

Identificação de Espécies Florestais com YOLO: Um Estudo Baseado em Imagens de Folhas

Isabela Yasmim Osowski*, Eduardo Marcon Gonçalves Fadel*, Larissa Sabbi*,
Agostinho Zanini*, Pedro Luiz de Paula Filho*

*UTFPR - Universidade Tecnológica Federal do Paraná, Medianeira, Paraná, Brasil
osowskiisabela@gmail.com, leduardofadel@gmail.com, larissasabbi@utfpr.edu.br,
agozanini@yahoo.com.br, plpf2004@gmail.com

Abstract—Managing and protecting natural resources is a primary concern for countries. The threat of climate change highlights the importance of forest conservation, and the first step towards preventing more damage is environmental education. The following work aims to create a deep learning model capable of identifying the species of a tree by its leaves. By utilizing an available dataset of 16 different species and a total of 698 images, with the samples taken from a controlled environment, to train the object detection model using YOLO, the goal is to provide a tool that can facilitate and provide extra assurance in identifying species on any location and making it possible to study forest areas that receive less attention from formal organizations and have fewer resources to put towards research. The model presents 93% accuracy with a confidence rate of 60%.

Keywords—species identification; object detection; deep learning.

Resumo—Gerenciar e proteger recursos naturais é uma das principais preocupações de um país, a ameaça da mudança climática traz ênfase para a preservação florestal, o primeiro passo na direção da prevenção é a educação ambiental. O seguinte trabalho procura criar um modelo de aprendizado profundo capaz de identificar a espécie de uma árvore por sua folha, utilizando um dataset disponibilizado com 16 espécies diferentes, totalizando 698 imagens, e amostras em um ambiente controlado para treinar o modelo de detecção de objetos com o software YOLO, procura-se disponibilizar uma ferramenta capaz de facilitar e garantir maior confiança na identificação de uma espécie florestal em qualquer lugar. Possibilitando o estudo ambiental de áreas que recebem menos atenção de organizações formais e possuem menos recursos e mão de obra especializada. O modelo apresenta taxas de acerto acima de 93% com uma confiança de 60% nas predições.

Palavras-chave—reconhecimento de espécie; detecção de objetos; aprendizado profundo.

I. INTRODUÇÃO

A preservação florestal é essencial na proteção dos recursos naturais, seus benefícios vão além de fronteiras delimitadas e auxiliam na manutenção da qualidade de vida no mundo inteiro. No total são 4.06 bilhões de hectares de área florestal, o que configura 31% de toda a superfície de terras emersas. As áreas florestais abrangem um grande número de ecossistemas e apresentam diferentes características entre si, desde a composição

de espécies até as modificações feitas pelo homem. Dito isso, a FRA (*Global Forest Resources Assessments*) identifica duas categorias de floresta, a de regeneração natural e a plantada. Na América do Sul 97% das espécies florestais em áreas plantadas não são nativas de suas regiões [1].

O impacto climático derivado das ações humanas é majoritariamente negativo, afetando primeiro os países em desenvolvimento. O esperado é que 30% de espécies de fauna e flora estejam em risco de extinção em algumas décadas [2]. Assim como a preservação florestal, a educação florestal também é crucial na preparação da sociedade para enfrentar os efeitos do aquecimento global, demandas de energia, degradação do meio ambiente, e mudanças nas condições mundiais sociais, econômicas e sustentáveis [1].

De acordo com Leme Machado [3] a aplicação do princípio de prevenção do direito ambiental começa com a identificação e inventário das espécies de um território. No Brasil existe o Inventário Florestal Nacional (IFN) que é responsável por coletar dados no país todo e assim levantar informações sobre os recursos naturais. Com isso auxilia na elaboração de políticas florestais que são de suma importância na prevenção de danos ambientais. Os dados do IFN são abertos e disponíveis ao público geral, no entanto a maioria dos inventários realizados são sobre áreas restritas.

“A sustentabilidade vai além da manutenção da natureza, deve envolver a comunidade de forma integral” [4], a educação ambiental deve estar presente em todo os níveis de instrução, garantindo assim a cooperação dos indivíduos de uma região com a preservação da natureza. Para esse fim projetos similares, que utilizam a tecnologia de visão computacional para a identificação de espécies como o de Pires *et al.* [5], usa a abordagem de classificação de imagens com modelos de redes neurais convolucionais, e o de da Silva *et al.* [6] que constrói seu modelo em imagens microscópicas de folhas, realizam um papel importante na democratização do conhecimento da área.

