

Aplicação de Modelos de Aprendizado Profundo para a Segmentação Semântica de Imagens de Colonoscopia

Rubens M. G. Aguiar¹, Michel H. Scheeren², Sandro L. de Araujo Junior²,
Eduardo Mendes², Pedro L. de Paula Filho², Ricardo A. P. Franco³

¹Centro de Excelência em Inteligência Artificial (CEIA)
Goiânia – GO – Brasil

²Universidade Tecnológica Federal do Paraná (UTFPR)
Medianeira – PR – Brasil

³Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

rubensgaguiar0@gmail.com, michelhanzenscheeren@gmail.com

Abstract. *Most cases of colorectal cancer originate from colorectal polyps. An increase in the polyp detection rate reduces the risk of developing this pathology. Therefore, the implementation of neural networks to detect and segment such polyps represents a strategy that generates impactful results. Therefore, there arises a need to compare the different models available in the literature and identify the best options for polyp segmentation. Indeed, from the Kvasir dataset, models with potential were selected, and their performance training and evaluations were carried out. Effectively, 8 models were trained, among which the most efficient architectures for the segmentation of polyps in the dataset were identified, for example, the ESFPNet model achieved the highest DICE (0.9008).*

Resumo. *Grande parte dos casos de câncer colorretal tem origem em pólipos colorretais. A elevação na taxa de detecção de pólipos reduz o risco de desenvolvimento dessa patologia. Por isso, a implementação de redes neurais para detectar e segmentar tais pólipos representa uma estratégia que gera resultados impactantes. Por isso, surge a necessidade de comparar os diferentes modelos disponíveis na literatura e identificar as melhores opções para segmentação de pólipos. De fato, a partir de do conjunto de dados Kvasir, foram selecionados modelos com potencial e foram executado os treinamentos e avaliações de seus desempenhos. Efetivamente, foram treinados 8 modelos, nos quais foram identificadas as arquiteturas mais eficientes para a segmentação de pólipos no conjunto de dados, por exemplo, o modelo ESFPNet alcançou o maior DICE (0,9008).*

1. Introdução

A maioria dos casos de câncer colorretal origina-se de pólipos nessa região, sendo que um aumento na identificação destes, especialmente dos adenomatosos, apresenta uma relação inversa com o risco de desenvolver a doença. Portanto, a detecção precoce desses pólipos é fundamental para prevenir o desenvolvimento de formas graves ou fatais de câncer colorretal [Hossain et al. 2022].

Um procedimento essencial para o diagnóstico dessas condições é a colonoscopia, um procedimento que utiliza câmeras para capturar imagens do cólon e do intestino. Ela permite aos profissionais de saúde identificar pólipos colorretais através da análise de vídeos e imagens, facilitando a detecção em fases iniciais. Tal diagnóstico precoce é crucial para o tratamento eficaz do câncer, aumentando significativamente as probabilidades de cura [Hossain et al. 2022].

Por conta dos benefícios da detecção precoce de pólipos por meio de imagens de colonoscopia, diversos trabalhos foram publicados visando aumentar a taxa de detecção de pólipos, como em [Rahman and Marculescu 2023], [Huang et al. 2021], entre outros. De fato, uma das principais estratégias para a detecção e segmentação de pólipos é a utilização de modelos de *machine learning* cujo objetivo é identificar as regiões contendo pólipos de forma automática, permitindo a otimização do processo de colonoscopia através da diminuição do tempo de exame, detecção de lesões pequenas e redução da necessidade de exames adicionais [Marques et al. 2023].

Com o crescente número de trabalhos de ponta propondo modelos para detecção de pólipos, surge a necessidade de compará-los em condições equivalentes para identificar as melhores arquiteturas em termos de acurácia. Neste contexto, este trabalho pretende realizar uma comparação abrangente de diferentes redes neurais profundas aplicadas a segmentação de pólipos. Para tal, propõe-se uma homogeneização das condições de treinamento dos modelos, permitindo a validação dos resultados publicados e a identificação das arquiteturas que apresentam melhor desempenho.

