

Ferramenta para auxílio na rotulação de *datasets* para segmentação semântica

Sandro Luis de Araujo Junior*, Jakeline da Silva Andrade*, Pedro Luiz de Paula Filho*, Ricardo Franco†

*Universidade Tecnológica Federal do Paraná (UTFPR)

Medianeira, Brasil

Email: sandro.1999@alunos.utfpr.edu.br, jakelineandrads@gmail.com, pedrol@utfpr.edu.br

†CEIA - Instituto de Informática, Universidade Federal de Goiás (INF/UFG)

Goiânia, Brasil

Email: ricardofranco@ufg.br

Abstract—This study aims to develop a tool to assist in annotating images for segmentation purposes, using SAM. To achieve this goal, we chose to use the Python programming language, which is SAM's standard language. The selected development and testing environment was Google Colab. To validate the developed tool, two sets of images were chosen. The first one is the Kvasir-SEG dataset, which contains 1000 images of polyps obtained from colonoscopy exams. The second set consists of images of lettuce grown in a hydroponic system. As a result, the tool achieved good performance in both sets of images, demonstrating its potential in aiding manual annotations for image segmentation.

Keywords—image segmentation; image annotation; deep learning.

Resumo—Este estudo tem como objetivo desenvolver uma ferramenta que auxilie na anotação de imagens para fins de segmentação, utilizando o SAM. Para alcançar esse propósito, optou-se por utilizar a linguagem de programação Python, que é a linguagem padrão do SAM. O ambiente de desenvolvimento e teste escolhido foi o Google Colab. Para validar a ferramenta desenvolvida, foram selecionados dois conjuntos de imagens. O primeiro é o *dataset* Kvasir-SEG, que contém 1000 imagens de pólipos obtidas de exames de colonoscopia. O segundo conjunto consiste em imagens de alfaces cultivadas em sistema hidropônico. Como resultado, a ferramenta obteve bom desempenho em ambos os conjuntos de imagens, demonstrando ter potencial no auxílio de anotações manuais para segmentação de imagens.

Palavras-chave—segmentação de imagens; anotação de imagens; aprendizado profundo..

I. INTRODUÇÃO

A capacidade inata de capturar, processar e interpretar informações adquiridas por meio da visão tem impulsionado a criação de técnicas sofisticadas na área da computação, com o intuito de permitir que a tecnologia exerça essas funções de maneira análoga ao ser humano, porém com uma eficiência consideravelmente maior. Entretanto, executar tais processos não é uma tarefa trivial. Por conta disto, são empregadas abordagens específicas com o objetivo de facilitar esses processos, entre as quais se inclui a segmentação de imagens [1].

A segmentação é um método que engloba a subdivisão de regiões ou objetos com atributos específicos em uma imagem digital [2]. Esta técnica desempenha um papel primordial e insubstituível no campo do processamento digital de imagens, uma vez que dela emerge a extração de informações que viabilizam análises subsequentes [3].

Dentro do âmbito da segmentação de imagens utilizando técnicas de aprendizado profundo, existem tipos distintos, como a segmentação semântica, por instâncias e panóptica. Na segmentação semântica, cada píxel da imagem é associado a uma classe correspondente [4]. Em uma abordagem mais clara, todos os elementos presentes na imagem são nomeados. Elementos que compõem um mesmo grupo recebem a mesma designação. No entanto, este método não efetua a separação entre objetos pertencentes à mesma classificação. Por exemplo, em uma imagem contendo a representação de vários cães, a segmentação semântica não seria capaz de separar individualmente cada animal, apenas os agruparia sob uma mesma classe.

Na segmentação por instâncias, ocorre uma abordagem mais refinada, na qual é realizada a separação por classes, de forma similar à segmentação semântica, com a capacidade adicional de separar individualmente os objetos. Em termos simples, elementos pertencentes ao mesmo grupo recebem rótulos distintos. Contudo, objetos que não podem ser contabilizados e que geralmente compõem o fundo da imagem, como grama, areia e outros, não recebem nenhum tipo de classificação [5], [6].

Por fim, a segmentação panóptica consiste na fusão das abordagens de segmentação semântica e segmentação por instâncias, gerando informações mais complexas, tais como a identificação do tipo do objeto representado, a separação entre esses objetos e também a rotulação de objetos como chão, grama, areia e outros [7].