Enquanto o reconhecimento é feito por profissionais, engenheiros ambientais, taxonomistas, botânicos e biólogos, ou

em casos de áreas densas com trabalho de campo se procura “mateiros”, nativos da região com conhecimento e familiaridade das espécies locais. No entanto, os especialistas mais qualificados são poucos e mais inacessíveis em pequenos projetos com menor orçamento, existe um desestímulo na formação de taxonomistas em virtude do tempo que se leva para acumular o nível de conhecimento necessário [7].

Com isso, o objetivo do trabalho a seguir é fornecer ferramentas adicionais e proporcionar maior interesse na população sobre a sustentabilidade local. Treinando um modelo de detecção de objetos, uma técnica de visão computacional flexível que pode ser aplicada em diversos contextos. Um exemplo é o reconhecimento facial, que possui trabalhos que exploram diferentes redes para medir a eficiência de cada uma, como é o caso de Jiang e Huaizu [8]. Liu *et al.* [9] também explora a arquitetura Faster R-CNN, com a inclusão de outras, entre elas o YOLOv3, foca na performance de reconhecimento de pequenos objetos.

Para o presente estudo foi escolhido o YOLOv8 [10] para a parte de detecção de objetos, dado a evolução da rede neural do YOLO com o tempo e sua alta performance em diferentes situações. Neste trabalho é realizada a aplicação da técnica de detecção de objetos para reconhecer as folhas de uma imagem e identificar a espécie florestal das mesmas. Conta com amostras da região de Medianeira no Paraná, tem foco nas espécies das árvores locais.

II. METODOLOGIA

O trabalho procura utilizar o método de detecção de objetos como sua abordagem, para isso é necessário anotar as imagens do dataset. Nas anotações foi utilizada a ferramenta do Roboflow [11]. O treinamento do modelo em si foi feito com o YOLO na versão 8, a figura 1 demonstra as etapas seguidas para chegar ao resultado final, seguido de uma explicação mais detalhada.

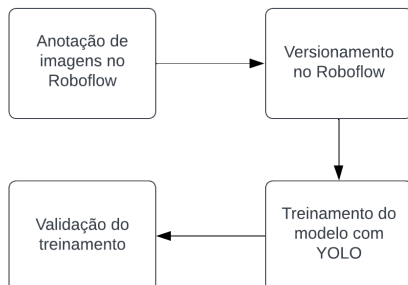


Fig. 1. Fluxograma de etapas do treinamento

A. Dataset

O conjunto de dados foi disponibilizado parcialmente para esse trabalho, contém 16 espécies florestais distintas com 698 imagens no total [5]. As fotos foram tiradas em um ambiente controlado, em cada amostra a folha foi achatada contra um fundo branco por um painel de vidro, apresenta iluminação uniforme tirando a interferência de possíveis sombras e ângulos, como exemplificado na figura 2. A relação do nome popular das espécies com a quantidade de amostras disponíveis para cada uma está na Tabela I.