Neste artigo, é apresentada uma comparação abrangente de diferentes modelos para detecção e segmentação de pólipos em imagens de colonoscopia. Para garantir uma comparação justa e confiável, todos os modelos foram treinados nas mesmas condições gerais, utilizando o mesmo conjunto de dados e as mesmas técnicas de aumento de dados. O estudo revela as melhores arquiteturas em termos de métricas de validação dos modelos, fornecendo informações valiosas para o desenvolvimento de ferramentas clínicas de diagnóstico. Enquanto a maioria dos estudos presentes na literatura se limita a comparar resultados já existentes, este trabalho se distingue ao realizar não apenas uma revisão sistemática desses, mas também a implementação e validação dos mesmos sob condições controladas, seguida de uma análise comparativa.

2. Trabalhos relacionados

A segmentação de objetos por meio de redes neurais tem sido um campo de pesquisa em constante evolução nos últimos anos. Uma das abordagens mais promissoras nesse campo é o uso de redes neurais convolucionais profundas (CNNs). A pesquisa de [Guo et al. 2017] revisa os conceitos de CNNs aplicadas à segmentação semântica e conclui que essa arquitetura de redes neurais tem o potencial de revolucionar como os algoritmos de *machine learning* interpretam imagens.

Nesse contexto, a segmentação de pólipos surge como uma tarefa a ser possivelmente completada por meio de CNNs. Assim, em [Guo et al. 2020], os autores propõem uma nova arquitetura de CNN para segmentação de pólipos chamada Dilated ResFCN. A nova arquitetura é comparada com outras CNNs de referência, e os resultados demonstram que a segmentação de pólipos é viável, pois a arquitetura proposta atinge DICE de 0,79. Além disso, o trabalho fornece uma visão geral do estado da arte em segmentação

de pólipos até 2020.

Semelhantemente, [Krishnendu et al. 2020] realizam uma revisão abrangente dos modelos de última geração para segmentação de pólipos. A revisão fornece uma visão geral das principais técnicas utilizadas, incluindo CNNs com diferentes variações, como redes neurais convolucionais piramidais (PCNNs), combinação de CNNs e redes neurais recorrentes (RNNs) e arquiteturas *encoder-decoder*. A comparação permite identificar os modelos mais eficazes e destaca que a segmentação de pólipos em imagens ainda é um desafio, apesar do alto desempenho já alcançado pelos modelos.

Ademais, [Gupta and Mishra 2024] realizaram uma revisão sistemática abrangente da literatura sobre segmentação de pólipos. A revisão fornece uma descrição detalhada das arquiteturas e técnicas utilizadas, seus respectivos resultados, os principais conjuntos de dados e suas características. A análise revela o estado da arte na segmentação de pólipos, identificando os melhores modelos e suas métricas de desempenho. A abordagem detalhada da revisão permite uma comparação coerente entre os modelos e contribui para o avanço da pesquisa na área.

Dessa forma, como observado, é fato que revisões extensivas dos principais algoritmos de redes neurais profundas já foram realizadas. Contudo, é importante ressaltar que, tais revisões somente referenciam artigos e não validam os resultados por meio de novos treinamentos em condições semelhantes. Por isso, este trabalho realiza uma confirmação dos resultados e uma comparação entre diferentes arquiteturas.

3. Materiais e métodos

Visando melhorar a organização geral do fluxo de trabalho, optou-se pela divisão do conjunto de tarefas em quatro etapas principais, que iniciam com a definição do conjunto de dados e finalizam na análise dos resultados obtidos nos treinamentos. O fluxograma mostrado na Figura 1 apresenta as etapas executadas com a finalidade de comparar de forma abrangente diferentes modelos na tarefa de segmentação de pólipos em imagens de colonoscopia.

