação faria a separação dos pixels correspondentes a estes elementos na imagem.

Dessa forma, o presente trabalho tem como objetivo aprimorar o processo de diagnóstico de pólipos ao realizar a aplicação de um algoritmo de segmentação em imagens de colonoscopia. Ao fazer isso, almeja-se alcançar um diagnóstico mais eficaz e ágil durante esses procedimentos, permitindo assim a identificação precoce dos pólipos, o que, por sua vez, pode ter um impacto substancial na saúde dos pacientes.

Este trabalho está estruturado da seguinte forma: a Seção 2 discorre sobre os trabalhos relacionados relevantes para este estudo; a Seção 3 detalha os materiais e métodos empregados; a Seção 4 apresenta e analisa os resultados obtidos; finalmente, a Seção 5 aborda as conclusões do estudo.

## **2. Trabalhos relacionados**

Algoritmos de diagnóstico auxiliado por computador estão presentes na literatura desde a metade da década de 1980. Entretanto, essa área de pesquisa recebeu grande destaque nos últimos anos devido ao desenvolvimento de soluções baseadas em *Deep Learning* (DL), por apresentarem resultados melhores do que os métodos clássicos de processamento de imagem [Pacal et al. 2020].

Os algoritmos de DL representam uma ferramenta poderosa na área de processamento de imagens médicas, como a detecção e diagnóstico de cânceres bem como na

segmentação de tumores, devido à sua capacidade de extrair características de alto nível diretamente das imagens [Pacal et al. 2020].

Dessa forma, diversos conjuntos públicos de imagens foram criados para o treinamento e avaliação de modelos de DL destinados à segmentação de imagens médicas. Alguns dos principais conjuntos de imagens de colonoscopia contendo pólipos rotulados para segmentação são:

- **Kvasir-SEG**<sup>1</sup>: constituído por 1000 imagens que foram coletadas e avaliadas por gastroenterologistas do Hospital Vestre Viken Health Trust, localizado na Noruega. A resolução das imagens varia entre  $332 \times 487$  e  $1920 \times 1072$  [Jha et al. 2020].
- **CVC-ClinicDB**<sup>2</sup>: desenvolvido a partir de vídeos de colonoscopia do Hospital Clínico de Barcelona, localizado na Espanha. No total o conjunto possui 612 imagens de pólipos em resolução  $384 \times 288$  [Bernal et al. 2015].
- **ETIS-Larib Polyp DB**<sup>3</sup>: foi criado a partir de quadros extraídos de vídeos de colonoscopia, totalizando 196 imagens de resolução  $1225 \times 966$ . Este conjunto de dados é de propriedade conjunta do *Lariboisière Hospital-APHP* e do *ETIS laboratory, ENSEA, University of Cergy-Pontoise*, ambos localizados na França [Silva et al. 2014].

A utilização de conjuntos públicos de imagens permite a comparação de desempenho entre diferentes modelos, facilita a reprodutibilidade de pesquisas, e contribui para o avanço e aprimoramento contínuo dessa área [Mei et al. 2023].

Dois métricas essenciais para avaliar modelos de segmentação são o coeficiente de dice e o Intersection over Union (IoU), que medem, respectivamente, a similaridade e a sobreposição entre a máscara de segmentação anotada por um especialista e a máscara predita pelo modelo. Os valores dessas métricas variam de 0 a 1, de forma que valores mais altos indicam melhor desempenho. Detalhes adicionais sobre essas métricas são discutidos na Seção 3.

Dentre os algoritmos de segmentação treinados e avaliados nos conjuntos apresentados, destaca-se o **Unet++**. A arquitetura deste modelo consiste em um *encoder-decoder* no qual as subredes são conectadas através de uma série de caminhos densos aninhados, permitindo assim uma aprendizagem facilitada devido à similaridade semântica das características extraídas presentes no *encoder* e no *decoder* [Zhou et al. 2018].