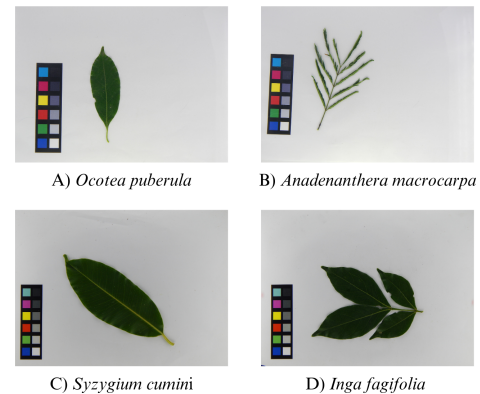


Fig. 2. Exemplo de imagens do dataset

TABELA I
RELAÇÃO DE ESPÉCIES E QUANTIDADE AMOSTRAS

Espécie	Nome Popular	Amostras
<i>Anadenanthera macrocarpa</i>	Angico Vermelho	22
<i>Calliandra sp.</i>	Caliandra Vermelha	23
<i>Ocotea puberula</i>	Canela Guaicá	38
<i>Averrhoa carambola</i>	Carambola	59
<i>Cedrela fissilis</i>	Cedro rosa	57
<i>Erythrina crista-galli</i>	Corticeira	34
<i>Campomanesia xanthocarpa</i>	Guabiroba	75
<i>Schizolobium parahyba</i>	Guapuruvu	32
<i>Hibiscus sp.</i>	Hibisco	68
<i>Inga fagifolia</i>	Inga	22
<i>Handroanthus albus</i>	Ipê-amarelo	19
<i>Tabebuia roseo-alba</i>	Ipê-branco	41
<i>Handroanthus heptaphyllus</i>	Ipê-rosa	40
<i>Syzygium cumini</i>	Jambolão	65
<i>Ceiba speciosa</i>	Paineira	60
<i>Liriodendron tulipifera</i>	Tulipeira	43

Com a variedade de espécies presente no estudo é necessário fazer uma avaliação das características das folhas que são

divididas em dois tipos de acordo com o formato: simples e compostas. As folhas simples como na figura 3A que possui margens lisas, a 3B que possui margem serreada e a folha na imagem 3C que apresenta lobos profundos são aquelas que tem apenas um limbo, já as compostas apresentam sub categorias de acordo com sua disposição [12].

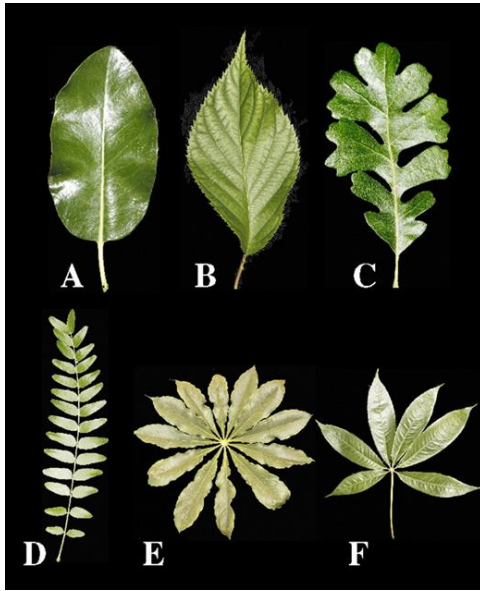


Fig. 3. Exemplo de folhas simples e compostas [12]

As folhas compostas são vistas como uma coleção de estruturas simples chamadas de folíolos, essa categoria é subdividida entre pinadas e palmadas, um exemplo de folha pinada é demonstrada na figura 3D. As folhas ilustradas em 3E e 3F são consideradas palmadas. O trabalho realizado por Mzoughi [13] propõe divisões de nomenclatura para essas folhas compostas além das duas sub categorias principais. Essa abordagem procura facilitar a identificação de particularidades nas folhas em trabalhos de visão computacional como este. A nomenclatura é simples e baseada na quantidade de folículos, tamanho e formato como é exemplificado na figura 4.

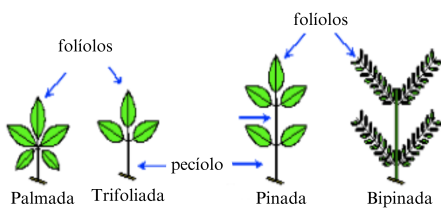


Fig. 4. Nomenclatura das folhas compostas [13]

B. Roboflow

O Roboflow [11] é uma ferramenta web que simplifica o processo de criar um modelo, ele possibilita a criação de um projeto entre 4 categorias: detecção de objetos, classificação, segmentação e identificação de *keypoints*. Neste projeto foi escolhido a detecção de objetos, com essa abordagem é possível identificar os objetos e suas localizações na imagem.

Um projeto no Roboflow guarda as imagens e anotações feitas, cada imagem do conjunto de dados foi registrada manualmente, essa anotação foi feita criando uma *bounding box* em volta de cada folha presente na imagem. Esse processo foi direto na maior parte das amostras pois cada imagem contém apenas uma folha simples, no entanto existe uma complicação com certas espécies como o Angico Vermelho. Espécies como essa possuem uma folha composta bipinada com até 30 pares de folíolos como exemplificado na figura 2B [14].