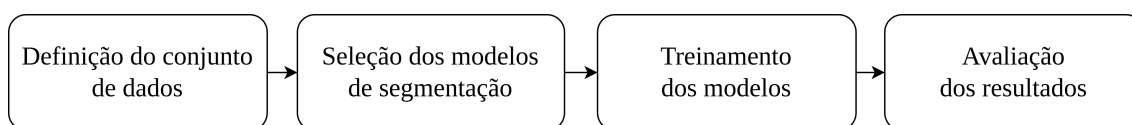


Figura 1. Fluxograma da metodologia proposta.

O conjunto de dados escolhido para os experimentos propostos foi o Kvasir-SEG¹ [Jha et al. 2020], formado por 1000 imagens de pólipos gastrointestinais e suas respectivas máscaras de segmentação, que foram manualmente anotadas e revisadas por médicos da área. As imagens foram divididas entre os conjuntos de treinamento, validação e teste, seguindo o critério 80-10-10: 80% das imagens para treinamento (800 imagens), 10% para validação (100 imagens) e 10% para teste (100 imagens). A Figura 2 apresenta alguns exemplos de imagens encontradas no Kvasir-SEG e suas respectivas máscaras de segmentação.

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

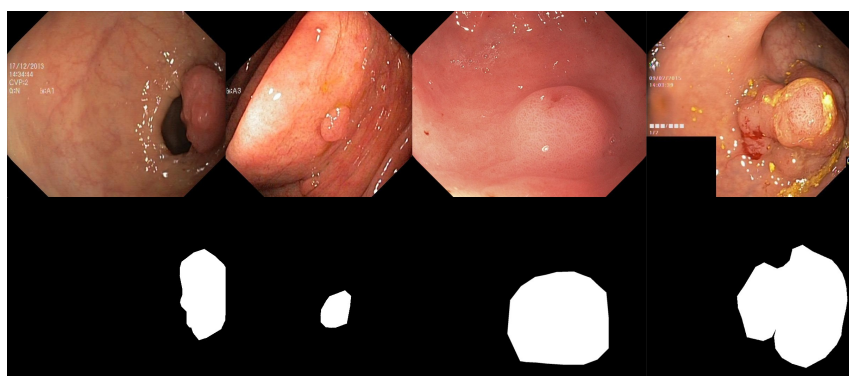


Figura 2. Exemplos do conjunto de dados Kvasir-SEG.

A etapa de definição dos modelos envolveu a pesquisa no estado da arte da tarefa de segmentação semântica de pólipos, visando identificar arquiteturas promissoras para a realização dos experimentos no conjunto de dados apresentado, incluindo modelos mais simples baseados em Redes Neurais Convolucionais (CNNs), até modelos mais avançados baseados em redes Transformers [Vaswani et al. 2017]. Foram selecionados oito modelos para a tarefa proposta: HarDNet-MSEG [Huang et al. 2021], HarDNet-DFUS [Liao et al. 2022], SSFormer [Wang et al. 2022], TGANet [Tomar et al. 2022], ColonFormer [Duc et al. 2022], ESFPNet [Chang et al. 2023], Meta-Polyp [Trinh 2023] e Cascade [Rahman and Marculescu 2023].

Para o treinamento dos modelos, optou-se pela utilização do *Google Colab*², um ambiente de desenvolvimento *online* e interativo que possibilita o acesso gratuito a GPUs como a T4, utilizada em todos os experimentos. Também foram padronizadas as técnicas de aumento de dados aplicadas no conjunto de treinamento usando a biblioteca *Albumentations*³, incluindo espelhamento (horizontal e vertical), movimentação pelos eixos (até 10%), variação de perspectiva (até 10%), escala (80 a 120% do tamanho original), ruído gaussiano, equalização, e variações dos valores originais de brilho, contraste e saturação das imagens (até 20%).

Dos modelos selecionados, apenas o Meta-Polyp possui implementação utilizando a biblioteca *Tensorflow*⁴, enquanto os demais utilizam *Pytorch*⁵. Uma vez que o hardware utilizado foi padronizado para todos os treinamentos, definiu-se que o tamanho do lote utilizado seria o maior suportado pelo ambiente para cada modelo, respeitando assim as características individuais de cada arquitetura: SSFormer e Cascade utilizaram um tamanho de lote igual a 16 por se tratarem de arquiteturas mais leves, enquanto os demais modelos foram treinados com tamanho de lote igual a 8.

