III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

VeritasAI

Uma Proposta para Detecção de Imagens Artificiais

Leonardo S. Santos
leo.solovijovas@gmail.com
Graduando, BCC
IFSP - SJBV

João Pedro M. Silva
jp.dausi@hotmail.com
Graduando, BCC
IFSP - SJBV

Gabriel M. Alves
gabriel.marcelino@ifsp.edu.br
Docente
IFSP - SJBV

INSTITUTO FEDERAL São Paulo
INSTITUTO FEDERAL Sul de Minas Gerais Campus Poços de Caldas
unesp UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
unifeob.

Introdução

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

- Hiper-realismo Generativo
- Impacto Duplo
- A resposta: Verificação de Autenticidade Visual

Introdução

- Objetivo Geral
  - Desenvolver um sistema de classificação de imagens, baseado em redes neurais convolucionais capaz de identificar com precisão se uma imagem é real ou gerada por IA.

- Objetivos Específicos
  - Reunir e organizar um conjunto balanceado de imagens reais e artificiais;
  - Implementar e validar o modelo CNN com EfficientNetB0 e *fine-tuning*, por meio da Acurácia, F1-Score e Matriz de Confusão.

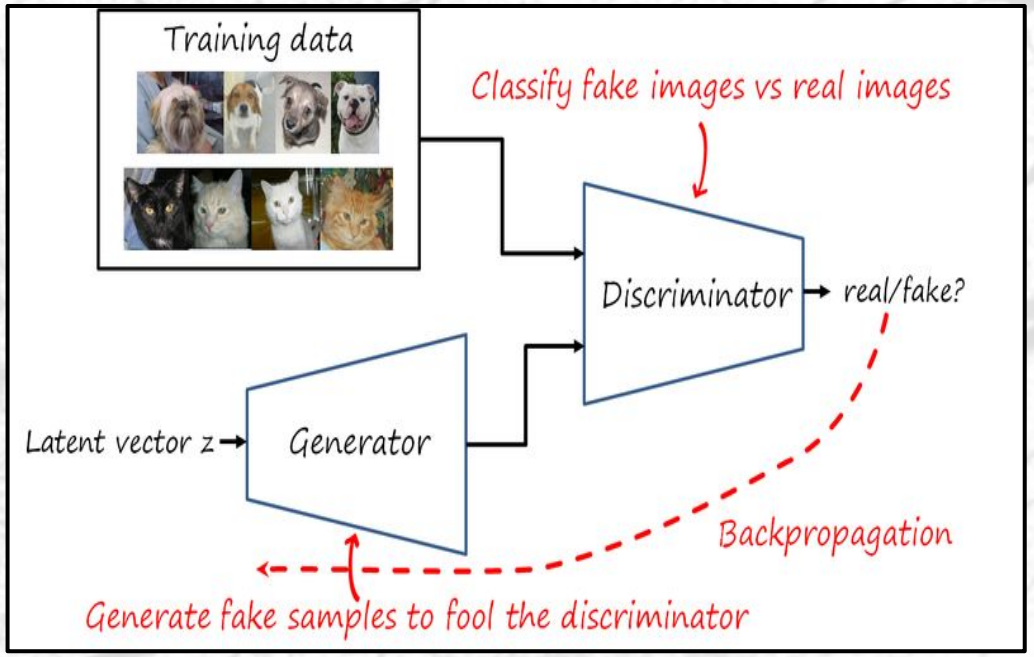
VeritasAI

Fundamentação teórica

State-of-Art

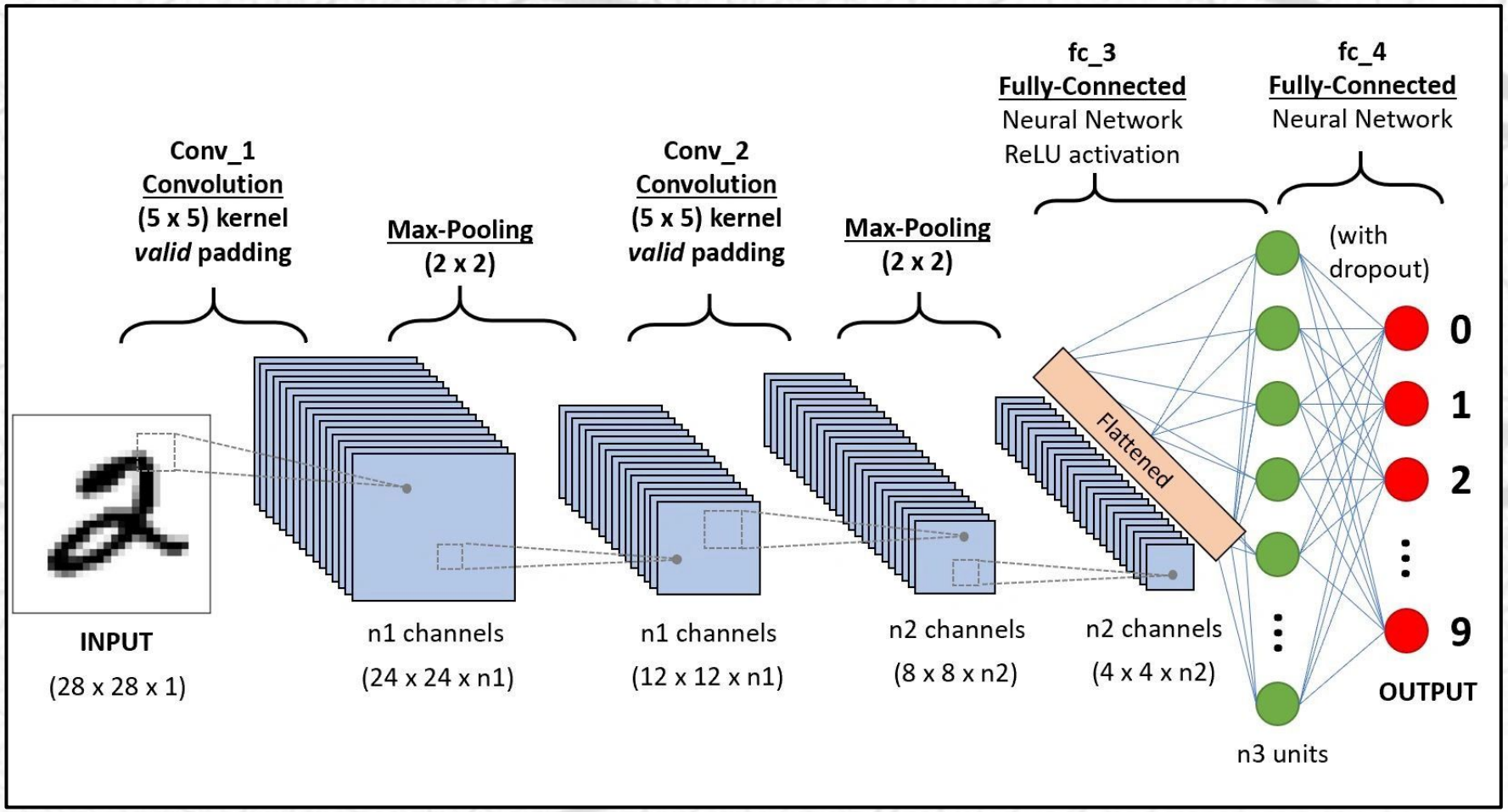
Diffusion Models x GANS

Fonte:https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/



Fonte:https://www.deeplearningbook.com.br/wp-content/uploads/2019/09/gan.png

Convolutional Neural Networks



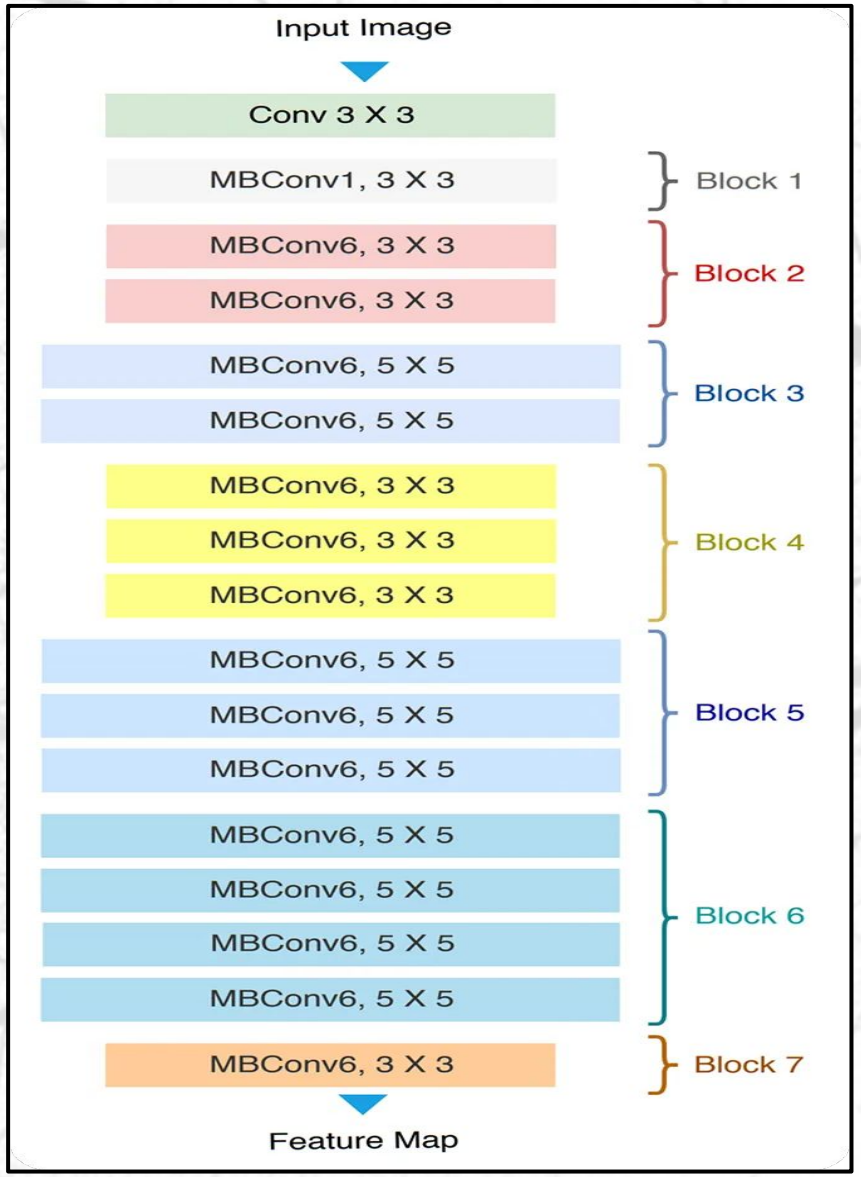
Fonte: https://cdn.analyticsvidhya.com/wp-content/uploads/2020/10/90650dnn2.webp

