

Mapeamento da diversidade florística a partir de imagens de drone usando poucas anotações

Tomas Ferranti¹, Danilo L. Cardoso², Raul L. Figueiredo³, Dario A. B. Oliveira⁴
EMAp/FGV, Rio de Janeiro, RJ
José Moneva⁵, Matheus Fuza⁶
ESALQ/USP, Piracicaba, SP

Os produtores rurais brasileiros são responsáveis por parte da alimentação da nação e do mundo com *commodities* agrícolas. São mais de 5 milhões de pessoas trabalhadores da zona rural, enfrentando desafios climáticos, econômicos e sociais para garantir o alimento na mesa de todos.

Apesar disso, sem vistoria apropriada eles não são recompensados pela manutenção e restauração da vegetação nativa. A remuneração desses ativos ambientais depende de um rigoroso processo de Monitoramento, Reporte e Verificação (MRV). Os métodos tradicionais de MRV são baseados em medições *in loco* e são onerosos, pois demandam uma logística complexa e levantamento em campo recorrentes. Por este motivo, a pesquisa por alternativas mais factíveis para o mapeamento da diversidade florística de uma região faz-se necessária.

O objetivo do presente trabalho é apresentar um fluxo de dados que faz uso de diversas ferramentas da área de *deep learning* para simplificar a etapa de mapeamento identificando espécies de copas de árvores. Usando imagens aéreas RGB coletadas por drones, dois bancos de dados, supervisionado e não supervisionado, foram criados para esse objetivo. Uma amostra de anotação presente no supervisionado consiste em um polígono que delimita a região da imagem correspondente à copa de uma árvore de uma espécie. O banco supervisionado consiste em 10 espécies de árvores com aproximadamente 30 anotações para cada uma delas. O banco não supervisionado consiste em imagens georreferenciadas que cobrem uma área de 25 hectares com resolução 1.6cm/pixel, abrangendo a região das anotações supervisionadas.

Muitos modelos de *deep learning* necessitam de uma grande quantidade de dados anotados. A solução proposta consiste em enriquecer o banco de dados supervisionado através do banco de dados não supervisionado usando ferramentas como o SAM [2] (modelo de segmentação a partir de um *prompt*) e SwAV [1] (modelo de *deep clustering* baseado em atribuições de protótipos). Após isso, podemos usar um modelo supervisionado como o YOLOv8 [4], que identifica objetos em imagens e segmenta a máscara de pixels correspondente.

O *Segment Anything Model* (SAM) da OpenAI consiste em um *vision foundation model* que permite gerar possibilidades de máscaras individuais *zero-shot* para cada objeto de uma imagem. No nosso caso, ele foi usado para gerar hipóteses de copas de árvores, usando como *prompt* uma malha de pontos 32×32 espalhada uniformemente na imagem.

A partir desse conjunto de possibilidades de máscaras, uma heurística de duas etapas foi criada para escolher apenas aquelas que representassem uma amostra semelhante à alguma das espécies do banco de dados supervisionado. Primeiramente, foram calculadas 5 propriedades geométricas do

¹tomas_ferranti_07@hotmail.com

²danilo.lemos.cardoso@hotmail.com

³raulmonte13@gmail.com

⁴dario.oliveira@fgv.br

⁵joemviveiros@usp.br

⁶matheus.fuza@usp.br

formato de máscara (área, área convexa, extensão, solidez, excentricidade) mais 6 propriedades da região coberta (média e variância de cada canal de cor) para cada registro do banco supervisionado. Com isso, para cada atributo de cada espécie obtemos um intervalo de confiança empírico de (5%, 95%). Uma nova máscara é incluída nesta etapa se ela estiver dentro de 8 ou mais intervalos de pelo menos uma espécie (incluindo obrigatoriamente área normal e convexa). Em seguida, descartamos todas as máscaras que possuísem interseção priorizando manter aquelas com maior probabilidade atribuída pelo SAM de estarem corretas. Este processo pode ser visualizado na Figura 1. No total foram obtidas 88 mil possíveis máscaras de copas de árvores.

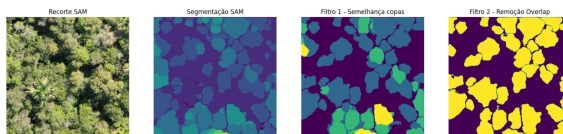


Figura 1: Heurística de duas etapas para filtragem das máscaras geradas pelo SAM.

Com as máscaras do banco de dados supervisionado acrescido das máscaras geradas pela etapa anterior semi-supervisionada, foi empregado um modelo de *deep clustering* chamado SwAV (*Swapping Assignments between Views*). A partir de recortes de imagem centralizados na segmentação de cada copa de árvore, este modelo possibilita a separação desse espaço latente de imagens. Usando protótipos para divisão, *clusters* são criados atribuindo cada imagem a um protótipo, maximizando a semelhança dentro destes. Dessa maneira, copas de árvores novas são rotuladas através da sua atribuição a um protótipo, considerando a espécie majoritária dos dados rotulados deste.

Para a segmentação de instancias supervisionada, podemos usar um modelo chamado YOLO (*You Only Look Once*) [3]. Este é capaz de identificar retângulos delimitadores de objetos em imagens, oferecendo informações espaciais das classificações da rede neural. O modelo em sua versão mais recente também faz a segmentação da máscara do objeto desejado na imagem.

Dentre o planejamento citado, o trabalho encontra-se na etapa de usar o modelo SAM para segmentar as imagens não supervisionadas, gerando as imagens de entrada do SWaV. Após o treinamento do SWaV, será elaborada uma heurística de uso dos resultados de *clustering* para enriquecimento do banco de dados supervisionado. Como resultado final, será realizada a comparação da performance entre o YOLO treinado apenas no banco de dados supervisionado anotado manualmente e treinado acrescido dessas novas amostras com possíveis ruídos.

Referências

- [1] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski e A. Joulin. “Unsupervised learning of visual features by contrasting cluster assignments”. Em: **Advances in neural information processing systems** 33 (2020), pp. 9912–9924.
- [2] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Lo, P. Dollar e R. Girshick. “Segment Anything”. Em: **Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)**. Out. de 2023, pp. 4015–4026.
- [3] J. Redmon, S. K. Divvala, R. B. Girshick e A. Farhadi. “You Only Look Once: Unified, Real-Time Object Detection”. Em: **CoRR** abs/1506.02640 (2015). arXiv: 1506.02640. URL: <http://arxiv.org/abs/1506.02640>.
- [4] Ultralytics. **Site oficial do Modelo YOLOv8**. Online. Acessado em 04/03/2024, <https://yolov8.com/>.