Embora não exista uma padronização definitiva na literatura de segmentação de pólipos, a resolução mais utilizada nos artigos da área é de 352×352 pixels, tendo sido adotada nos experimentos. No entanto, por conta de limitações técnicas relacionadas à arquitetura do modelo HarDNet-DFUS, não foi possível utilizar a resolução de 352×352 pixels no treinamento. Desse modo, foi utilizada a resolução mais próxima suportada pelo

²<https://colab.research.google.com/>

³<https://albumentations.ai/>

⁴<https://www.tensorflow.org/?hl=pt-br>

⁵<https://pytorch.org/>

modelo (384x384 *pixels*).

Quanto ao número de épocas, o padrão geral adotado foi treinar os modelos por 200 épocas, realizando o salvamento dos pesos sempre que um valor menor de erro era obtido no conjunto de validação. As duas únicas exceções são dos modelos Cascade e Colonformer, que foram treinados por 100 e 30 épocas, respectivamente, seguindo as recomendações descritas pelos autores dos artigos originais desses modelos. A Tabela 1 resume os hiperparâmetros utilizados nos treinamentos.

Tabela 1. Resumo dos hiperparâmetros utilizados nos treinamentos

Modelo	Tamanho do lote	Resolução	Épocas
HarDNet-MSEG	8	352x352	200
HarDNet-DFUS	8	384x384	200
SSFormer	16	352x352	200
TGANet	8	352x352	200
ColonFormer	8	352x352	30
ESFPNet	8	352x352	200
Meta-Polyp	8	352x352	200
Cascade	16	352x352	100

Avaliar com precisão o desempenho dos modelos é uma etapa crucial para identificar a real eficácia das arquiteturas treinadas, e diversas métricas foram propostas na literatura para este fim. Para a avaliação dos modelos treinados, foram selecionadas quatro métricas amplamente utilizadas na literatura de segmentação de pólipos: o coeficiente DICE, a intersecção sobre união (IoU), a precisão e a cobertura. Nas equações apresentadas, VP representam os Verdadeiros Positivos, FP representam os Falsos Positivos e FN os Falsos Negativos.

O coeficiente DICE, também conhecido como *F1 score*, calcula a divisão entre a área comum e a área de união de dois conjuntos, representados pela máscara de segmentação anotada e pela máscara predita pelo modelo. Apresenta resultados que variam entre 0 e 1, em que 1 significa máscaras idênticas e 0 indica máscaras completamente diferentes. A Equação 1 apresenta a fórmula matemática do coeficiente de Dice [Liu et al. 2021]

$$DICE = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \quad (1)$$

A métrica intersecção sobre união (IoU), ou ainda *Jaccard Index*, também mede a região de sobreposição entre dois conjuntos, assim como o coeficiente de DICE. No entanto, o IoU expressa essa sobreposição como uma razão entre o tamanho da intersecção e o tamanho da união dos dois conjuntos. A Equação 2 apresenta a fórmula matemática do IoU [Dumitru et al. 2023].

$$IoU = \frac{VP}{VP + FP + FN} \quad (2)$$

A precisão indica a proporção de *pixels* preditos que correspondem aos *pixels* anotados, ou seja, mede a capacidade do modelo de identificar corretamente os objetos na imagem. Já a cobertura mede a proporção dos *pixels* anotados que foram realmente preditos pelo modelo, ou seja, indica a capacidade do modelo de encontrar todos os objetos relevantes na imagem. As Equações 3 e 4 apresentam as fórmulas da precisão e cobertura, respectivamente [Faceli et al. 2011].