Fundamentação teórica

EfficientNetB0

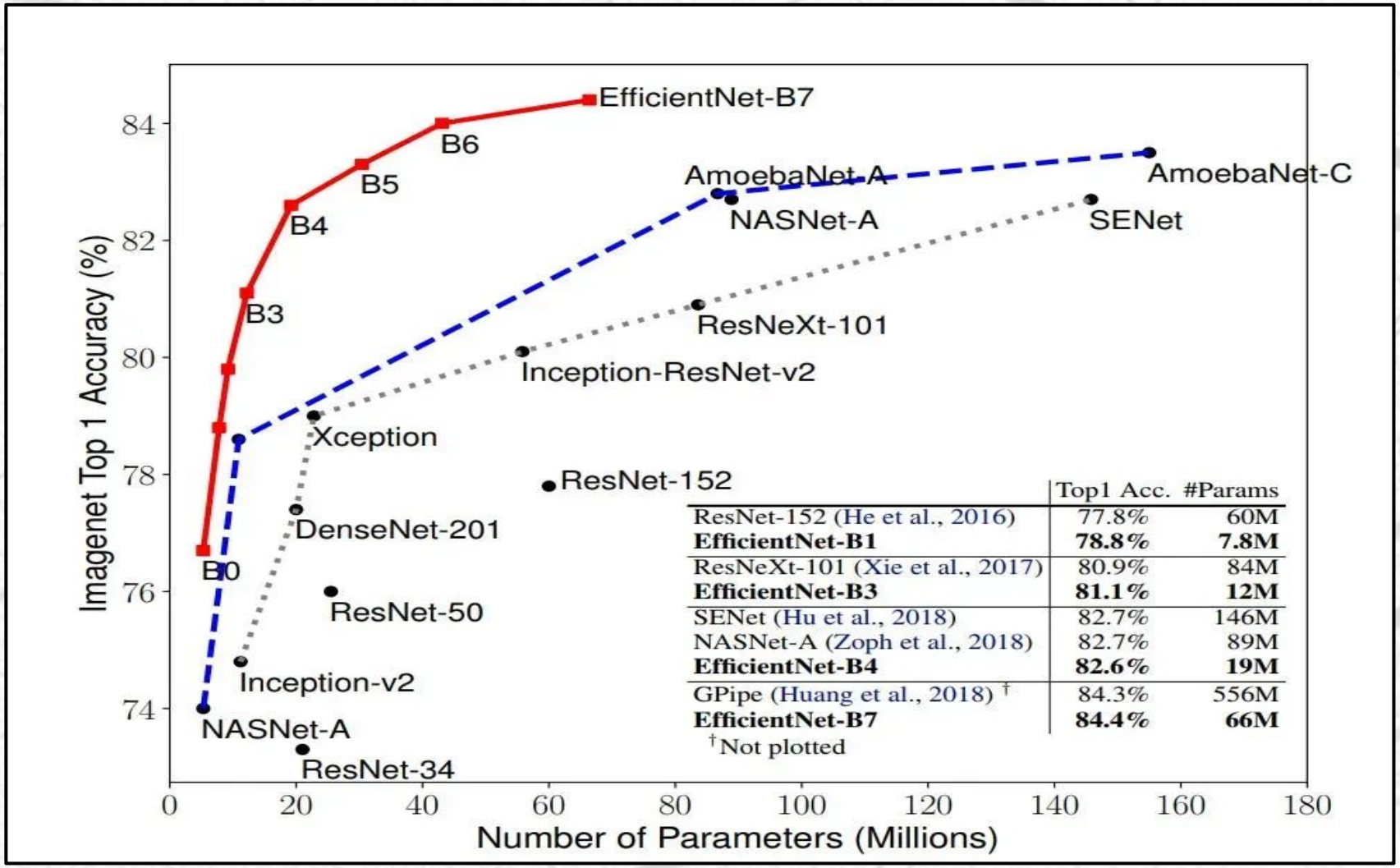
Um Novo Paradigma para Escalar Redes Convolucionais de Forma Eficiente

$$\begin{aligned} \text{depth: } & d = \alpha^\phi \\ \text{width: } & w = \beta^\phi \\ \text{resolution: } & r = \gamma^\phi \\ \text{s.t. } & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ & \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$



