

Comparativo entre modelos de segmentação para identificação de espécies florestais

Rômulo Diógenes Moro

Universidade Tecnológica Federal do Paraná - UTFPR
Medianeira, Brasil
romulodiogenesmoro@alunos.utfpr.edu.br

Pedro Luiz de Paula Filho

Universidade Tecnológica Federal do Paraná - UTFPR
Medianeira, Brasil
pedrol@utfpr.edu.br

Abstract—Deforestation is a growing concern, as its impacts are directly related to global warming. One way to reduce forest degradation is to correctly identify wood species. However, this identification becomes difficult due to the large number of species present in the Brazilian biome. This article aims to explore the concepts for developing a solution based on image segmentation techniques such as You Only Look Once (YOLO) and Segment Anything Model (SAM) to optimize the performance of a Resnet-50 classifier. Upon completion of this solution, we hope to contribute to the accurate identification of wood species, assisting law enforcement agents in combating illegal deforestation.

Keywords—YOLO; SAM; Resnet-50; Forest Species.

Resumo—O desmatamento das florestas é uma preocupação crescente, pois seus impactos estão diretamente relacionados ao aquecimento global. Uma das maneiras de diminuir a degradação das florestas é a correta identificação das espécies de madeira. Porém, essa identificação se torna difícil devida à grande quantidade de espécies presentes no bioma brasileiro. Este artigo tem como objetivo explorar os conceitos para o desenvolvimento de uma solução baseada em técnicas de segmentação de imagens como o *You Only Look Once* (YOLO) e o *Segment Anything Model* (SAM) para otimizar o desempenho de um classificador Resnet-50. Ao concluir esta solução, espera-se contribuir para a identificação precisa das espécies de madeira, auxiliando com os agentes fiscalizadores no combate ao desmatamento ilegal.

Palavras-chave—YOLO; SAM; Resnet-50; Espécies Florestais.

I. INTRODUÇÃO

A preocupação com o aquecimento global tem crescido nos últimos anos, sendo a emissão de gases do efeito estufa uma das principais causas [1]. Entre os importantes agentes que contribuem para a diminuição do aquecimento global, estão as florestas [2]. As florestas desempenham um papel importante para o equilíbrio do ecossistema da terra, pois é nas florestas que encontramos uma grande quantidade de plantas e animais. Dentre as várias contribuições das florestas, pode ser destacado a capacidade de captação de gases que causam o efeito estufa, como o dióxido de carbono. A captação desses gases auxilia para a diminuição do efeito estufa, reduzindo o aquecimento global [3].

Entre as inúmeras florestas, a Amazônica se destaca por sua biodiversidade, porém, vem sofrendo degradação, entre os principais motivos, está a pecuária, e a exploração ilegal da madeira [4]–[6]. A madeira contribuiu para o desenvolvimento humano ao longo da história [7]. O reconhecimento de espécies através da madeira tem impacto em diversas áreas, como comércio, construção civil, indústria, infraestrutura, entre outros. A correta classificação das espécies é importante para o manejo florestal adequado [8].

Uma das medidas que pode contribuir para a diminuição do desmatamento ilegal é uma maior fiscalização. Entretanto, os agentes fiscalizadores encontram dificuldades para realizar o processo de identificação da madeira, pois à uma grande quantidade de espécies presentes nos biomas brasileiro [9]. Visando criar soluções que auxiliem na identificação de espécies florestais, diversas soluções foram propostas, utilizando técnicas como *Support Vector Machine* (SVM), Redes Neurais Artificiais (RNA) e Redes Neurais Convolucionais (CNN) [7], [9], [10].

Entretanto essas soluções utilizam a imagem da madeira sem a presença de diversos objetos, como rochas, solo, folhas, partes do corpo humano, entre outras. A presença desses objetos pode prejudicar o desempenho das soluções, diante disso, surge a demanda por alternativas que minimizem o impacto desses objetos. Entre essas alternativas, está uma solução que utilize imagens que contenham apenas a madeira, uma abordagem que segmente a madeira da imagem. Este estudo busca desenvolver uma solução baseada na segmentação de imagens para correta identificação de espécies florestais.