$$Precisão = \frac{VP}{VP + FP} \quad (3)$$

$$Cobertura = \frac{VP}{VP + FN} \quad (4)$$

4. Resultados e discussão

Esta seção visa apresentar e analisar os resultados obtidos pelos modelos de segmentação HarDNet-MSEG, HarDNet-DFUS, SSFormer, TGANet, ColonFormer, ESFPNet, Meta-Polyp e Cascade no conjunto Kvasir-SEG usando as métricas Dice, IoU, precisão e cobertura. A Tabela 2 sintetiza os resultados obtidos nos experimentos realizados. Para facilitar a visualização dos dados, o melhor resultado de cada métrica foi negrito e o segundo melhor resultado foi sublinhado. Além disso, a tabela foi ordenada do maior para o menor DICE, por tratar-se de uma das métricas mais completas e utilizadas na literatura.

Tabela 2. Resultados obtidos pelos modelos no conjunto de teste do Kvasir-SEG

Modelo	DICE	IoU	Precisão	Cobertura
ESFPNet	0,9008	0,8465	0,8967	0,9458
Meta-Polyp	<u>0,8991</u>	0,8399	<u>0,9224</u>	0,9118
HarDNet-MSEG	0,8977	<u>0,8451</u>	0,9319	0,9064
SSFormer	0,8965	0,8421	0,9007	<u>0,9341</u>
Cascade	0,8940	0,8430	0,9123	0,9123
TGANet	0,8923	0,8392	0,9015	0,9274
HarDNet-DFUS	0,8896	0,8336	0,9133	0,9148
ColonFormer	0,8620	0,7928	0,9013	0,8779

O modelo que obteve o melhor desempenho geral foi o ESFPNet, apresentando o maior DICE (0,9008) e o maior IoU (0,8465) dentre os experimentos realizados. Embora também tenha obtido a maior Cobertura (0,9458), foi o modelo que apresentou a menor precisão geral (0,8967), o que indica que embora apresente a melhor capacidade geral de identificar o maior número de pólipos nas imagens do conjunto de testes, também é o

responsável pelo maior índice de falsos positivos, que são regiões que o modelo acredita pertencerem à classe de pólipos, mas não pertencem. Porém, isso pode ser vantajoso pois torna o modelo mais cauteloso, optando por prever casos positivos em situações de incerteza.

A segunda colocação ficou com o modelo Meta-Polyp, com DICE de 0,8991 e IoU de 0,8399, resultados ligeiramente abaixo do ESFPNet (0,0017 pontos e 0,0066 pontos, respectivamente). Por outro lado, o modelo apresentou uma variabilidade menor nos valores de precisão e cobertura (0,9224 e 0,9118), o que pode indicar uma alternativa mais equilibrada em termos de capacidade de identificação de pólipos e confiabilidade nos resultados identificados.

Completando o ranking dos três melhores resultados obtidos, tem-se o modelo HarDNet-MSEG que apresentou um DICE de 0,8977 e um IoU de 0,8451 (apenas 0,0031 e 0,014 pontos abaixo do primeiro colocado, respectivamente). Este foi o modelo que apresentou a maior precisão geral dos experimentos, embora tenha a penúltima menor cobertura, o que pode fornecer um menor índice geral de identificação dos pólipos relevantes nas imagens, embora apresente um maior percentual de assertividade nos objetos identificados.

O modelo com o pior desempenho nos experimentos realizados foi o ColonFormer, com DICE de 0,8620 e IoU de 0,7928, sendo respectivamente 0,0388 e 0,0537 pontos abaixo dos resultados obtidos pelo ESFPNet, o que representa uma variação significativa. Embora o próprio artigo original recomende que o modelo seja treinado por um número baixo de épocas, isso pode ter atrapalhado os resultados obtidos no experimento, uma vez que ele pode não ter convergido totalmente.

De forma geral, os resultados foram próximos ao estado da arte da área de segmentação de pólipos para o conjunto de dados Kvasir-SEG, assim como também se assemelham aos resultados originais apresentados pelos artigos dos modelos que foram utilizados nos experimentos. Podem ser citados como exemplos os modelos HarDNet-MSEG (DICE de 0,9040 e um IoU de 0,8480), HarDNet-DFUS (DICE de 0,918) e TGA-Net (DICE igual a 0,9029 e IoU de 0,8330).