Fonte:https://www.ultralytics.com/pt/blog/what-is-efficientnet-a-quick-overview

Fundamentação teórica



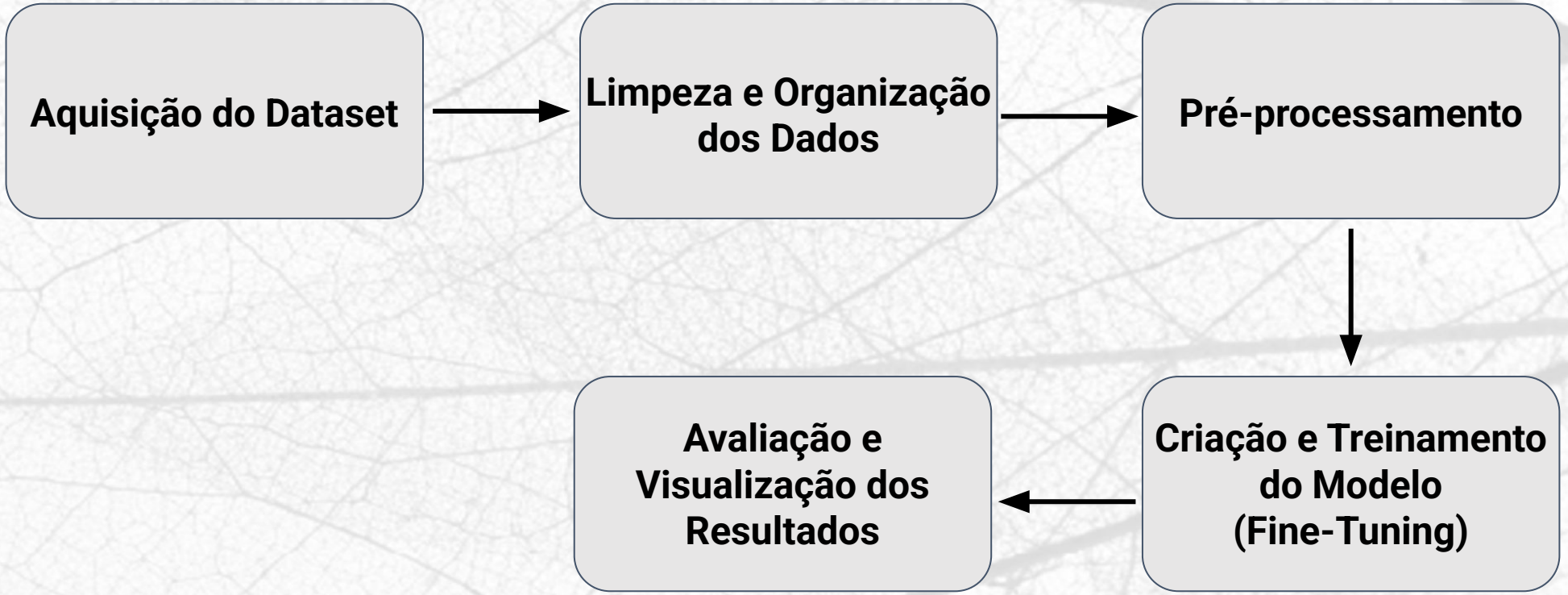
	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**78.8%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.1%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.6%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.4%**	**66M**