Embora a segmentação seja um método facilitador, no qual auxilia na obtenção de dados, existem diversos obstáculos que



dificultam sua implementação. Conforme apontado por [8], a segmentação é um desafio de percepção psicofísica, sendo, portanto, uma questão que não pode ser abordada somente por meio de soluções analíticas.

Outra questão que apresenta desafios significativos na segmentação é a rotulação de *datasets*, devido à necessidade de encontrar materiais específicos, o que muitas vezes depende de fontes externas. Quando os dados disponíveis não atendem às necessidades para resolver o problema, torna-se necessário criar conjuntos de dados próprios. Além disso, a rotulação manual é um processo que tende a ser lento e custoso. Realizar esse procedimento por meio da intervenção humana consome tanto tempo quanto recursos financeiros, o que pode impactar negativamente a eficiência e a viabilidade de outras tarefas mais relevantes [9].

Para reduzir a ocorrência desse tipo de cenário, é interessante a utilização de tecnologias que realizam esse processo de maneira eficaz e eficiente, sem que seja necessário que o algoritmo aprenda a gerar máscaras e classes para novas imagens. Nesse sentido, a adoção do *Segment Anything Model* (SAM) pode fornecer suporte e gerar resultados significativos.

O SAM é uma abordagem de Inteligência Artificial utilizada para segmentar objetos, adotando uma perspectiva de generalização *zero-shot*. Isso implica que o modelo não requer treinamento específico para reconhecer objetos inteiramente novos. Desenvolvido pela empresa Meta, o SAM foi concebido para realizar a segmentação de qualquer objeto previamente especificado nos *prompts* de entrada, sejam eles provenientes de imagens ou vídeos. Vale ressaltar que esse tipo de segmentação não se encaixa em nenhuma das classificações mencionadas anteriormente (segmentação semântica, panóptica e por instâncias). Além disso, o SAM se destaca por permitir uma integração flexível com outros sistemas e a capacidade de efetuar segmentação de texto para objetos [10].

Existem diversos estudos relacionados à utilização do SAM para a segmentação de objetos. Um desses estudos é apresentado por [11], no qual foi avaliada a competência do SAM em segmentar imagens médicas, utilizando uma composição de 19 conjuntos de imagens distintos. Os resultados alcançados mostraram diversas observações significativas, como a constatação de um melhor desempenho na segmentação de objetos circunscritos quando *prompts* com menor ambiguidade foram empregados. Além disso, o estudo identificou um desempenho superior do SAM em comparação com métodos similares, como RITM [12], SimpleClick [13] e FocalClick [14].

Em [15], é conduzida uma análise comparativa entre as abordagens SAM e *Brain Extraction Tool* (BET) [16] para a segmentação de imagens de ressonância magnética. Nesse estudo, as métricas utilizadas para realizar o paralelo entre

essas técnicas incluem o *Intersection over Union* (IoU) e o Coeficiente Dice. Os resultados obtidos pelo SAM foram um IoU de 0,842 e um dice de 0,914, contra o BET que atingiu 0,518 de IoU e 0,628 de Dice. Mesmo com o SAM sendo uma abordagem relativamente nova, sua capacidade de produzir resultados promissores sugere que ele pode emergir como uma ferramenta poderosa e precisa para auxiliar na segmentação de elementos presentes no cérebro.

No estudo conduzido por [17], é realizada a detecção universal de crateras utilizando o método SAM. Segundo essa pesquisa, as crateras são características morfológicas de grande importância no estudo de planetas. No entanto, contar e identificar manualmente esses objetos, é uma tarefa trabalhosa. Para solucionar esse problema, foi aplicada tecnologia para mapear e detectar crateras automaticamente. Como resultado, a utilização do método SAM possibilitou a detecção de objetos com aparência de crateras de maneira notável, em diferentes tipos de dados, incluindo imagens de satélite e com variações de ângulos de observação.

Os exemplos fornecidos evidenciam a eficácia do método SAM na segmentação de diversas categorias de objetos, ultrapassando em desempenho outros tipos de modelos de segmentação e eliminando a necessidade de um conjunto de dados previamente rotulado para executar tais tarefas. Com o intuito de contribuir para essa área de estudo, esta pesquisa tem como objetivo desenvolver uma ferramenta destinada a auxiliar na anotação de imagens para fins de segmentação utilizando o SAM.