II. METODOLOGIA

Para melhorar a organização e demonstrar de forma clara o plano deste trabalho, foi desenvolvido o fluxograma presente na Figura 1 com as etapas a serem realizadas.

A primeira etapa foi a seleção de um conjunto de dados, este está sendo desenvolvido pelo Grupo de inteligência Computacional (GIC) da Universidade Tecnológica Federal do Paraná

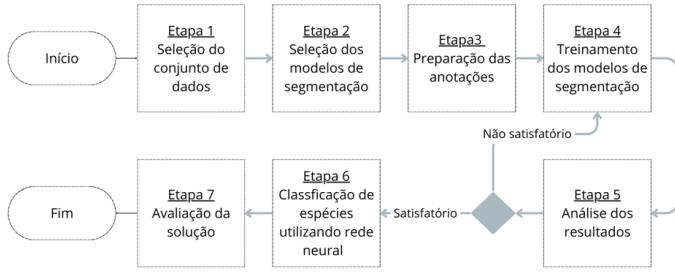
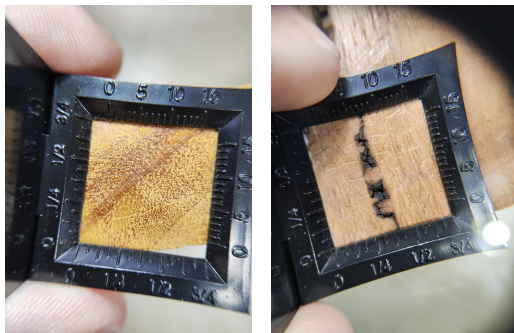


Fig. 1. Fluxograma da organização geral do projeto. Autoria própria (2024).

(UTFPR), campus Medianeira. O conjunto de dados contém imagens de espécies do bioma Amazônico, as amostras foram fornecidas pela Polícia Federal (PF) do Amazonas. As fotografias foram tiradas de diferentes celulares, com iluminação controlada. Conforme a Figura 2a, é apresentada uma imagem da espécie peroba, enquanto a Figura 2b mostra uma imagem da espécie castanheira.



(a) Peroba (b) Castanheira
Fig. 2. Tipos de máscara. Autoria própria (2024).

A Tabela I demonstra informações como nome vulgar e nome científico das espécies florestais, também estão presentes na tabela a quantidade de imagens de cada espécie.

A segunda etapa consiste em escolher os modelos para realizar a tarefa de segmentação de imagens. O primeiro modelo é o YOLO em sua versão 8, a sua escolha se deu devido a sua alta precisão, facilidade de uso e sua capacidade de realizar tarefas de segmentação [11]. O YOLOv8, é um modelo que pode ser utilizado para tarefas de detecção de objetos, segmentação de instâncias e classificação de imagens. O modelo é composto por duas funções principais, são elas: a *backbone* e a *head*.

A função *backbone* extrai características da imagem de entrada, fornecendo uma representação das informações contidas na imagem. A função *head* recebe as características geradas pela função *backbone* e as processa para gerar uma saída para o modelo [12].

TABELA I
SÍNTESE DE INFORMAÇÕES DOS CONJUNTOS DE DADOS USADOS NO TRABALHO

Número	Nome vulgar	Nome científico	Quantidade
01	Açoita cavalo	<i>Luehea divaricata</i>	50
02	Astronium levantei	<i>Astronium lecointei</i>	50
03	Castanheira	<i>Bertholletia excelsa</i>	50
04	Sucupira-preta	<i>Bowdichia nitida</i>	50
05	Tanambuca	<i>Buchenavia oxycarpa</i>	50
06	Piquiá	<i>Caryocar villosum</i>	50
07	Tauari	<i>Couratari tauari</i>	50
08	Cumaru	<i>Dipteryx odorata</i>	50
09	Angelim-pedra	<i>Dinizia excelsa</i>	50
10	Baru	<i>Dipterix</i>	50
11	Ducke	<i>Manilkara huberi</i>	50
12	Baru	<i>Eperva oleifera</i>	50
13	Garapeira	<i>Apuleia leiocarpa</i>	50
14	Maracatiara	<i>Astronium lecointei</i>	50
15	Peroba	<i>Aspidosperma polyneuron</i>	50
16	Roxinho	<i>Peltogyne maranhensis</i>	50
17	Sextonia rubra	<i>Sextonia rubra</i>	50
18	Tajuvá/Taiúva	<i>Maclura tinctoria</i>	50
19	Timbuíla	<i>Enterolobium ontortisiliquum</i>	50
20	Vochysia	<i>Vochysia</i>	50
Total			1000

Autoria própria (2024).