†Not plotted

Fonte:https://arxiv.org/abs/1905.11946

Metodologia

Etapas do Projeto



Metodologia

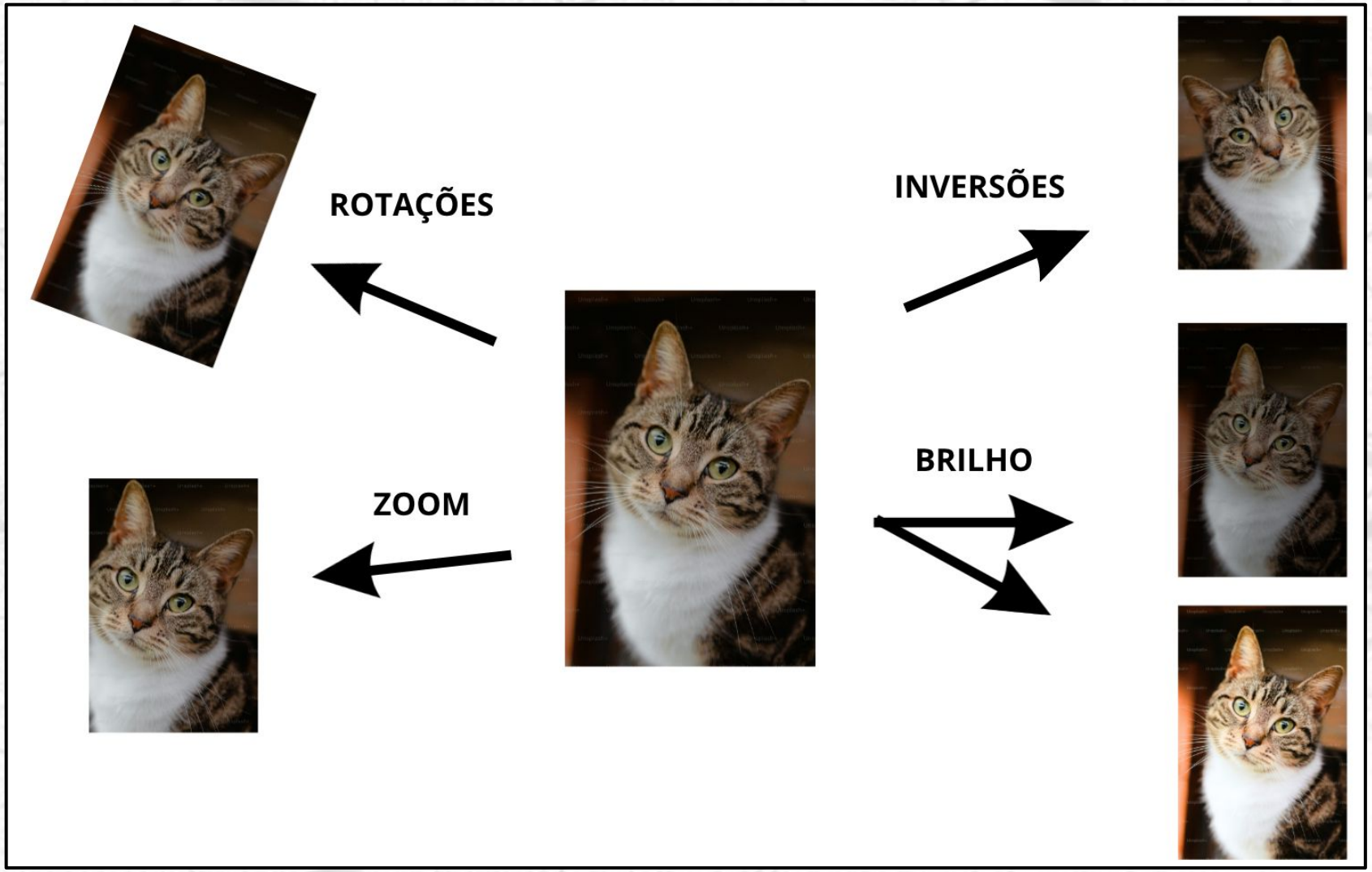
AI vs. Human-Generated Images

- 70k Imagens Balanceadas
- Imagens em Pares
- ⅓ de Imagens de Humanos

Fonte:https://www.kaggle.com/datasets/alessandrasala79/ai-vs-human-generated-dataset