A principal preocupação nessa decisão foi com o usuário do dia a dia sem conhecimento aprofundado de plantas, pois em folhas compostas de tipo trifoliada é fácil de selecionar apenas um folíolo e confundir ele pela folha inteira, como no caso da Inga (figura 2D) que apresenta folhas paripinadas. A solução aplicada foi de marcar a *bounding box* na folha inteira, para o Angico Vermelho que tem muitos folíolos pequenos se torna inviável marcar cada um individualmente. No caso separado da Inga como é fácil confundir seu folíolo por uma folha completa cada um foi anotado separadamente.

Foram criadas classes para cada espécie de folha presente nas amostras. Entre o grupo de imagens elas foram divididas em 3 tipos, treinamento, teste e validação. Foi aplicada a proporção padrão da ferramenta que define 10% das imagens para teste, 20% para validação e os remanescente 70% para treinamento.

O Roboflow permite ao usuário criar múltiplas versões de um mesmo projeto, foram feitas 3 versões distintas com diferentes quantidades de espécies em cada, sendo a terceira a única contendo todas as amostras disponíveis. Essas diferentes versões foram treinadas com os mesmos parâmetros a fim de verificar se havia alguma diferença na eficiência do modelo. A princípio o número de espécies não gerou alteração significativa nos resultados.

C. YOLO

O YOLO (*You Only Look Once*) é uma série de algoritmos muito utilizados em trabalhos similares a este. Para o treinamento do modelo de detecção de objetos, foi feito um código na linguagem Python no Google Colab¹, importando a chave de usuário e o banco de dados anotado no Roboflow.

O funcionamento geral do YOLOv8 [10], versão utilizada neste trabalho, se resume na separação da imagem de análise

¹<https://colab.research.google.com/>

em uma grade, e para cada elemento da grade ele prediz um conjunto de *bounding boxes* e os classifica. Com isso, ele avalia esses grupos de *bounding boxes* com um algoritmo NMS (*Non-Maximum-Suppression*), que seleciona quais delas têm maior probabilidade de acerto nas áreas de maior sobreposição [15].

O resultado final de uma imagem apresenta uma *bounding box* ao redor de cada objeto, a predição de sua classe, e o grau de confiança do algoritmo em seu resultado. Para referência a figura 5 traz o resultado de uma validação do modelo treinado com a terceira versão do dataset. Ao finalizar um treinamento com o YOLOv8 [10] ele salva dois modelos, o último e o que apresentou melhores resultados.



Fig. 5. Exemplo de predição de um Angico Vermelho

D. Treinamento

Foram realizados três treinamentos distintos, a única diferença sendo o tempo de treinamento, isso é causado por uma alteração dos parâmetros no código. O primeiro treinamento apresentava um número de épocas menor com a intenção de evitar o *overfitting*, quando um modelo é mais flexível do que necessário e possui elementos irrelevantes para realizar seu trabalho, essa condição aumenta o risco de falha na predição de um modelo [16]. O tempo de treinamento também influenciou a decisão referente ao número de épocas; assim, definiu-se inicialmente o limite de 20 épocas para o treinamento do modelo.

O segundo modelo teve 100 épocas de treinamento e foi adicionado um parâmetro de paciência, que indica a quantidade de épocas sem melhorias que ele tolera antes de terminar o processo, que estava em 50. Essa adição foi feita para prevenir um tempo de execução longo sem progresso significativo. Mesmo com essa medida adicional o treinamento foi realizado

sem nenhum problema ou sinal de desestabilização, finalizando em 100 épocas. Por esse motivo, o terceiro modelo foi treinado com 200 épocas e o mesmo parâmetro de paciência, buscando a maior taxa de eficiência que ele pode alcançar.

III. RESULTADOS E DISCUSSÕES

Os modelos foram avaliados primeiramente pela matriz de confusão, tabela que compara as predições corretas e os erros assim avaliando a eficiência, gerada na validação. A princípio a validação foi feita com o padrão de 0.6 de confiança o que resultou em porcentagens de acerto acima de 93%, a confiança indica a porcentagem de certeza que o modelo tem na identificação feita. No terceiro modelo feito com 200 épocas apenas duas espécies de árvore não apresentaram 100% de acerto, o Hibisco e o Jambolão, que mediante maior observação na Tabela II são as únicas espécies que não chegaram a 100% em nenhum modelo.