O segundo modelo escolhido foi SAM, pois é um modelo capaz realizar a segmentação de imagens de forma rápida, como também a faz sem a necessidade de treinamento prévio nas imagens. é um modelo de segmentação de imagem *zero-shot*, ou seja, é capaz de realizar a segmentação com precisão para qualquer objeto e em qualquer imagem, incluindo imagens que não encontrou durante o seu treinamento. Diferente dos conceitos de segmentação panóptica, instância e semântica que associam uma classe as áreas segmentadas, o SAM utiliza um novo conceito, denominado de segmentação baseada em *prompt*, no qual significa que por meio do *prompt* uma máscara é gerada. Os *prompts* de entrada podem ser caixas delimitadoras, cliques, textos, entre outros. Essas entradas dizem ao modelo qual a região deve ser segmentada [13]. A arquitetura do modelo é composta por três componentes principais:

- *Image encoder*: realiza a extração das características da imagem [14];
- *Prompt encoder*: possibilita que diversas tarefas de segmentação sejam executadas passando a entrada correta, essas entradas podem ser caixas delimitadoras, cliques, texto [14];
- *Mask decoder*: utiliza as informações do codificador de imagem e do codificador para gerar as máscaras segmentadas [14].

A terceira etapa consiste em realizar as anotações nas ima-

gens, estas serão realizadas utilizando a ferramenta *RoboFlow*. A técnica de anotação será a anotação por polígonos, onde estes polígonos circundam a região da madeira.

Na quarta etapa, dentre os diversos fatores que podem impactar o desempenho de um modelo, testar todas as opções de ajustes dos hiperparâmetros não é uma tarefa viável. Como não foram encontrados estudos que determinem os melhores parâmetros para a identificação de espécies, são necessários experimentos preliminares. Esses experimentos servirão para encontrar as configurações que melhor se adaptam ao modelo, gerando um melhor desempenho.

Para auxiliar na tarefa de ajuste dos hiperparâmetros, dois algoritmos serão utilizados. O primeiro é *Grid Search*, um algoritmo que busca encontrar os melhores hiperparâmetros testando e avaliando todas as combinações possíveis de um conjunto de hiperparâmetros especificados previamente. O segundo algoritmo é o *Random Search*. Diferente do primeiro, este testa apenas quantidades específicas e limitadas de combinações, esse limite é determinado previamente. Ambos algoritmos estão disponíveis através da biblioteca *Scikit-learn*.

Na quinta etapa, para realizar o cálculo das métricas, os conjuntos de pixels classificados corretamente para classe positiva, são denominados verdadeiros positivos (VP), aqueles atribuídos incorretamente para classe positiva são falsos positivos (FP), não atribuídos a classe positiva, mas que pertencem a ela falsos negativos (FN) e os que foram classificados corretamente como não pertencentes a classe positiva, os verdadeiros negativos (VN).

- **Intersection over Union (IoU):** é o resultado da divisão da área de intersecção pela área da união entre a caixa delimitadora predita pelo modelo e uma caixa anotada. A Equação 1 demonstra o cálculo, o resultado varia entre 0 e 1 [15].

$$IoU = \frac{|Predita \cap Anotada|}{|Predita \cup Anotada|} \quad (1)$$

- **Precision:** é a porcentagem de *pixels* classificados corretamente, o resultado varia entre 0 e 1. A Equação 2 demonstra o cálculo [16].

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

- **Recall:** é o número de *pixels* positivos que foram classificados corretamente em relação a todos os *pixels* da classe positiva. O resultado varia entre 0 e 1, conforme o cálculo da Equação 3 [16].

$$Recall = \frac{VP}{VP + FN} \quad (3)$$

- **F1-score:** é a média harmônica entre *precision* e o *recall* conforme a Equação 4. O resultado varia entre 0 e 1 [16].

$$F1 = \frac{2Precision * Recall}{Precision + Recall} \quad (4)$$

- **Dice Coefficient:** mede a similaridade entre as áreas predita e anotada, conforme a Equação 5. O resultado varia entre 0 e 1 [16].