Os resultados obtidos por Mei et al. (2023) utilizando este algoritmo sobre métricas *dice* e *Intersection over Union* (IoU) foram 0,794 e 0,729 no conjunto CVC-Clinic DB, 0,821 e 0,743 no conjunto Kvasir-SEG, e 0,401 e 0,344 no conjunto ETIS-Larib respectivamente.

Outro algoritmo de DL capaz de gerar a segmentação precisa de pólipos é o **Pra-Net**. Este algoritmo primeiro agrega as características de alto nível da imagem para formar uma área aproximada inicial dos pólipos. Em seguida, as bordas desses elementos são modeladas implicitamente através de um mecanismo de atenção reversa. Através dessa

---

<sup>1</sup><https://datasets.simula.no/kvasir-seg/>

<sup>2</sup><https://polyp.grand-challenge.org/CVCClinicDB/>

<sup>3</sup><https://polyp.grand-challenge.org/ETISLarib/>

abordagem, foram atingidos os valores de 0,899 de *dice*, e 0,849 de IoU sobre o conjunto de imagens CVC Clinic DB, 0,898 e 0,840 sobre Kvasir-SEG, e 0,628 e 0,567 sobre o ETIS-Larib [Fan et al. 2020].

### 3. Materiais e métodos

Para alcançar o objetivo de auxiliar no diagnóstico de pólipos em imagens de colonoscopia através da aplicação de um algoritmo de segmentação, o método adotado neste trabalho foi dividido em quatro etapas, conforme ilustrado na Figura 1, e detalhado na sequência.

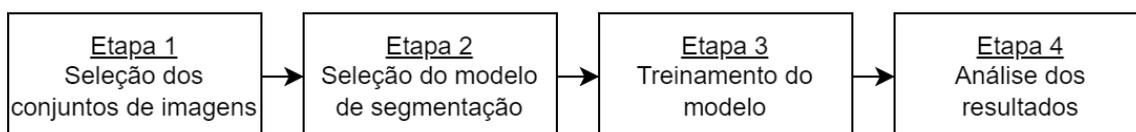


Figura 1. Fluxograma do método utilizado

Para conduzir o treinamento do modelo de aprendizado profundo na tarefa de segmentação, é essencial utilizar conjuntos de imagens que incluam exemplos das classes a serem segmentadas. Dessa forma, durante a primeira etapa, foram selecionadas três bases públicas de imagens de colonoscopia que contêm máscaras de segmentação anotadas para os pólipos, sendo elas: Kvasir-SEG, CVC-ClinicDB e ETIS-Larib Polyp DB.

A segunda etapa consistiu na seleção de um modelo de segmentação. Nesse sentido optou-se pelo *You Only Look Once* (YOLO) em sua oitava versão, que além da segmentação, também é capaz de realizar outras tarefas de visão computacional como detecção de objetos, classificação e estimativa de pose [Jocher et al. 2023].

O desempenho deste algoritmo já foi avaliado no processamento de imagens de diversas áreas médicas, obtendo bons resultados em imagens de ultrassom [Inui et al. 2023], tomografia computadorizada e raio-X [Pandey et al. 2023] e angiografia [Osama et al. 2023]. No contexto de pólipos, este algoritmo obteve bons resultados na tarefa de detecção de objetos, alcançando 0,956 de *precision*, 0,917 de *recall*, e 0,924 de *dice* [Lalinia and Sahafi 2023].

A arquitetura do YOLOv8 é composta por dois componentes principais: o *backbone* e o *head*, conforme representado de forma simplificada na Figura 2.