Na matriz de confusão é indicado que o modelo confunde essas duas espécies entre si, geralmente erra o Jambolão quando a amostra da folha é menor que a média da espécie, no entanto para o Hibisco ele tende a errar indicando que a folha pertence a uma Guabiroba que é mais parecida esteticamente. As três espécies indicadas são as que contêm a maior quantidade de amostras entre todas, para remediar isso seria possível aumentar a quantidade de imagens das outras espécies para alcançar maior uniformidade de dados.

TABELA II
PORCENTAGEM DE ACERTO DA PREDIÇÃO DE CADA MODELO

Nome popular	Modelo 1	Modelo 2	Modelo 3
Angico Vermelho	1	1	1
Caliandra Vermelha	1	1	1
Canela Guaicá	1	1	1
Carambola	1	0.92	1
Cedro rosa	0.8	0.9	1
Corticeira	0.86	1	1
Guabiroba	1	1	1
Guapuruvu	0.86	1	1
Hibisco	0.82	0.94	0.94
Inga	0.95	1	1
Ipê-amarelo	0.75	1	1
Ipê-branco	0.75	1	1
Ipê-rosa	0.5	1	1
Jambolão	0.93	0.93	0.93
Paineira	1	1	1
Tulipeira	1	1	1

Adicionalmente o desempenho do modelo sozinho pode ser avaliado pelo gráfico da curva de precisão e *recall*, a precisão avalia a quantidade de positivos verdadeiros entre todas as

previsões realizadas. O *recall* faz uma proporção entre os positivos verdadeiros e a soma de todos, validando a capacidade do modelo de detectar uma classe.

A figura 6 representa a curva de precisão e *recall* do primeiro modelo treinado, serve de comparação para avaliar a evolução do modelo quando comparado com a figura 7. Que por sua vez apresenta o mesmo gráfico porém seus dados pertencem ao terceiro modelo. Esse resultado era esperado pois há uma diferença de 180 épocas entre ambos os modelos. Enquanto no primeiro temos uma eficiência mais variada entre as espécies, no segundo exemplo temos um atraso inicial apenas no Hibisco, Jambolão e na Guabirola como já foi observado anteriormente através da matriz de confusão.

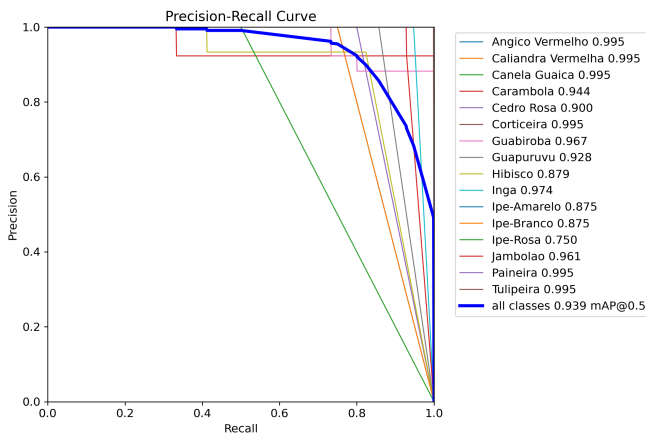


Fig. 6. Gráfico de confiança e *recall* do primeiro modelo

A principal distinção que valida a melhoria do modelo é a linha azul que representa todas as classes. O valor de mAP50 aumenta de 0.939 para 0.987. O mAP50 é a média de precisão calculada considerando as interseções das *bounding boxes* em relação ao IoU (*Intersection over Union*) de 0.5. É feita uma validação se a área de sobreposição entre a caixa prevista e a caixa anotada cobre pelo menos metade da área total combinada. Em suma, o modelo considera que uma previsão está correta se a sobreposição entre a área prevista e a área real for de, no mínimo, 50%.

Na figura 8 temos dois gráficos diferentes do treinamento do terceiro modelo, ambos descrevem a taxa de *box_loss* com o número do treinamento atual. O *box_loss* monitora o quão bem as *bounding boxes* da predição se alinham com o objeto na imagem, mede a precisão geral do modelo. A partir deles é possível afirmar que a cada iteração o modelo teve uma tendência de erro menor, a comparação das taxas do treinamento e da validação de um modelo é importante pois é através delas que se identifica a possibilidade do *overfitting*.