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (5)$$

A sexta etapa, após os treinamentos dos modelos apresentarem resultados satisfatórios, o próximo passo é verificar se o processo de segmentação tem impacto positivo na classificação das espécies. Para esta tarefa, será utilizado o modelo de classificação Resnet-50, este modelo apresenta bom desempenho, baixo consumo de memória e é um modelo adequado para um treinamento rápido. A escolha baseou-se nos resultados obtidos por [17]. Para uma base de dados com imagens microscópicas a acurácia foi de 99,43%. Para a base de dados com imagens macroscópicas, a acurácia atingiu 99,12% após quatro épocas de treinamento. É uma rede convolucional utilizada para visão computacional, é um modelo que venceu o desafio *ImageNet* em 2015. Essa rede permite treinar com profundidade e eficiência em problemas de classificação de imagens [18].

Sua abordagem permitiu superar problemas de gradiente evanescente em redes neurais, ao se tornarem mais profundas, a precisão da rede satura e se degrada rapidamente. Por meio de conexões residuais, a propagação entre conexões evita que falhe por esvanecimento [19].

A arquitetura da *Resnet-50* permite que a função de ativação *ReLU*, continue processando valores resultantes da camada anterior, bem como continue recebendo os valores de entrada das funções [19].

Na sétima etapa será avaliada toda a solução desenvolvida, considerando a qualidade e representatividade das espécies que compõem o conjunto de dados. Para os modelos de segmentação, são avaliados a precisão, a eficiência computacional e a capacidade de lidar com diferentes condições de iluminação, fundos, objetos, entre outros. Avaliar se a predição das espécies utilizando modelo de classificação teve um bom desempenho com imagens segmentadas em relação ao modelo com imagens sem segmentação. Por fim, analisar a aplicabilidade da solução desenvolvida e a eficiência para auxiliar na tarefa de identificação de espécies.

III. CONCLUSÃO

Neste artigo, foram apresentadas algumas características da madeira, além de conceitos de visão computacional, modelos de segmentação, aprendizado de máquina e redes neurais convolucionais. A pesquisa aborda, de forma introdutória um projeto

de Trabalho de Conclusão de Curso (TCC) em andamento. No entanto, experimentos preliminares ainda não foram realizados e estão previstos para as próximas etapas da pesquisa, conforme o cronograma. Entre essas etapas, estão a aplicação prática, a comparação dos modelos e sua avaliação na execução do classificador.

Com resultados satisfatórios em relação à acurácia, o próximo passo será buscar otimizar o modelo, visando sua execução em dispositivos móveis. Com esse ajuste, será possível testar a aplicação em ambientes remotos, sem acesso a internet, por exemplo, aumentando as possibilidades de uso em campo e simulando condições reais de uso. Buscando desenvolver uma ferramenta prática e eficaz que auxilie os agentes fiscalizadores na identificação correta das espécies florestais.

AGRADECIMENTOS

Agradecemos ao CNPq pela bolsa de Produtividade em Desen. Tec. e Extensão Inovadora (DT-II - Processo: 303279/2023-0).

