



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Autorizada pelo Decreto Federal nº 77.496 de 27/04/76
Recredenciamento pelo Decreto nº 17.228 de 25/11/2016



PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
COORDENAÇÃO DE INICIAÇÃO CIENTÍFICA

XXVII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS **SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2023**

Transfer Learning com Vision Transformers para Reconhecimento Automático de Espécies de Plantas em Imagens

Luciano Araújo Dourado Filho¹; Rodrigo Tripodi Calumby² e João Batista Rocha Junior³

1. Bolsista PIBIC/CNPq, Graduando em Engenharia de Computação, UEFS, e-mail: lucianoadfilho@ecomp.uefs.br

2. Coordenador do Projeto, Departamento de Ciências Exatas, UEFS, e-mail: rcalumby@uefs.br

3. Orientador do projeto, Departamento de Ciências Exatas, UEFS, e-mail: joao@uefs.br

PALAVRAS-CHAVE: Plantas; Classificação; Vision Transformers; Transfer Learning.

INTRODUÇÃO

Um número crescente de estudos envolvendo reconhecimento de plantas em imagens para aplicações na agricultura indicam o potencial que as tecnologias de computação visual apresentam para aplicações do mundo real (Santos & Gebler 2021). Essas tecnologias estão evoluindo para dar suporte, principalmente por meio da implantação em sistemas embarcados e dispositivos móveis, e auxiliar em diversas tarefas como detecção de doenças em plantas, deficiência de nutrientes, reconhecimento de pragas, etc (Xu & Tian. 2022). Apesar disso, grande parte dos modelos desenvolvidos para realização dessas tarefas têm sido obtidos via transfer learning (transferência de aprendizado) a partir de conjuntos de dados de propósito geral de larga escala, como ImageNet (Russakovsky et al. 2015). Transfer learning (TL) viabiliza o aprendizado de modelos em situações em que a quantidade de dados do domínio para o qual se deseja realizar a transferência é escassa. Isso acontece pois parte do processo de aprendizagem dos modelos é dedicada ao aprendizado de características genéricas utilizadas na composição de padrões mais complexos e que podem ser reaproveitadas por meio de TL. Assim, uma vez que esse processo tenha sido feito, realizá-lo novamente é custoso e até mesmo desnecessário.

Embora seja uma abordagem promissora e até mesmo necessária em muitas situações, podem existir aplicações em que as diferenças entre as características das imagens do domínio de transferência em relação ao domínio alvo sejam significativas a ponto da realização de TL não ser eficaz (Matsoukas et al. 2022). Além disso, para alguns domínios, como o de espécies de plantas, por exemplo, já existem conjuntos de dados de proporções significativamente maiores que o ImageNet (Göeau et al. 2022). Isso significa que, para tarefas relacionadas a este domínio, realizar TL a partir destes datasets específicos pode ser uma abordagem mais eficaz e eficiente. Além disso, o surgimento de modelos chamados *Vision Transformers* (Xu et al. 2022), reconhecidos pela capacidade de aprendizado e eficiência, tem possibilitado um grande avanço para a área de identificação de espécies de plantas. Em (Xu et al. 2022), por exemplo, a utilização da arquitetura ViT-MAE permitiu alcançar o primeiro lugar na competição

PlantCLEF2022 (PC2022) (Göeau et al. 2022), além disso, em (Xu et al. 2022), os autores também puderam demonstrar que a realização de TL a partir do PC2022 apresentou benefícios em relação ao ImageNet.

Diante disso, neste trabalho investigamos como o treinamento de modelos de deep learning sobre um conjunto de dados para identificação de espécies de plantas de larga escala pode potencializar tarefas similares no contexto de aplicações na agricultura. Com isso, objetivou-se a disponibilização de modelos mais eficazes a serem utilizados via transfer learning para desenvolvimento de novas aplicações.

METODOLOGIA

O conjunto PC2022 é composto por dois subconjuntos: web (imagens coletadas da web) e trusted (imagens confiáveis/curadas). Com isso, investigamos o impacto que o treinamento com as imagens possivelmente ruidosas (porção web) pode exercer sobre a qualidade do modelo. Para realizar uma avaliação de alta confiabilidade, optamos por utilizar um único conjunto de validação para os dois experimentos, sendo derivado apenas das imagens confiáveis do dataset. Para apresentar o maior nível de compatibilidade possível entre os resultados, nossos experimentos foram conduzidos conforme as configurações experimentais propostas em (Xu et al. 2022). Duas variações do ViT-MAE foram avaliadas sobre o conjunto de testes do PC2022, sendo a de maior eficácia submetida à uma análise comparativa de desempenho de transfer learning em relação ao ViT-MAE pré-treinado sobre o ImageNet.

Para a avaliação comparativa do transfer learning utilizou-se 4 tarefas com diferentes níveis de complexidade num contexto de plantas (ver Tabela 1). A avaliação dos modelos foi realizada considerando principalmente a acurácia nos conjuntos isolados de teste (quando disponíveis) ou no conjunto de validação e a medida Macro Averaged Mean Reciprocal Rank (MA-MRR) foi utilizada para cômputo da eficácia dos modelos treinados sobre o PC2022.

Tabela 1. Conjuntos de dados utilizados para avaliação do TL dos modelos pré-treinados.

Nome	Qtd. de Imagens	Qtd. Classes	Amostragem	Descrição
Cassava (Sambasivam & Opiyo, 2021)	9.436	5	80-20	Identificação de doenças em folhas de mandioca.
JMuBEN (Jepkoech et al. 2021)	58.555	5	70-20-10	Identificação de doenças em folhas de café Arábica.
Spinach (Ahila et al. 2022)	2.244	25	80-20	Identificação de espécies de espinafre.
CoLeaf (Tuesta-Monteza et al. 2023)	1.006	10	80-20	Identificação de deficiência de nutrientes em folhas de café.