Metodologia

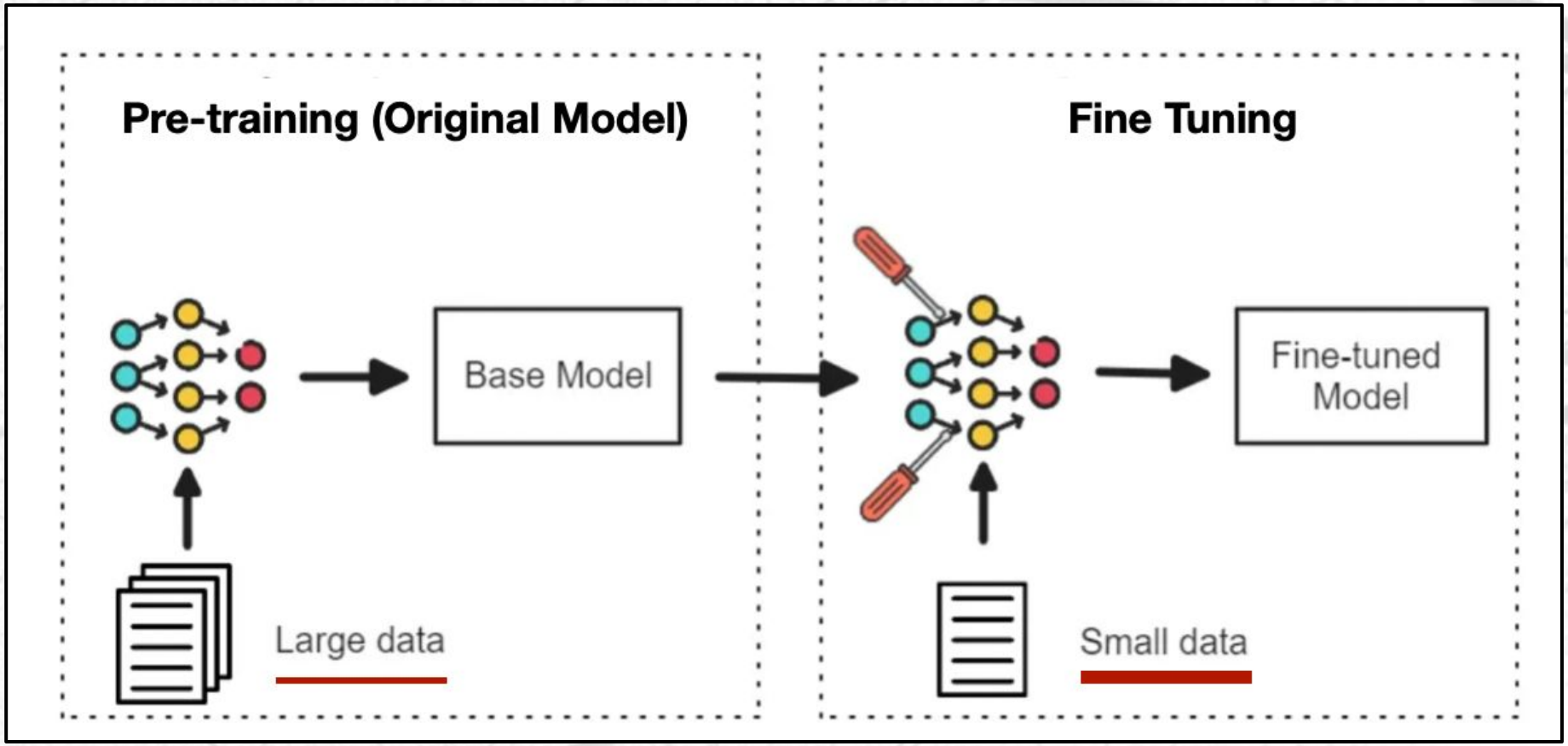
Pré-Processamento e Data Augmentation



Fonte:https://unsplash.com/pt-br

Metodologia

FINE TUNING



Fonte:https://miro.medium.com/1*y9mXfWfxvqHk55TNrP2CXg.png

Metodologia

CNN TREINADA DO ZERO

- Requer grande quantidade de dados.
- Treinamento mais demorado e custoso.
- Não capta padrões complexos.

FINE-TUNING

- Melhor generalização e desempenho.
- Treinamento mais rápido e eficiente.
- Capta padrões complexos desde o início.

Resultados e discussão

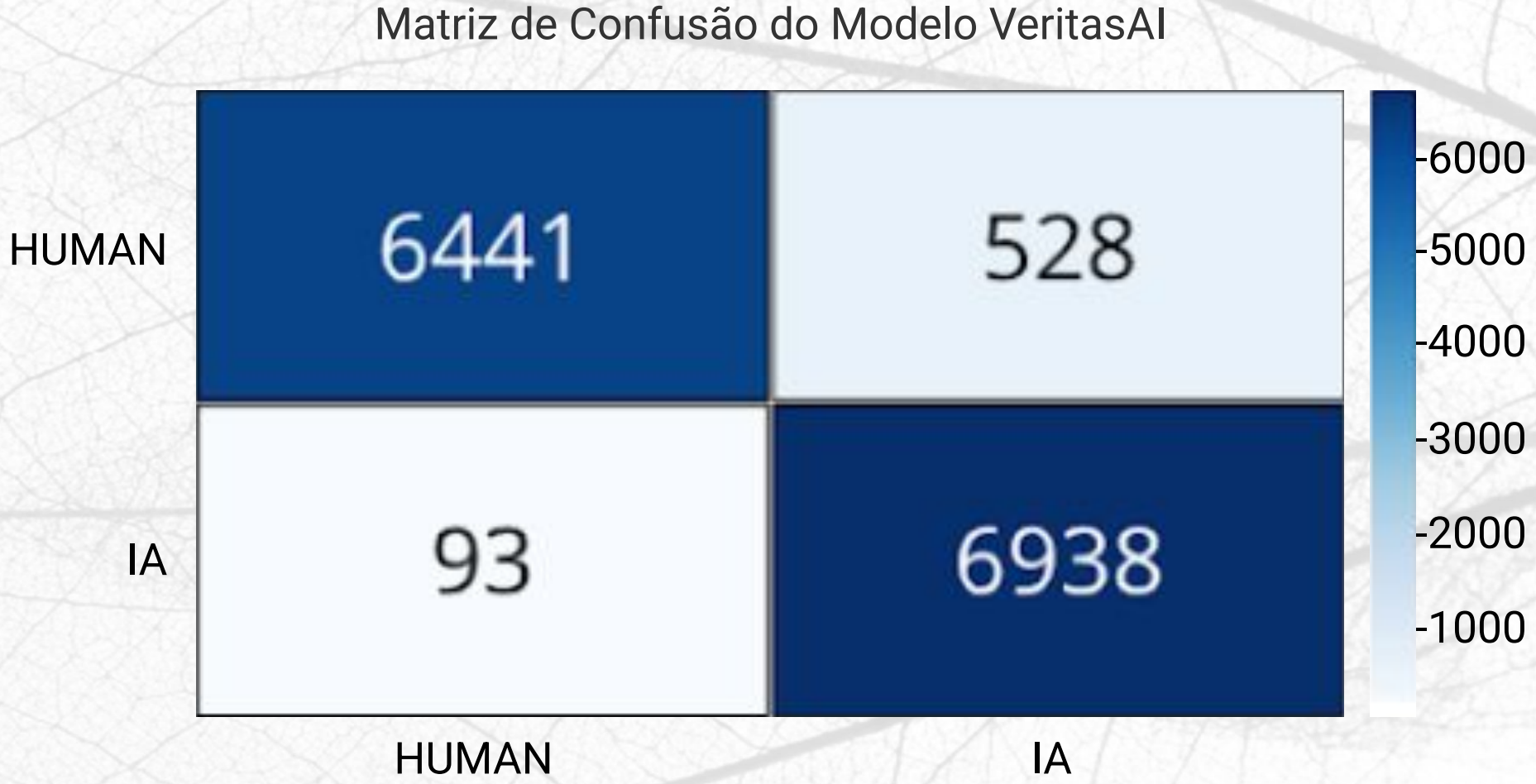
III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