Nas próximas seções, são detalhados os materiais e métodos empregados, expondo a metodologia adotada e os passos seguidos para a efetuação deste estudo, na sequência os resultados obtidos e, por fim, a conclusão alcançada.

II. MATERIAIS E MÉTODOS

A metodologia utilizada para o desenvolvimento da ferramenta que facilite a anotação de imagens para fins de segmentação pode ser observada no fluxograma apresentado na Figura 1.

O primeiro passo foi estabelecer um modelo de aprendizado profundo capaz de gerar máscaras a partir de entradas mais simples. Nesse contexto, optou-se pela utilização do SAM [18], uma arquitetura de segmentação de código aberto e abrangente.

O SAM traz ao contexto de segmentação um novo tipo de tarefa, chamada de segmentação baseada em *prompt* (*promptable segmentation task*), que diferentemente das segmentações semântica, de instâncias e panóptica, não associa uma classe às áreas segmentadas, mas sim gera uma máscara de segmentação válida a partir de um *prompt* [18].

O *prompt* é a especificação de quais áreas devem ser segmentadas, e pode ser expresso na forma de pontos, caixas

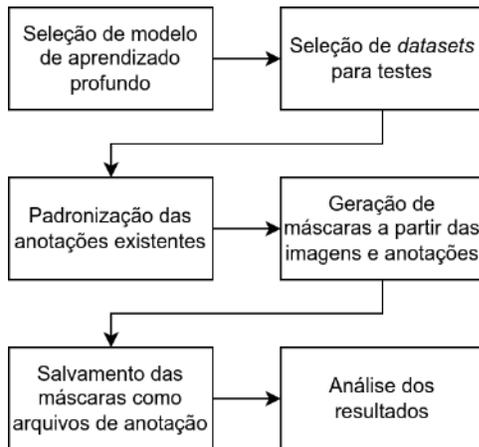


Fig. 1. Autoria Própria - Fluxograma de atividades

delimitadoras (conhecidas como *bounding boxes*), ou ainda como um texto descritivo. Mesmo em casos no qual um *prompt* é ambíguo e pode se referir a mais de um objeto, o SAM produz uma máscara para pelo menos um desses objetos [18].

O SAM se destaca por ser um modelo pré-treinado, possuindo a capacidade de generalização para objetos que não foram apresentados durante a fase de treinamento, na qual foram utilizadas 11 milhões de imagens, contendo mais de 1 bilhão de máscaras [18].

Para realizar a codificação da ferramenta nos processos de leitura das imagens, processamento das anotações existentes, comunicação com o SAM e processamento dos resultados, foi utilizada a linguagem Python¹, que é uma linguagem de programação de código aberto, alto nível, multiparadigma, com tipagem dinâmica, e multiplataforma. A escolha se deu por essa ser a linguagem utilizada pelo SAM.

O desenvolvimento e os testes foram realizados no ambiente Google Colab². Esse ambiente proporciona ao usuário a possibilidade de escrever e executar código Python diretamente no navegador, sem necessidade de configuração adicional. Além disso, uma das vantagens do Colab é a possibilidade de acesso gratuito a GPUs (*Graphics Processing Unit*), o que permite a execução de tarefas paralelas em um menor tempo de execução.

Outra etapa importante realizada foi a escolha dos conjuntos de imagens utilizados para realização dos testes da ferramenta desenvolvida. Pelo fato de uma das principais áreas de utilização de técnicas de segmentação ser a análise de imagens médicas [5], optou-se pela utilização do conjunto de imagens Kvasir-SEG³ [19], que contém imagens de colonoscopia nas

¹<https://www.python.org/>

²<https://colab.research.google.com/>

³<https://datasets.simula.no/kvasir-seg/>

quais pode ser observada a presença de pólipos.

Os pólipos são crescimentos anormais da mucosa intestinal que podem evoluir para câncer colorretal [20], o terceiro tipo de câncer mais comum no mundo, e o segundo mais letal [21]. Portanto, a tarefa de segmentação de pólipos pode desempenhar um papel crucial no diagnóstico, permitindo a identificação e remoção precoce dos pólipos, contribuindo para a prevenção do câncer colorretal.

O *dataset* Kvasir-SEG é constituído por 1000 imagens que foram coletadas e avaliadas por gastroenterologistas do Hospital Vestre Viken Health Trust, localizado na Noruega. Além das imagens, o conjunto de dados também inclui máscaras de segmentação anotadas manualmente por um engenheiro e um médico, que foram posteriormente avaliadas por um gastroenterologista para garantir a precisão.