A Figura 3 mostra a curva ROC de oito modelos no conjunto de testes do Kvasir-SEG, permitindo avaliar o desempenho dos modelos em relação aos falsos positivos e verdadeiros positivos conforme variam os limiares de decisão aplicados às predições dos modelos. Ademais, um classificador ideal busca maximizar a taxa de verdadeiros positivos e minimizar a de falsos positivos. Assim, as curvas dos modelos ESFPNet, HarDNet-MSEG e SSFormer se aproximam mais de um classificador ideal, enquanto as curvas de Cascade, Colonformer e HardNet-DFUS se distanciam, indicando menor eficácia. Por outro lado, nota-se que a escolha do limiar de predição influencia na comparação entre os modelos, como visto no Meta-Polyp, que, apesar de ter o segundo melhor DICE, de acordo com a Tabela 2, não está entre os três melhores modelos da curva ROC.

A Figura 4 faz uma comparação da imagem original e máscara verdadeira (1ª e 2ª colunas) com o resultado gerado em cada um dos experimentos para quatro imagens do conjunto de testes, com os modelos ordenados da esquerda para a direita do maior para o menor DICE. De forma geral, é possível perceber que as primeiras colunas apresentam maior semelhança em relação à máscara verdadeira, incluindo os contornos e detalhes

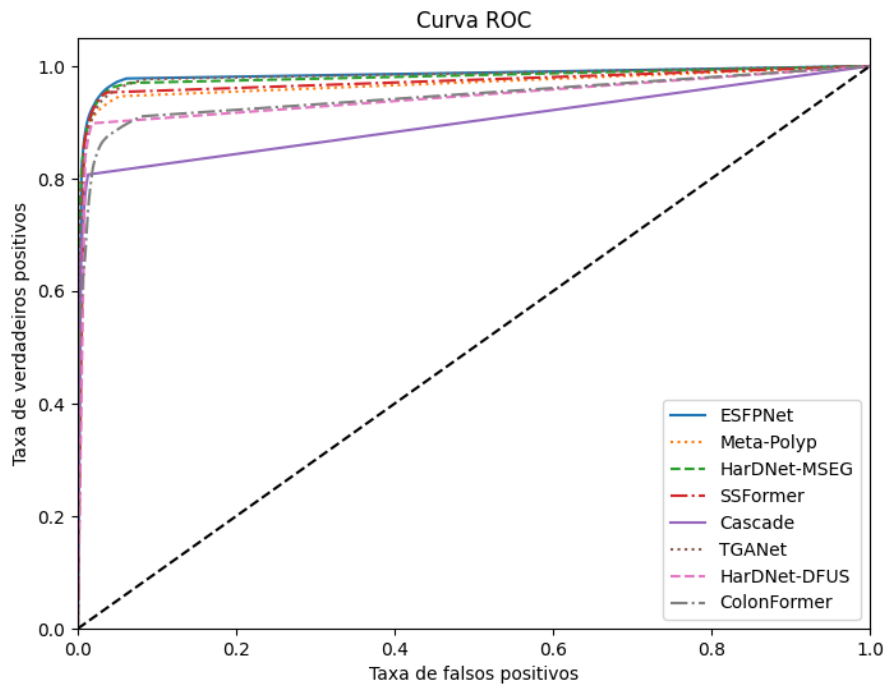


Figura 3. Curva ROC dos modelos no conjunto de testes do Kvasir-SEG

presentes na imagem original, assim como todos os pólipos identificados nos casos em que existem mais de um na imagem (terceira linha). Por outro lado, as inferências mais à direita apresentam diferenças mais significativas, incluindo elementos ausentes na imagem, pólipos identificados parcialmente ou confusão com o fundo da imagem.