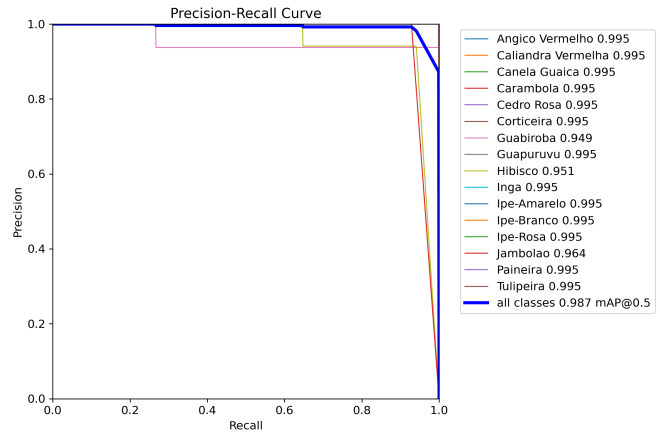


Fig. 7. Gráfico de confiança e *recall* do terceiro modelo

Como foi estabelecido anteriormente, o segundo e terceiro modelo tiveram um parâmetro de paciência para evitar que o modelo passe do ponto adequado e passe a perder performance. No treinamento de 200 épocas é possível ver um indício que o algoritmo estava passando do ponto adequado, ao observar as últimas iterações do modelo. No gráfico de treinamento da figura 8 sua métrica está em queda constante, enquanto a validação tem um leve pico de piora perto do final após a estabilização da *box_loss*. Em casos que a taxa começa a subir novamente na validação enquanto no treinamento está decrescendo qualificam como *overfitting*.

IV. CONCLUSÃO

Em síntese, o estudo buscou desenvolver um modelo de detecção de objetos capaz de detectar diferentes espécies florestais da região por meio de suas folhas. Foi baseado no conjunto de dados utilizado no trabalho de Pires *et al.* [5], não contendo a mesma quantidade total de amostras e espécies. A anotação e o treinamento foram realizados utilizando YOLO e Roboflow respectivamente.

No final do estudo os resultados obtidos satisfizeram as expectativas, demonstrando o grande potencial para um estudo mais aprofundado da detecção de objetos para esse meio. Os modelos treinados com 100 épocas ou mais apresentam porcentagens de acerto acima de 90%, na maioria das ocasiões com a confiança padrão de 0.6 ele acerta 14 das 16 das espécies todas as vezes. Isso é um resultado promissor, porém deve se levar em consideração que todas as imagens utilizadas são em um ambiente controlado, o que inviabiliza de certo modo a aplicação para campo onde se torna necessário um processo para preparar a amostra de forma correta. Uma outra complicação está em algumas espécies que possuem a copa da

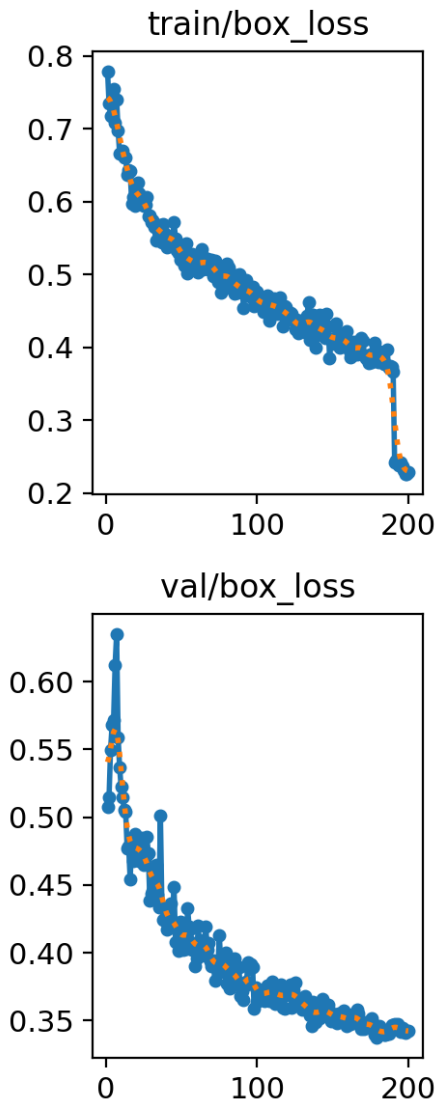


Fig. 8. Gráficos de *box_loss* do treinamento e da validação

árvore muito alta para a coleta das folhas mesmo com o auxílio de ferramentas.