RESULTADOS E DISCUSSÃO

As Tabelas 2 e 3 apresentam os resultados para os experimentos realizados. Conforme a Tabela 2, a utilização dos dados auxiliares provenientes da porção web do PC2022 possibilitou a obtenção de **MA-MRR=0.67514**. Isso indicou melhoria significativa em relação aos resultados obtidos apenas a partir dos dados *trusted*, além de melhoria expressiva em relação àquilo obtido em (Xu et al. 2022). Assim, o modelo *PC2022 Trusted + Web* foi escolhido os experimentos de TL para os outros contexto de dados.

Tabela 2. Resultados dos experimentos de treinamento ViT-MAE (100 épocas).

Dataset	Imagens	Resultado
ImageNet	1.2M	Acurácia = 85.9
<i>PC2022 Trusted</i> (Xu et al. 2022)	2.9M	MA-MRR = 0.62692
<i>PC2022 Trusted</i>	2.9M	MA-MRR = 0.64617
<i>PC2022 Trusted + Web</i>	4M	MA-MRR = 0.67514

A Tabela 3 reporta os resultados dos experimentos de *TL* a partir do PC2022 em comparação a ImageNet. Observou-se que, realizar TL a partir do PC2022 possibilitou uma melhoria significativa em relação ao ImageNet. Outro aspecto decorrente da reutilização do conhecimento proveniente do treinamento a partir do PC2022 diz respeito ao custo de treinamento. O TL a partir do PC2022 possibilitou de modo geral uma melhoria significativa no tempo de convergência dos modelos (e.g., CoLeaf). Em termos práticos, isso significa que além de superior em termos de eficácia, o TL a partir do PC2022 foi realizado com menor demanda de recursos computacionais.

Tabela 3. Resultados dos experimentos de *Transfer Learning* com ViT-MAE sobre os conjuntos relacionados.

Dataset	Pré-treino ImageNet	Pré-Treino PC2022	Número de épocas
Cassava	88.41%	90.52%	50
JMuBEN	100%	100%	20
Spinach	85.38%	99.54%	50
CoLeaf	68.13% (convergência em approx. 130 épocas)	70.58% (convergência em approx. 25 épocas)	150

CONCLUSÃO

Neste trabalho foi realizado um estudo comparativo de desempenho de transfer learning envolvendo eficácia e eficiência de modelos de Vision Transformer pré-treinados. Considerou-se dados de contextos de diferentes similaridades em relação a um conjunto de tarefas alvo de interesse do setor agrícola. Os resultados demonstraram que, em comparação ao uso do conjunto ImageNet, o conjunto PC2022 possibilitou o

treinamento de modelos mais eficazes para reutilização eficiente em tarefas relacionadas ao reconhecimento de plantas. Este trabalho abordou novas possibilidades de experimentos até então não consideradas por outros autores. Para isso, estendeu-se os experimentos realizados em (Xu et al. 2022) por meio do treinamento dos modelos com dados “ruidosos” do conjunto PC2022 até então desconsiderada. Por fim, considerando a eficácia obtida pelos modelos produzidos neste trabalho, códigos, scripts e os parâmetros dos modelos serão compartilhados com a comunidade por meio de repositório on-line¹ juntamente com as instruções apropriadas para reutilização.

REFERÊNCIAS

- RUSSAKOVSKY, O., DENG, J., SU, H., KRAUSE, J., SATHEESH, S., Ma, S., ... & FEI-FEI, L. (2015). Imagenet large scale visual recognition challenge. *IJCV*, 115, 211-252.
- SANTOS, T., & GEBLER, L. (2021). A methodology for detection and localization of fruits in apples orchards from aerial images. In *Anais do XIII SBIAGRO*, (pp. 1-9). Porto Alegre: SBC.
- SAMBASIVAM, G. A. O. G. D., & OPIYO, G. D. (2021). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian informatics journal*, 22(1), 27-34.
- JEPKOECH, J., MUGO, D. M., KENDUYWO, B. K., & Too, E. C. (2021). Arabica coffee leaf images dataset for coffee leaf disease detection and classification. *Data in brief*, 36, 107142.
- GÖEAU, H., BONNET, P., & JOLY, A. (2022). Overview of PlantCLEF 2022: Image-based plant identification at global scale. In *CLEF 2022* (Vol. 3180, No. 153, pp. 1916-1928).
- XU, J., Gu, B., & TIAN, G. (2022). Review of agricultural IoT technology. *Artif. Intell. Agric*, 6, 10-22.
- MATSOUKAS, C., HASLUM, J. F., SORKHEI, M., SODERBERG, M., & SMITH, K. (2022). What makes transfer learning work for medical images: Feature reuse & other factors. In *Proceedings of the IEEE/CVF* (pp. 9225-9234).
- XU, M., YOON, S., JEONG, Y., Lee, J., & PARK, D. S. (2022). Transfer learning with self-supervised vision transformer for large-scale plant identification. In *CLEF (Springer;)* (pp. 2253-2261).
- AHILA PRIYADHARSHINI, R., ARIVAZHAGAN, S., & ARUN, M. (2022, November). A Curated Dataset for Spinach Species Identification. In *International Conference on Computer Vision and Image Processing* (pp. 216-228). Cham: Springer Nature Switzerland.
- TUESTA-MONTEZA, V. A., MEJIA-CABRERA, H. I., & ARCILA-DIAZ, J. (2023). CoLeaf-DB: Peruvian coffee leaf images dataset for coffee leaf nutritional deficiencies detection and classification. *Data in Brief*, 48, 109226.

¹ <https://github.com/rtealumby/plantTL2023>