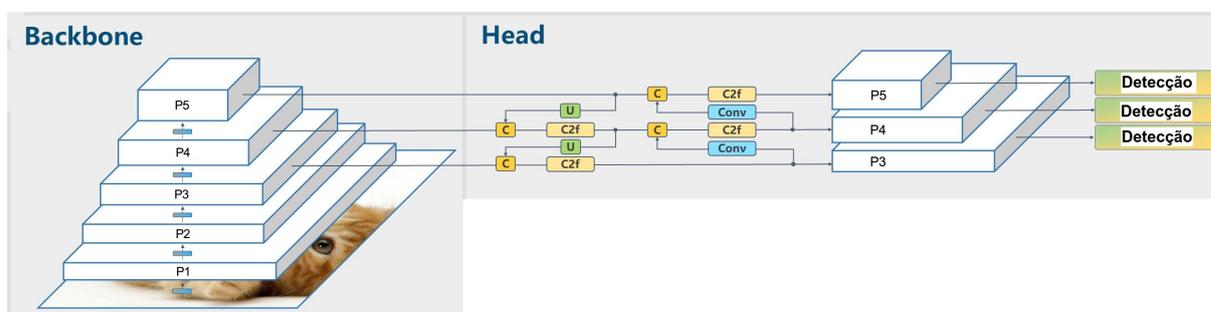


Figura 2. Arquitetura simplificada do YOLOv8 [Jocher et al. 2023]

A função do *backbone* consiste na extração de características hierárquicas em diferentes escalas. Inicialmente, nas primeiras camadas, são extraídas características de

baixo nível, como bordas e texturas. À medida que se avança nas camadas mais profundas, é realizada a extração de características de mais alto nível, como partes dos objetos e informações semânticas [Terven and Cordova-Esparza 2023].

O *head*, por sua vez, assume a responsabilidade de receber as *feature maps* geradas pelo *backbone* e processá-las a fim de produzir a saída final do algoritmo. Uma característica distintiva do *head* é o seu desacoplamento, que permite que ele conduza as tarefas de pontuação de objeto, classificação e regressão de maneira independente. Isso significa que cada ramificação concentra-se exclusivamente em sua tarefa, o que, por sua vez, aprimora a precisão geral do modelo [Jocher et al. 2023].

O YOLOv8 apresenta cinco variantes com diferentes tamanhos, de forma que quanto maior o tamanho, melhor tende a ser o seu desempenho, entretanto, o treinamento e a realização das inferências possuem um custo computacional maior. As variantes disponíveis são as seguintes: YOLOv8n (nano), YOLOv8s (*small*, ou pequeno), YOLOv8m (*medium*, ou médio), YOLOv8l (*large*, ou grande) and YOLOv8x (*extra large*, ou extra grande).

A escolha deste modelo foi motivada pelo fato de que, apesar de seu desempenho ter sido previamente avaliado na tarefa de detecção de pólipos [Lalinia and Sahafi 2023], que visa demarcar esses crescimentos celulares em uma imagem utilizando retângulos delimitadores, existe uma escassez de trabalhos na literatura que avaliem este algoritmo na tarefa de segmentação, a qual tem por objetivo realizar a classificação à nível de pixel das áreas da imagem correspondentes aos pólipos.

Na terceira etapa, o modelo de segmentação escolhido foi treinado utilizando a separação proposta por Fan et al. 2020, na qual são utilizados os conjuntos Kvasir-SEG e CVC-ClinicDB, dividindo-os em três partes: 80% para treinamento, 10% para validação e 10% para testes. Dessa forma, o *dataset* ETIS-Larib Polyp DB foi utilizado apenas para testes. A separação dos dados é detalhada na Tabela 1.

**Tabela 1. Conjunto de imagens utilizado**

Conjunto	Treinamento	Validação	Testes	Total
Kvasir	800	100	100	1000
CVC	490	61	61	612
ETIS	0	0	196	196
Total	1290	161	357	1808

Os hiperparâmetros utilizados para o treinamento das diferentes variantes do YOLOv8 foram mantidos com seus valores padrões, com exceção do número de épocas que foi configurado para 300. Vale também ressaltar o valor 50 de *patience*, que indica que o treinamento seria interrompido caso não fossem observadas melhorias durante 50 épocas. Todos os treinamentos foram realizados utilizando o ambiente Google Colaboratory<sup>4</sup>, utilizando uma *Graphics Processing Unit* (GPU) Nvidia Tesla T4.