REFERÊNCIAS

- [1] I. Dickmann and L. Liotti, *Educação Ambiental Crítica: Mudanças Climáticas*, 2024, p. 189. [Online]. Available: <https://www.researchgate.net/publication/3803>
- [2] M. do Meio Ambiente, “Prevenção e controle do desmatamento,” 2024. [Online]. Available: <https://www.gov.br/mma/pt-br/assuntos/combate-ao-desmatamento>
- [3] M. N. dos Santos Andreza, C. L. F. Diniz, J. V. da Silva Dantas, N. L. Cartaxo, C. T. de Macedo, V. S. dos Santos, W. J. de Santana, and R. L. Gonçalves, “Interferência do desmatamento na qualidade ambiental e social,” *Brazilian Journal of Development*, vol. 10, pp. 1420–1436, 2024. [Online]. Available: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/66410>
- [4] A. Castro and D. Andrade, “O custo econômico do desmatamento da floresta amazônica brasileira (1988-2014),” *Perspectiva Econômica*, vol. 12, p. 21, 2016. [Online]. Available: https://www.researchgate.net/publication/305313317_O_custo_economico_do_desmatamento_da_Floresta_Amazonica_brasileira_1988-2014
- [5] P. Fearnside, *Desmatamento na Amazônia Brasileira: História, Índices e Consequências*, 2020. [Online]. Available: https://www.researchgate.net/publication/340923558_Desmatamento_na_Amazonia_brasileira_Historia_indices_e_consequencias
- [6] M. B. Diniz, “Desmatamento e ausência de riqueza na amazônia,” 2017, <https://www.gbv.de/dms/zbw/1668938774.pdf>. [Online]. Available: <https://www.econbiz.de/Record/desmatamento-e-aus%C3%AAncia-de-riqueza-na-amaz%C3%B4nia-diniz-marcelo-bentes/10012025755>
- [7] J. Achatz, M. Lukovic, S. Hilt, T. Lädach, and M. Schubert, “Convolutional neural networks for quality and species sorting of roundwood with image and numerical data,” *Expert Systems with Applications*, vol. 246, p. 123117, 2024. [Online]. Available: https://www.researchgate.net/publication/377167038_Convolutional_neural_networks_for_quality_and_species_sorting_of_roundwood_with_image_and_numerical_data
- [8] W. de Oliveira, “Software para reconhecimento de espécies florestais a partir de imagens digitais de madeiras utilizando deep learning,” *Medianeira*, PR, Brasil, p. 103, 2018. [Online]. Available: <https://repositorio.utfpr.edu.br/jspui/handle/1/3189>
- [9] P. L. de Paula Filho, “Reconhecimento de espécies florestais através de imagens macroscópicas,” Ph.D. dissertation, Universidade Federal do Paraná, 2012. [Online]. Available: <https://acervodigital.ufpr.br/handle/1884/29781>
- [10] S.-W. Hwang and J. Sugiyama, “Computer vision-based wood identification and its expansion and contribution potentials in wood science: A review,” *Plant Methods*, vol. 17, 2021. [Online]. Available: https://www.researchgate.net/publication/351137832_Computer_vision-based_wood_identification_and_its_expansion_and_contribution_potentials_in_wood_science_A_review
- [11] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [12] J. Terven, D.-M. C. Esparza, and J.-A. Romero-Gonzalez, “A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, nov 2023. [Online]. Available: <http://dx.doi.org/10.3390/make5040083>
- [13] 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,” *ArXiv*, vol. abs/2304.02643, p. 30, 2023. [Online]. Available: <https://arxiv.org/abs/2304.02643>
- [14] Meta, “Introducing segment anything: Working toward the first foundation model for image segmentation,” 2024. [Online]. Available: <https://ai.meta.com/blog/segment-anything-foundation-model-image-segmentation/>
- [15] 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 Machine Intelligence*, vol. abs/2305.06422, p. 44, 2022. [Online]. Available: <https://www.computer.org/csdl/journal/tp/2022/07/09356353/1rigXK0s5Ak>
- [16] I. R. I. Haque and J. Neubert, “Deep learning approaches to biomedical image segmentation,” *Informatics in Medicine Unlocked*, vol. 18, p. 100297, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S235291481930214X>
- [17] G. Silva, M. K. C. Huang, T. F. Naves, and P. L. de Paula Filho, “Reconhecimento de espécies florestais utilizando visão computacional,” in *Anais do XII Seminário de Extensão e Inovação & XXVII Seminário de Iniciação Científica e Tecnológica da UTFPR*. Santa Helena (PR): UTFPR Santa Helena, 2022. [Online]. Available: <https://www.even3.com.br/anais/seisicite2022/549155-reconhecimento-de-especies-florestais-utilizando-visao-computacional>
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *ArXiv*, vol. abs/1512.03385, p. 12, 2015. [Online]. Available: <https://arxiv.org/abs/1512.03385>
- [19] V. Souza, L. Silva, L. Araujo, and A. Santos, “Análise comparativa de redes neurais convolucionais no reconhecimento de cenas,” 2020. [Online]. Available: https://www.researchgate.net/publication/346593579_Analise_Comparativa_de_Redes_Neurais_Convolutivas_no_Reconhecimento_de_Cenas