As máscaras estão no formato de imagens binárias, nas quais os pixels brancos representam os pólipos, e os pixels pretos o fundo da imagem. Além disso, o conjunto de dados disponibiliza *bounding boxes*, que foram geradas a partir das máscaras de segmentação. Na Figura 2 são apresentadas algumas imagens do Kvasir-SEG, na qual a primeira coluna apresenta a imagem original, a segunda coluna as máscaras de segmentação, e a terceira coluna as *bounding boxes* sobre a imagem original.

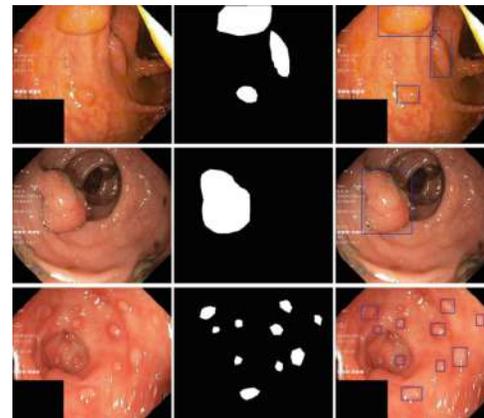


Fig. 2. Exemplo de imagens do Kvasir-SEG juntamente com as máscaras anotadas e as *bounding boxes* [19]

Com o objetivo de testar a eficácia da ferramenta em um cenário de uso completamente diferente, também foi utilizada uma base de imagens de alfaces (*Lactuca sativa L.*) cultivadas por meio de hidroponia, disponibilizada por [22]. A aquisição das imagens foi realizada utilizando uma câmera multi HD do modelo INTELBRAS VHD 3230 B G4, disposta a uma altitude aproximada de 2,70 metros. Este conjunto possui anotações no formato de *bounding boxes* que foram geradas a partir de técnicas de processamento de imagens. Na Figura 3 é

apresentada uma imagem deste conjunto juntamente com as *bounding boxes*.



Fig. 3. Exemplo de imagem do conjunto de imagens de alface com *bounding boxes* [22].

Dentre as hortaliças cultivadas no Brasil, a alface ganha destaque devido à sua alta presença na mesa dos brasileiros. Isso confere à mesma grande relevância no âmbito social e econômico, contribuindo para a geração de empregos [23]. A alface é uma hortaliça com uma taxa elevada de perecibilidade, e por isso é cultivada geralmente próxima aos grandes centros consumidores. Utilizar a alface como um produto implica a necessidade de que ela atenda aos parâmetros de qualidade, pois caso não esteja dentro desses padrões, pode resultar em prejuízos para o produtor [24]. Para evitar esse cenário, é interessante realizar um diagnóstico visual das plantas durante o estágio de desenvolvimento, a fim de identificar possíveis deficiências em seu crescimento. Portanto, a tarefa de segmentação de alfaces pode auxiliar nessa área, uma vez que, por meio da segmentação, é possível extrair informações e, posteriormente, realizar análises de seu desenvolvimento.

Conforme destacado por [11], o SAM demonstra um desempenho superior quando se utilizam caixas delimitadoras em comparação aos demais tipos de *prompt*. Portanto, foi decidido empregar as caixas delimitadoras que acompanham os conjuntos de imagens como entrada para o SAM, a fim de realizar a previsão das máscaras.

Contudo, as *bounding boxes* dos dois conjuntos não se encontram em um formato padronizado. Logo, foi essencial uniformizar sua representação para um formato de anotação amplamente reconhecido na literatura. Nesse contexto, foi decidido adotar o formato empregado pelo conjunto de imagens COCO⁴ [25].

A escolha desse formato se justifica pela ampla adoção deste conjunto na comunidade de visão computacional, sendo frequente utilizado para tarefas como segmentação [5] e detecção de objetos [26].

⁴<https://cocodataset.org/>

Com essa padronização implementada, a ferramenta agora se torna versátil, capaz de trabalhar não apenas com os conjuntos de imagens específicos deste projeto, mas também com qualquer conjunto que siga o formato COCO.

O código que realiza esta conversão foi implementado na linguagem Python em conjunto com a biblioteca json⁵, utilizada para fazer a leitura e escrita dos arquivos de anotação.