Por último, na quarta etapa foi realizada a avaliação da solução proposta por meio de quatro métricas de segmentação amplamente utilizadas na literatura, sendo elas: coeficiente de similaridade *dice*, *Intersection over Union* (IoU), *precision*, e *recall*.

<sup>4</sup><https://colab.google/>

O coeficiente de similaridade *dice* (Equação 1) é a métrica mais utilizada para a análise de imagens médicas no contexto de segmentação. Ele mede a similaridade entre duas áreas, sendo estas a máscara de segmentação anotada e a predita pelo algoritmo. Seu valor varia de 0 a 1, de forma que 0 indica máscaras completamente diferentes, e 1 representa máscaras idênticas [Liu et al. 2021].

$$Dice(A, B) = 2 \frac{|A \cap B|}{|A| + |B|} \quad (1)$$

O IoU, também conhecido como Índice de Jaccard, é utilizado para medir a sobreposição entre duas áreas A e B. Ele é calculado dividindo-se a intersecção pela união entre as duas áreas (conforme a Equação 2), e seu valor varia de 0 a 1. No caso da segmentação de imagens, o IoU é calculado entre as máscaras preditas e anotadas [Minaee et al. 2022].

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

A métrica *precision*, definida na Equação 3, representa a proporção de pixels preditos que correspondem aos pixels anotados. Já o *recall*, apresentado na Equação 4 mede a proporção de pixels anotados que foram corretamente preditos pelo algoritmo [Haque and Neubert 2020]. Ambas as equações utilizam a seguinte nomenclatura:

- TP (verdadeiro positivo): pixels anotados como pertencentes a uma classe A e corretamente preditos como A;
- FP (falso positivo): pixels não anotados como A, mas preditos como A;
- FN (falso negativo): pixels anotados como A, mas preditos como não A.

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

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

#### 4. Resultados e discussões

Após o treinamento das diferentes variantes do YOLOv8 utilizando os conjuntos de imagens Kvasir-SEG e CVC Clinic DB, o desempenho dos modelos treinados foi avaliado sobre as separações de testes dos mesmos, e sobre o conjunto ETIS Larib Polyp DB. Os resultados obtidos são apresentados na Tabela 2, na qual o melhor resultado atingido por cada métrica sobre determinado conjunto é apresentada em negrito, e o segundo melhor sublinhado.

Conforme evidenciado na Tabela 2(a), que expõe os resultados referentes à métrica *dice* (considerada a mais crucial), a variante *n* apresentou o melhor resultado para o *dataset* CVC-Clinic DB. Além disso, obteve o segundo melhor resultado nos demais conjuntos. No caso dos *datasets* Kvasir-SEG e ETIS-Larib, os melhores resultados foram alcançados pelos modelos *m* e *s* respectivamente.

YOLO	Kvasir	CVC	ETIS	YOLO	Kvasir	CVC	ETIS
n	0,859	<b>0,919</b>	0,647	n	0,795	<b>0,877</b>	0,585
s	0,845	0,910	<b>0,652</b>	s	0,791	0,867	<b>0,592</b>
m	<b>0,869</b>	0,906	0,584	m	<b>0,808</b>	0,859	0,530
l	0,839	0,911	0,567	l	0,777	0,872	0,519
x	0,847	0,822	0,559	x	0,783	0,844	0,507

(a) Dice				(b) IoU			
YOLO	Kvasir	CVC	ETIS	YOLO	Kvasir	CVC	ETIS
n	0,908	0,942	0,744	n	<b>0,875</b>	<b>0,930</b>	<b>0,729</b>
s	<b>0,936</b>	0,951	<b>0,823</b>	s	0,845	0,911	0,650
m	0,925	0,927	0,790	m	0,874	0,926	0,650
l	0,909	<b>0,955</b>	0,784	l	0,847	0,911	0,624
x	0,914	0,923	0,752	x	0,850	0,914	0,633