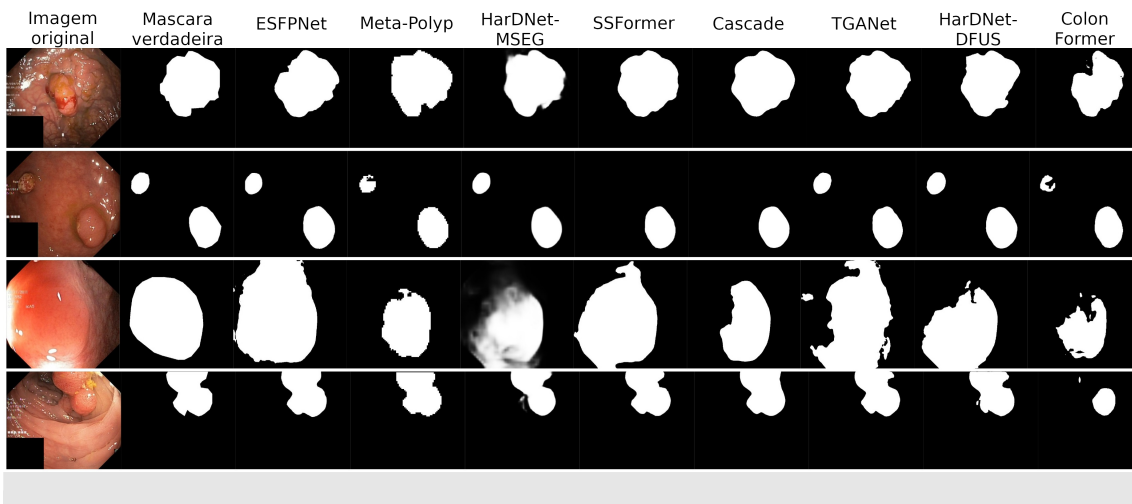


Figura 4. Inferência dos modelos em quatro imagens do conjunto de testes

5. Conclusão e Trabalhos Futuros

A identificação precoce dos pólipos colorretais é crucial para permitir um tratamento eficaz e mitigar os eventuais riscos envolvidos, e para isso os modelos de aprendizado

profundo se mostram uma solução promissora. Dessa forma, surge a necessidade de comparar os diversos modelos disponíveis na literatura para permitir a seleção das melhores opções para aplicações na área de segmentação de pólipos. Assim, é perceptível que a metodologia adotada para treinar os modelos, utilizando um conjunto de dados pré-estabelecido e condições semelhantes de treinamento, possibilita a validação dos resultados e uma comparação efetiva entre os diferentes modelos.

De fato, foi constatado que a rede ESFPNet obteve o melhor desempenho geral entre os experimentos realizados neste artigo quanto ao DICE, IoU e cobertura, embora o modelo HarDNet-MSEG tenha apresentado a maior precisão. Além disso, os objetivos do artigo foram alcançados, uma vez que todos os modelos foram devidamente validados e os resultados alcançados aproximam-se do estado da arte da área de segmentação semântica de pólipos, permitindo comparar eficientemente os desempenhos entre os modelos e identificar os mais eficazes.

Dessa forma, os resultados apresentados no presente trabalho podem ser utilizados como uma fonte de comparação abrangente de diversos modelos de aprendizado de máquina na tarefa de segmentação semântica de pólipos em imagens de colonoscopia, fornecendo resultados confiáveis e com potencial para guiar a definição de alternativas viáveis para o desenvolvimento de soluções na área da saúde, capazes de auxiliar em diagnósticos clínicos e agilizar a realização de exames por imagem, contribuindo assim para a identificação precoce de pólipos.

Como uma forma de melhorar ainda mais a acurácia e a confiabilidade dos resultados apresentados neste artigo, é imprescindível incluir outras bases de dados no treinamento e avaliação dos modelos, com objetivo de aumentar a variabilidade de exemplos disponíveis para a construção do aprendizado, além de permitir a identificação da capacidade de generalização do conhecimento. Essa é uma prática comum na literatura da área, em que também são utilizadas bases de dados como o CVC-ClinicDB⁶ [Vázquez et al. 2016], ETIS-Larib Polyp DB⁷ [Silva et al. 2014], entre outros. Esse é o caso por exemplo dos resultados apresentados pelos artigos originais dos modelos MetaPolyp (DICE de 0,9590 e um IoU de 0,921) e ESFPNet (DICE de 0,931 e IoU de 0,8870), em que tanto o Kvasir-SEG quanto o CVC-ClinicDB foram utilizados no treinamento.