Após o preparo das anotações, estas já podem ser carregadas pela ferramenta principal, que realiza a sua leitura e itera pelos arquivos de imagem por ela indicados. Os arquivos de imagem são carregados do disco para a memória principal utilizando a biblioteca OpenCV⁶ (*Open Source Computer Vision Library*), que é uma biblioteca de código aberto amplamente utilizada para o desenvolvimento de aplicativos de visão computacional.

As imagens lidas pelo sistema são então enviadas para o SAM juntamente com as *bounding boxes*, para que este realize a predição de máscaras dentro das áreas delimitadas. As máscaras geradas são concatenadas usando a biblioteca NumPy⁷ e, posteriormente, o OpenCV é empregado para salvá-las como imagens binárias em disco. Essas máscaras podem ser refinadas manualmente em outras ferramentas de anotação, como por exemplo, o CVAT⁸, ou então serem utilizadas diretamente para o treinamento de algoritmos de segmentação semântica.

No caso do conjunto de imagens Kvasir-SEG, que já possui máscaras anotadas manualmente, foi possível fazer uma análise comparativa entre as máscaras originais e aquelas geradas pelo SAM para cada arquivo. Para essa avaliação, foram utilizadas o *Intersection over Union* (IoU) e o Coeficiente de Similaridade Dice, duas das principais métricas de avaliação de segmentação de imagens conforme [5], e indicadas pelos próprios desenvolvedores do *dataset*.

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 interseção pela união entre as duas áreas (conforme a Equação 1), e seu valor varia de 0 a 1. No caso da segmentação de imagens, o IoU é calculado entre as máscaras preditas e anotadas [5].

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

O Coeficiente de Similaridade Dice (Equação 2) é a métrica mais utilizada para a análise de imagens médicas no contexto de segmentação. Ele mede a similaridade entre duas áreas, e seu valor também varia de 0 a 1 [27].

⁵<https://docs.python.org/pt-br/3/library/json.html>

⁶<https://opencv.org/>

⁷<https://numpy.org/>

⁸<https://www.cvat.ai/>

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

No caso das imagens de alface, não foi possível o cálculo das métricas, uma vez que esse *dataset* não oferece máscaras de segmentação anotadas. Entretanto, para realizar uma análise visual da qualidade da anotação gerada, foi utilizada a biblioteca Matplotlib⁹, para desenhar as *bounding boxes* e as máscaras preditas em cima da imagem original, conforme demonstrado na Figura 4, na qual as máscaras geradas são apresentadas em coloração alaranjada.



Fig. 4. Exemplo de máscaras geradas para o conjunto da alface [22].

III. RESULTADOS

Após o processamento das imagens do conjunto Kvasir-SEG juntamente com as respectivas caixas delimitadoras, foi realizado o cálculo das métricas de avaliação entre as máscaras de segmentação geradas pelo algoritmo e as máscaras anotadas manualmente disponíveis no *dataset*.

O Índice de Jaccard (IoU) médio obtido foi de 0,849, enquanto o Coeficiente de Similaridade Dice médio atingiu 0,907, evidenciando uma taxa significativa para apoiar o processo de anotação de imagens. Esses resultados implicam que os anotadores humanos teriam apenas algumas imagens e áreas para realizarem ajustes manuais, destacando a eficácia do método proposto.

Ao fazer uma análise visual das predições, foi possível identificar que nos casos em que o pólipo possui limites bem definidos, como mostrado na Figura 5, a máscara gerada pelo SAM possui contornos mais suaves do que a máscara original.

Entretanto, quando o pólipo possui uma superfície mais plana, que se assemelha ao fundo da imagem, o SAM apresentou dificuldade em contornar o objeto de interesse, conforme exibido na Figura 6.

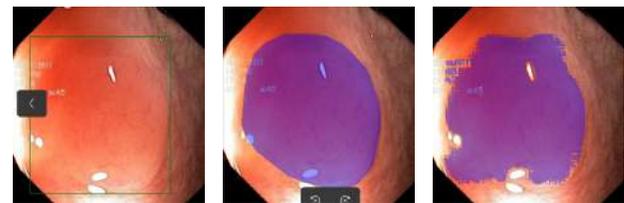
No conjunto de imagens de alfaces, ao se realizar a análise visual das máscaras geradas, foi constatado que as mesmas

⁹<https://matplotlib.org/>



(a) *Bounding box* (b) *Máscara original* (c) *Máscara predita*

Fig. 5. Predição realizada sobre pólipo bem definido



(a) *Bounding box* (b) *Máscara original* (c) *Máscara predita*

Fig. 6. Predição realizada sobre pólipo semelhante ao fundo

possuem ótima qualidade nos casos em que caixas delimitadoras cercaram precisamente o objeto desejado (Figura 7a), ou quando apenas uma pequena área da planta não está dentro da *bounding box* (Figura 7b).