(c) Precision				(d) Recall			
YOLO	Kvasir	CVC	ETIS	YOLO	Kvasir	CVC	ETIS
n	0,908	0,942	0,744	n	<b>0,875</b>	<b>0,930</b>	<b>0,729</b>
s	<b>0,936</b>	0,951	<b>0,823</b>	s	0,845	0,911	0,650
m	0,925	0,927	0,790	m	0,874	0,926	0,650
l	0,909	<b>0,955</b>	0,784	l	0,847	0,911	0,624
x	0,914	0,923	0,752	x	0,850	0,914	0,633

**Tabela 2. Resultados dos experimentos**

Apesar de os resultados na métrica IoU, conforme apresentados na Tabela 2(b), terem registrado valores inferiores aos obtidos na métrica *dice*, o padrão observado no *dice* também é válido para o IoU:

- A variante *n* obteve o melhor resultado para o conjunto de dados CVC-Clinic DB e o segundo melhor desempenho nos demais conjuntos;
- As variante *m* e *s* alcançaram os melhores resultados para os conjuntos Kvasir-SEG e ETIS-Larib, respectivamente.

Ao analisar os resultados para a métrica *precision*, apresentados na Tabela 2(c), é possível observar que essa foi a única métrica na qual o melhor (ou segundo melhor) resultado não foi obtido pela variante *n*. Entretanto, o melhor resultado para os *datasets* Kvasir-SEG e ETIS (e o segundo melhor para o CVC-ClinicDB) foi obtido pelo modelo *s*, caracterizado por ser a segunda menor variante do YOLOv8. O melhor resultado para o CVC-CLinicDB foi obtido pela variante *l*.

Para a métrica *recall*, é notável que a variante *n* obteve o melhor resultado para todos os conjuntos, com o segundo melhor resultado sendo alcançado pela variante *m* (e *s* no caso do ETIS-Larib).

Também é possível observar que os valores obtidos de *recall* foram inferiores aos obtidos de *precision*, indicando assim que todos os modelos avaliados apresentaram um número mais alto de falsos negativos do que falsos positivos. Isso significa que os modelos erraram mais por excesso do que por falta, predizendo como “pólipo” áreas normais da mucosa intestinal.

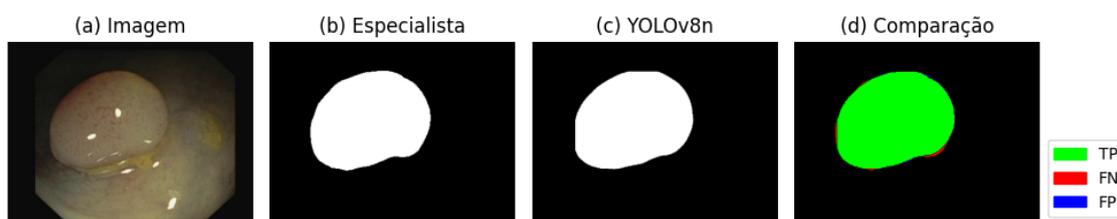
Nas quatro métricas houve um valor diminuto nos resultados obtidos sobre o conjunto ETIS-Larib Polyp DB quando comparados aos resultados dos demais *datasets*. Essa diferença pode ser explicada pela ausência de imagens deste conjunto nos exemplos de treinamento dos modelos.

No geral, é possível observar que houve poucas diferenças entre os resultados dos diferentes modelos, sendo que os modelos menores apresentaram um desempenho

ligeiramente superior. Uma vantagem adicional dos modelos menores é a necessidade reduzida de recursos computacionais e tempo tanto para o treinamento quanto para a realização das inferências.

Dessa forma, é possível apontar a variante  $n$  como a mais adequada para a segmentação de pólipos, uma vez que obteve o melhor (ou segundo melhor) resultado em três das quatro métricas avaliadas, incluindo a métrica mais relevante, o coeficiente *dice*.