Métricas de Desempenho do Modelo VeritasAI

	PRECISION	RECALL	F1-SCORE	SUPPORT
HUMAN	0.99	0.92	0.95	6969
IA	0.93	0.99	0.96	7031
ACCURACY	-	-	0.96	14000
MACRO AVG	0.96	0.96	0.96	14000
WEIGHTED AVG	0.96	0.96	0.96	14000

Resultados e discussão

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira



Resultados e discussão

- Zona de Incerteza

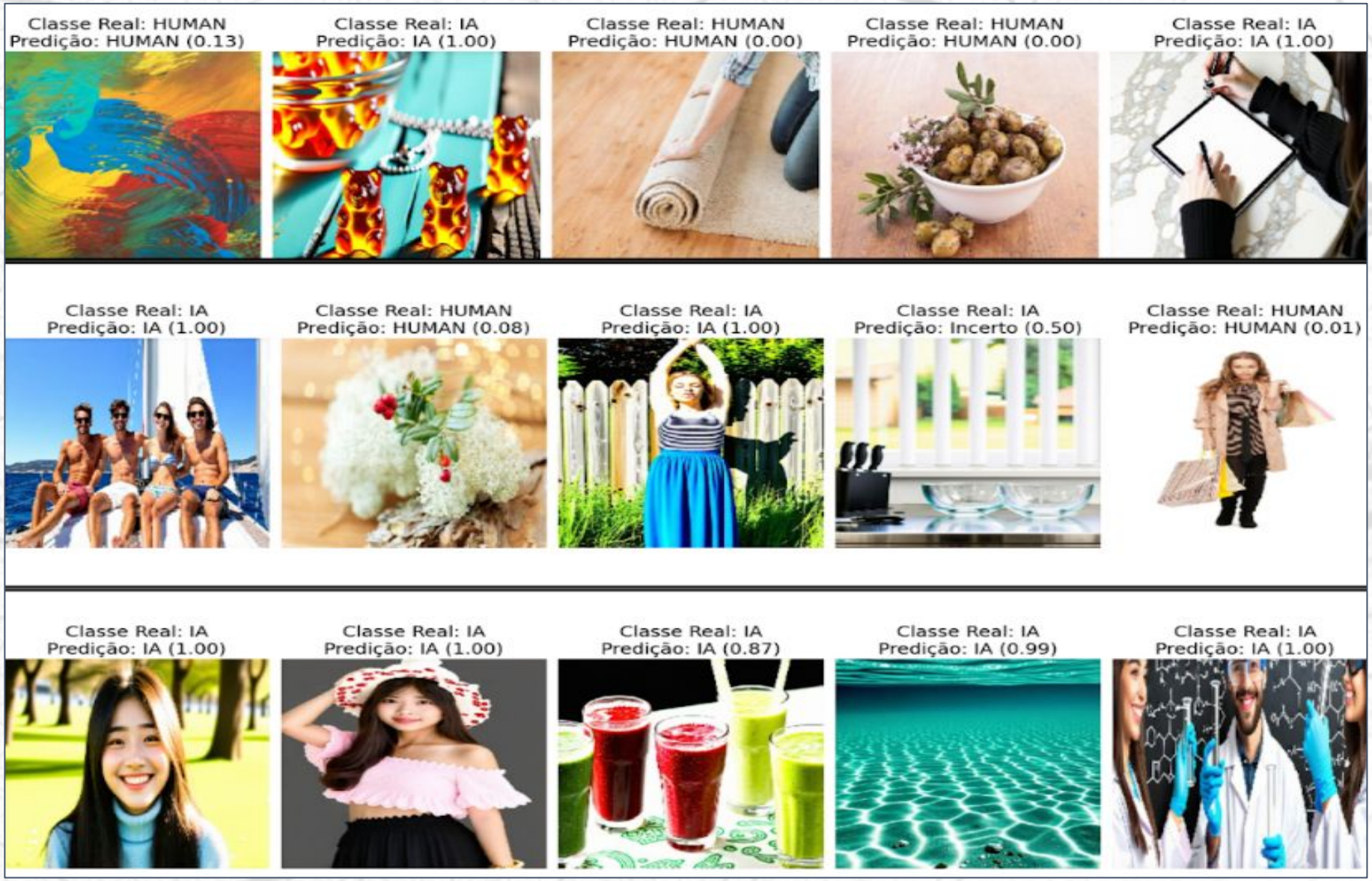
  0.4 < output < 0.6

- Classificação: Incerto

- Sugere Revisão Humana

Resultados e discussão

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira



Considerações finais

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

- Os objetivos do trabalho foram atingidos e o sistema desenvolvido demonstrou capacidade de diferenciar imagens reais de artificiais.
- O uso da EfficientNetB0 permitiu que o modelo generalizasse bem para dados inéditos, confirmando a viabilidade de uma abordagem automatizada para verificação de autenticidade visual.
- A implementação da zona de incerteza aumentou a confiabilidade das previsões.
- Foram observadas limitações em imagens de baixa qualidade ou muito estilizadas o que indica a necessidade de melhorias futuras.
- Trabalhos Futuros
  - Arquiteturas Multimodais
  - Mecanismos de Atenção
  - Ampliar o Dataset

VeritasAI

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

Referências

- ALAM, I.N., Kartowisastro, I.H., Wicaksono, P. (2022). **Transfer learning technique with EfficientNet for facial expression recognition system.** Revue d'Intelligence Artificielle, Vol. 36, No. 4, pp. 543-552. https://doi.org/10.18280/ria.360405
- GOODFELLOW, Ian J.; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WARDE-FARLEY, David; OZAIR, Sherjil; COURVILLE, Aaron; BENGIO, Yoshua. **Generative Adversarial Networks**. arXiv, 2014. Disponível em: https://arxiv.org/abs/1406.2661. Acesso em: 20 out. 2025.
- HOPF, Benedikt; TIMOFTE, Radu. **Practical Manipulation Model for Robust Deepfake Detection**. arXiv, 2025. Disponível em: https://arxiv.org/abs/2506.05119. Acesso em: 20 out. 2025.