No entanto, em situações em que a *bounding box* fornecida não abrange completamente a alface, houve ocorrência de partes da planta sem a devida máscara de segmentação (Figura 7c). Além disso, foi possível identificar um problema adicional: nas imagens em que a planta possui um tamanho pequeno, o modelo de segmentação incluiu a região correspondente aos espaços vazios onde a planta é posicionada no sistema hidropônico (Figura 7d).

IV. CONCLUSÃO

Neste estudo, foi desenvolvida uma ferramenta para auxiliar na rotulagem de conjuntos de imagens destinados à segmentação semântica. Foram analisadas 1000 imagens contendo pólipos e cerca de 600 mudas de alface, nos quais ambos os objetos foram previamente delimitados por *bounding boxes*.

No conjunto de dados contendo pólipos, obteve-se um alto IoU e um coeficiente Dice elevado. Já no conjunto de imagens de alfaces, foram extraídos resultados qualitativos satisfatórios. Entretanto, vale-se ressaltar que a qualidade das máscaras geradas está diretamente relacionada à precisão das *bounding boxes* fornecidas.

Portanto, conclui-se que a ferramenta possui o potencial de auxiliar no processo de rotulagem de imagens para a tarefa de segmentação em conjuntos que já possuam caixas delimitadoras anotadas, de forma que anotadores humanos precisariam apenas

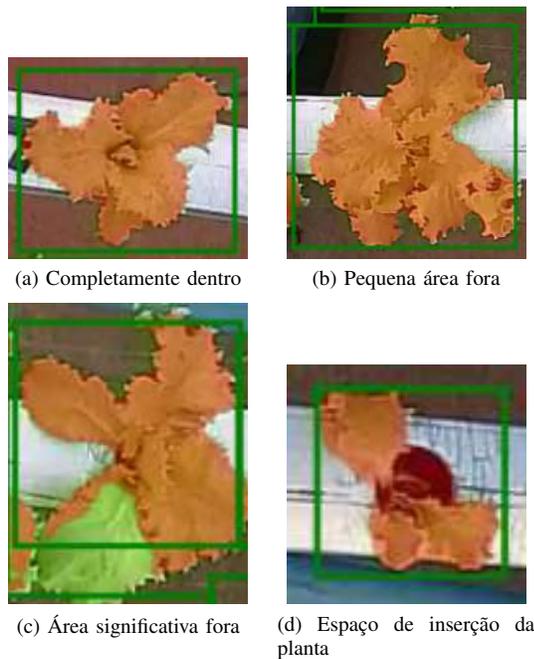


Fig. 7. Comparação entre diferentes cenários de *bounding boxes*

revisar e ajustar as máscaras geradas ao invés de desenhá-las em sua totalidade.

Como trabalho futuro é possível a avaliação da ferramenta desenvolvida em novos conjuntos de imagens, a exploração das demais formas de *prompt* do SAM para a geração das anotações, e também a utilização de algoritmos de detecção de objetos para a geração de *bounding boxes* em conjuntos de imagens que não possuam essa informação disponível, para na sequência serem utilizadas com o SAM.

AGRADECIMENTOS

A Universidade Tecnológica Federal do Paraná (UTFPR) e ao Centro de Excelência em Inteligência Artificial (CEIA) da Universidade Federal de Goiás (UFG).