Referências

- Chang, Q., Ahmad, D., Toth, J., Bascom, R., and Higgins, W. E. (2023). Espfnet: Efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video. In *Medical Imaging 2023, Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, United States. SPIE.
- Duc, N. T., Oanh, N. T., Thuy, N. T., Triet, T. M., and Dinh, V. S. (2022). Colonformer: An efficient transformer based method for colon polyp segmentation. *IEEE Access*, 10:80575–80586.
- Dumitru, R.-G., Peteleaza, D., and Craciun, C. (2023). Using duck-net for polyp image segmentation. *Scientific Reports*.

⁶<https://polyp.grand-challenge.org/CVCClinicDB/>

⁷<https://polyp.grand-challenge.org/ETISLarib/>

- Faceli, K., Lorena, A. C., Gama, J., and Carvalho, A. C. P. L. F. D. (2011). *Inteligência Artificial: uma abordagem de aprendizado de máquina*. LTC, Rio de Janeiro, RJ.
- Guo, Y., Bernal, J., and Matuszewski, B. J. (2020). Polyp segmentation with fully convolutional deep neural networks - extended evaluation study. *Journal of Imaging*.
- Guo, Y., Liu, Y., Georgiou, T., and Lew, M. S. (2017). A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*.
- Gupta, M. and Mishra, A. (2024). A systematic review of deep learning based image segmentation to detect polyp. *Artificial Intelligence Review*.
- Hossain, M., Karuniawati, H., Jairoun, A., Urbi, Z., Ooi, J., John, A., Lim, Y., Kibria, K., Mohiuddin, A., Ming, L., Goh, K., and Hadi, M. (2022). Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. *Cancers (Basel)*.
- Huang, C.-H., Wu, H.-Y., and Lin, Y.-L. S. (2021). Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. *ArXiv*, abs/2101.07172.
- 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.
- Krishnendu, S., Geetha, S., and Gopakumar, G. (2020). A review on polyp detection and segmentation in colonoscopy images using deep learning. *International Journal of Engineering Research & Technology*.
- Liao, T.-Y., Yang, C.-H., Lo, Y.-W., Lai, K.-Y., Shen, P.-H., and Lin, Y.-L. (2022). Hardnet-dfus: An enhanced harmonically-connected network for diabetic foot ulcer image segmentation and colonoscopy polyp segmentation.
- Liu, X., Song, L., Liu, S., and Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, 13(3).
- Marques, A. F., Marques, K. F., dos Santos Beraldo, M. N. M., Lima, T. B., Sasaki, L. Y., and Beraldo, R. F. (2023). Inteligência artificial na colonoscopia no rastreamento do câncer colorretal: revisão de literatura. *Brazilian Journal of Health Review*.
- Rahman, M. M. and Marculescu, R. (2023). Medical image segmentation via cascaded attention decoding. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6211–6220.
- 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*.
- Tomar, N. K., Jha, D., Bagci, U., and Ali, S. (2022). Tganet: Text-guided attention for improved polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 151–160, Cham. Springer Nature Switzerland.
- Trinh, Q.-H. (2023). c. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 742–747, Los Alamitos, CA, USA. IEEE Computer Society.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Vázquez, D., Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., López, A. M., Romero, A., Drozdal, M., and Courville, A. C. (2016). A benchmark for endoluminal scene segmentation of colonoscopy images. *Journal of Healthcare Engineering*, 2017.
- Wang, J., Huang, Q., Tang, F., Meng, J., Su, J., and Song, S. (2022). Stepwise feature fusion: Local guides global. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 110–120, Cham. Springer Nature Switzerland.