- LIU, Jiarui; ZHU, Kaiman; LU, Wei; LUO, Xiangyang; ZHAO, Xianfeng. **A lightweight 3D convolutional neural network for deepfake detection.** International Journal of Imaging Systems and Technology, v. 31, n. 1, p. 200-210, 2021. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/int.22499. Acesso em: 17 ago. 2025.
- NASKAR, Gourab; MOHIUDDIN, Sk; MALAKAR, Samir; CUEVAS, Erik; SARKAR, Ram. **Deepfake detection using deep feature stacking and meta-learning**. Heliyon, v. 10, n. 4, p. e25933, 2024. https://doi.org/10.1016/j.heliyon.2024.e25933

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

Referências

- ROMBACH, Robin; BLATTMANN, Andreas; LORENZ, Dominik; ESSER, Patrick; OMMER, Björn. **High-Resolution Image Synthesis with Latent Diffusion Models.** In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. p. 10684-10695. Disponível em: https://arxiv.org/abs/2112.10752. Acesso em: 17 ago. 2025.
- SALA, Alessandra. **AI vs Human-Generated Images.** 2025. Disponível em: kaggle.com/datasets/alessandrasala79/ai-vs-human-generated-dataset. Acesso em: 17 mai. 2025.
- TAN, Mingxing; LE, Quoc V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.** arXiv, 2019. Disponível em: https://arxiv.org/abs/1905.11946. Acesso em: 22 out. 2025.
- UNESCO. **Synthetic content and its implications for AI policy: a primer.** 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000392181. Acesso em: 17 ago. 2025.

III SECITEM
Simpósio de Educação, Ciência e Tecnologia da Mantiqueira

VeritasAI: Uma Proposta para Detecção de Imagens Artificiais

Obrigado!

Leonardo S. Santos
leo.solovijovas@gmail.com
Graduando, BCC
IFSP - SJBV

João Pedro M. Silva
jp.dausi@hotmail.com
Graduando, BCC
IFSP - SJBV

Gabriel M. Alves
gabriel.marcelino@ifsp.edu.br
Docente
IFSP - SJBV

INSTITUTO FEDERAL São Paulo
INSTITUTO FEDERAL Sul de Minas Gerais Campus Poços de Caldas
unesp UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
unifeob.