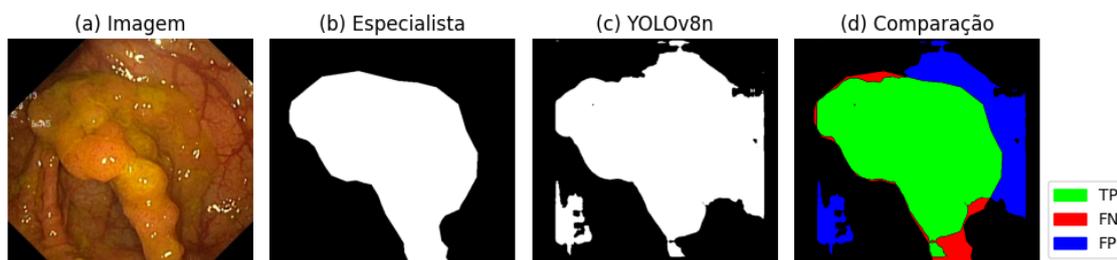
Quando comparadas com as anotações realizadas por especialistas, as predições realizadas pelo YOLOv8n ficaram bastante semelhantes em imagens com pólipos bem definidos, como é o caso da Figura 3.



**Figura 3. Comparação entre anotação realizada por especialista e predição do YOLOv8n sobre imagem do conjunto CVC-Clinic DB contendo pólipo bem definido**

Entretanto, uma limitação observada no modelo treinado ocorre em imagens com pólipos de formato irregular. Nessas situações, o algoritmo enfrentou dificuldades em realizar predições semelhantes às anotações, resultando em máscaras com áreas maiores do que o ideal, conforme ilustrado na Figura 4. Esse fenômeno pode ser a causa do *recall* mais baixo observado nas métricas de avaliação.

Todavia, é importante destacar que, nesses casos específicos, embora o algoritmo não reproduza exatamente o formato do pólipo, ele ainda pode ser útil aos profissionais de colonoscopia ao fornecer uma localização aproximada do mesmo.



**Figura 4. Comparação entre anotação realizada por especialista e predição do YOLOv8n sobre imagem do conjunto Kvasir-SEG contendo pólipo de formato irregular**

Ao realizar a comparação dos resultados obtidos pelo YOLOv8n com os modelos Unet++ [Zhou et al. 2018] e Pranet [Fan et al. 2020], conforme a Tabela 3, é possível observar que o YOLOv8n superou o desempenho dos outros modelos nos *datasets* CVC-Clinic DB e ETIS-Larib, e obteve um resultado intermediário no Kvasir-SEG.

	CVC-Clinic DB		Kvasir-SEG		ETIS-Larib	
	Dice	IoU	Dice	IoU	Dice	IoU
Unet++	0,794	0,729	0,821	0,743	0,401	0,344
Pranet	0,899	0,849	<b>0,898</b>	<b>0,840</b>	0,628	0,567
YOLOv8n	<b>0,919</b>	<b>0,877</b>	0,859	0,795	<b>0,647</b>	<b>0,585</b>

**Tabela 3. Comparação entre Unet++, Pranet e YOLOv8n**

## 5. Conclusões

Este trabalho teve como objetivo a aplicação de um modelo de segmentação em imagens de colonoscopia para a localização e delimitação de pólipos, visando auxiliar em diagnósticos médicos. Para isso, o YOLOv8 foi treinado em suas diferentes variantes em dois conjuntos de imagens, sendo seu desempenho avaliado nesses conjuntos e em um conjunto adicional.

Ao contrário do esperado, os modelos menores do YOLOv8 apresentaram resultados melhores do que os de tamanho maior, com destaque para a variante *nano*, cujos resultados conseguiram alcançar até 0,919 de *dice* e 0,877 de IoU. Além dos bons resultados atingidos por este modelo, ele também se destaca por ser o modelo mais leve em comparação com as demais variantes do YOLOv8.