REFERÊNCIAS

- [1] T. C. d. M. Nascimento, "Segmentação de imagens utilizando elementos de morfologia matemática," Recife, Abril 2013, - Universidade Federal de Pernambuco, Dissertação de Mestrado.
- [2] R. C. Gonzalez and R. E. Woods, "Processamento digital de imagens," vol. 3 ed., 2009.
- [3] M. F. S. Saldanha and C. C. Freitas, "Segmentação de imagens digitais: Uma revisão," in *Anais do 9º Workshop dos Cursos de Computação Aplicada do INPE*. São José dos Campos: INPE, 2009. [Online]. Available: <http://urlib.net/ibi/8JMKD3MGP8W/37NC86S>
- [4] S. Rajalingappaa, *Deep Learning for Computer Vision*. Birmingham, UK: Packt Publishing, Jan. 2018.
- [5] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.
- [6] X. Li and D. Chen, "A survey on deep learning-based panoptic segmentation," *Digital Signal Processing: A Review Journal*, vol. 120, p. 103283, 2022. [Online]. Available: <https://doi.org/10.1016/j.dsp.2021.103283>
- [7] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9396–9405.
- [8] R. P. Nikhil and K. P. Sankar, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, 1993.
- [9] M. Stewart. (2021, Feb.) Five big problems with labeled data. Zumo Labs. [Online]. Available: <https://www.zumulabs.ai/post/five-big-problems-with-labeled-data>
- [10] SAM. (2023) Segment anything. Acesso em 30/08/2023. [Online]. Available: <https://segment-anything.com/>
- [11] M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, and Y. Zhang, "Segment anything model for medical image analysis: An experimental study," *Medical Image Analysis*, vol. 89, p. 102918, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841523001780>
- [12] K. Sofiuk, I. A. Petrov, and A. Konushin, "Reviving iterative training with mask guidance for interactive segmentation," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3141–3145.
- [13] Q. Liu, Z. Xu, G. Bertasius, and M. Niethammer, "Simpleclick: Interactive image segmentation with simple vision transformers," 2023.
- [14] X. Chen, Z. Zhao, Y. Zhang, M. Duan, D. Qi, and H. Zhao, "Focalclick: Towards practical interactive image segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 1300–1309.
- [15] S. Mohapatra, A. Gosai, and G. Schlaug, "SAM vs BET: A comparative study for brain extraction and segmentation of magnetic resonance images using deep learning," Apr. 2023, v1, last revised 19 Apr 2023 (this version, v3). [Online]. Available: <https://arxiv.org/abs/2304.04738>
- [16] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.10062>
- [17] I. Giannakis, A. Bhardwaj, L. Sam, and G. Leontidis, "Deep learning universal crater detection using segment anything model (SAM)," *Cornell University*, Apr. 2023. [Online]. Available: <https://arxiv.org/abs/2304.07764>
- [18] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023.
- [19] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-seg: A segmented polyp dataset," in *MultiMedia Modeling*, Y. M. Ro, W.-H. Cheng, J. Kim, W.-T. Chu, P. Cui, J.-W. Choi, M.-C. Hu, and W. De Neve, Eds. Cham: Springer International Publishing, 2020, pp. 451–462.
- [20] P. Rawla, T. Sunkara, and A. Barsouk, "Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors," *Gastroenterology Review/Przegląd Gastroenterologiczny*, vol. 14, no. 2, pp. 89–103, 2019. [Online]. Available: <http://dx.doi.org/10.5114/pg.2018.81072>
- [21] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021. [Online]. Available: <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660>
- [22] J. Andrade, A. Cordeiro, J. A. Junior, and P. P. Filho, "Reconhecimento de alfabetos em sistema hidropônico utilizando técnicas de processamento digital de imagens," in *Anais do XIX Congresso Latino-Americano de Software Livre e Tecnologias Abertas*. Porto Alegre, RS, Brasil: SBC,



- 2022, pp. 137–140. [Online]. Available: <https://sol.sbc.org.br/index.php/latinoware/article/view/22981>
- [23] C. R. Brzezinski, J. Abati, A. Geller, F. Werner, and C. Zucareli, “Production of iceberg lettuce cultivars under two cropping systems,” *Revista Ceres*, vol. 64, pp. 83–89, 2017.
- [24] F. B. A. D. MEDEIROS, “Produção e qualidade de cultivares de alface americana em função do espaçamento de plantio,” 2015.
- [25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.
- [26] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, “A survey of modern deep learning based object detection models,” *Digital Signal Processing*, vol. 126, p. 103514, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1051200422001312>
- [27] X. Liu, L. Song, S. Liu, and Y. Zhang, “A review of deep-learning-based medical image segmentation methods,” *Sustainability*, vol. 13, no. 3, 2021. [Online]. Available: <https://www.mdpi.com/2071-1050/13/3/1224>