Conclui-se que a aplicação ideal do modelo é para fins educacionais, pois apresenta ser ideal para a identificação de árvores em áreas públicas ou menos densas, como bosques locais com árvores mais novas e de fácil acesso. Devido ao alto desempenho e a grande margem de acerto, é seguro aplicar esse conhecimento para áreas florestais da região oeste do Paraná a fim de realizar pesquisas e aprofundar na educação básica e envolver a comunidade desde cedo, aumentando o interesse e

contribuindo para a preservação ambiental.

Para outros trabalhos que podem utilizar técnicas e abordagens similares é promissor explorar a possibilidade de um conjunto de dados com amostras mais “naturais”. Por exemplo, imagens das folhas soltas com iluminação natural e um fundo de grama, diversas folhas agrupadas no galho e soltas no chão. Também é possível expandir o modelo atual com mais espécies para ter uma área geográfica de maior impacto.

AGRADECIMENTOS

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

REFERÊNCIAS

- [1] Food and Agriculture Organization of the United Nations, “Global forest resources assessment 2020,” 2020. [Online]. Available: <https://doi.org/10.4060/ca9825en>
- [2] R. A. Kerr, “Global warming is changing the world,” *Science*, vol. 316, no. 5822, pp. 188–190, 2007.
- [3] P. A. L. MACHADO and M. A. d. S. ARAGÃO, *Princípios de Direito Ambiental*. Editora Jus Podivm, 2022.
- [4] A. C. P. Mazzei, T. M. F. Floripes, and M. T. M. Feitosa, “A importância da educação ambiental no período operatório concreto para a sustentabilidade existencial,” in *PSICOLOGIA: UM OLHAR DO MUNDO REAL-VOLUME 1*, vol. 1. Editora Científica Digital, 2020, pp. 18–25.
- [5] W. O. Pires, R. C. Fernandes, P. L. de Paula Filho, A. Candido Junior, and J. P. Teixeira, “Leaf-based species recognition using convolutional neural networks,” in *Optimization, Learning Algorithms and Applications: First International Conference, OL2A 2021, Bragança, Portugal, July 19–21, 2021, Revised Selected Papers 1*. Springer, 2021, pp. 367–380.
- [6] N. R. Da Silva, M. W. d. S. Oliveira, H. A. d. A. Filho, L. F. S. Pinheiro, D. R. Rossatto, R. M. Kolb, and O. M. Bruno, “Leaf epidermis images for robust identification of plants,” *Scientific reports*, vol. 6, no. 1, p. 25994, 2016.
- [7] R. C. V. Martins-da Silva, M. G. Hopkins, and I. S. Thompson, “Identificação botânica na amazônia: situação atual e perspectivas.” *Belém, PA: Embrapa Amazônia Oriental*, 2003.
- [8] H. Jiang and E. Learned-Miller, “Face detection with the faster r-cnn,” in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*. IEEE, 2017, pp. 650–657.
- [9] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, “A survey and performance evaluation of deep learning methods for small object detection,” *Expert Systems with Applications*, vol. 172, p. 114602, 2021.
- [10] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics yolov8,” 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [11] B. Dwyer, J. Nelson, T. Hansen, and et al., “Roboflow (version 1.0),” 2024, computer vision software. [Online]. Available: <https://roboflow.com>
- [12] C. Champagne and N. Sinha, “Compound leaves: equal to the sum of their parts?” *Oxford University Press for The Company of Biologists Limited*, 2004.
- [13] O. Mzoughi, I. Yahiaoui, N. Boujemaa, and E. Zagrouba, “Multiple leaflets-based identification approach for compound leaf species.” in *EMR@ ICMR*, 2014, pp. 53–60.
- [14] P. E. R. Carvalho, *Espécies arbóreas brasileiras*. Embrapa, 2008.
- [15] R. Rothe, M. Guillaumin, and L. Van Gool, “Non-maximum suppression for object detection by passing messages between windows,” in *Computer Vision—ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1–5, 2014, Revised Selected Papers, Part I 12*. Springer, 2015, pp. 290–306.
- [16] D. M. Hawkins, “The problem of overfitting,” *Journal of chemical information and computer sciences*, vol. 44, no. 1, pp. 1–12, 2004.