Conclui-se, portanto, que a metodologia proposta tem o potencial de auxiliar profissionais de saúde no diagnóstico de pólipos, favorecendo intervenções mais precisas e oportunas. Como trabalho futuro, é interessante incluir mais imagens no conjunto de treinamento do modelo, visando aprimorar ainda mais seu desempenho.

## Referências

- Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., and Vilariño, F. (2015). Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111.
- Fan, D.-P., Ji, G.-P., Zhou, T., Chen, G., Fu, H., Shen, J., and Shao, L. (2020). Pranet: Parallel reverse attention network for polyp segmentation. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pages 263–273, Cham. Springer International Publishing.
- Haque, I. R. I. and Neubert, J. (2020). Deep learning approaches to biomedical image segmentation. *Informatics in Medicine Unlocked*, 18:100297.
- Heresbach, D., Barrioz, T., Lapalus, M.-G., Coumaros, D., Bauret, P., Potier, P., Sautereau, D., Boustière, C., Grimaud, J.-C., Barthélémy, C., Sée, J., Serraj, I., D’Halluin, P. N., Branger, B., and Ponchon, T. (2008). Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy*, 40(4):284–290.
- Inui, A., Mifune, Y., Nishimoto, H., Mukohara, S., Fukuda, S., Kato, T., Furukawa, T., Tanaka, S., Kusunose, M., Takigami, S., Ehara, Y., and Kuroda, R. (2023). Detection

- of elbow ocd in the ultrasound image by artificial intelligence using yolov8. *Applied Sciences*, 13(13).
- Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., and Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In Ro, Y. M., Cheng, W.-H., Kim, J., Chu, W.-T., Cui, P., Choi, J.-W., Hu, M.-C., and De Neve, W., editors, *MultiMedia Modeling*, pages 451–462, Cham. Springer International Publishing.
- Jocher, G., Chaurasia, A., and Qiu, J. (2023). Ultralytics yolov8.
- Keum, N. and Giovannucci, E. (2019). Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat Rev Gastroenterol Hepatol*, 16(12):713–732.
- Lalinia, M. and Sahafi, A. (2023). Colorectal polyp detection in colonoscopy images using yolo-v8 network. *Signal, Image and Video Processing*.
- Lee, S.-H., Park, Y.-K., Lee, D.-J., and Kim, K.-M. (2014). Colonoscopy procedural skills and training for new beginners. *World J Gastroenterol*, 20(45):16984–95.
- Liu, X., Song, L., Liu, S., and Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3).
- Mei, J., Zhou, T., Huang, K., Zhang, Y., Zhou, Y., Wu, Y., and Fu, H. (2023). A survey on deep learning for polyp segmentation: Techniques, challenges and future trends. *ArXiv*, abs/2311.18373.
- Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3523–3542.
- Osama, M., Kumar, R., and Shahid, M. (2023). Empowering cardiologists with deep learning yolov8 model for accurate coronary artery stenosis detection in angiography images. In *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, pages 1–6.
- Pacal, I., Karaboga, D., Basturk, A., Akay, B., and Nalbantoglu, U. (2020). A comprehensive review of deep learning in colon cancer. *Computers in Biology and Medicine*, 126:104003.
- Pandey, S., Chen, K.-F., and Dam, E. B. (2023). Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2592–2598.
- Rex, D. K., Cutler, C. S., Lemmel, G. T., Rahmani, E. Y., Clark, D. W., Helper, D. J., Lehman, G. A., and Mark, D. G. (1997). Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology*, 112(1):24–8.
- Silva, J., Histace, A., Romain, O., Dray, X., and Granado, B. (2014). Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and 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.

Terven, J. and Cordova-Esparza, D. (2023). A comprehensive review of yolo: From yolov1 and beyond.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Stoyanov, D., Taylor, Z., Carneiro, G., Syeda-Mahmood, T., Martel, A., Maier-Hein, L., Tavares, J. M. R., Bradley, A., Papa, J. P., Belagiannis, V., Nascimento, J. C., Lu, Z., Conjeti, S., Moradi, M., Greenspan, H., and Madabhushi, A., editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11, Cham. Springer International Publishing.