

Implementação de um modelo de classificação de áreas irregulares na Floresta Amazônica baseado em Transformers Visuais

Victor Moraes

UFMG

2022

Motivação



Figura:

Garimpo ilegal na Terra Indígena Munduruku, município de Jacareacanga.

Foto: Marizilda Cruppe/Amazônia Real

Revisão bibliográfica

- ▶ Domínio do problema: Sensoriamento remoto
- ▶ Aprendizado de máquina
- ▶ Redes convolucionais
- ▶ Transformers
- ▶ Trabalhos anteriores

Revisão bibliográfica - Sensoriamento remoto

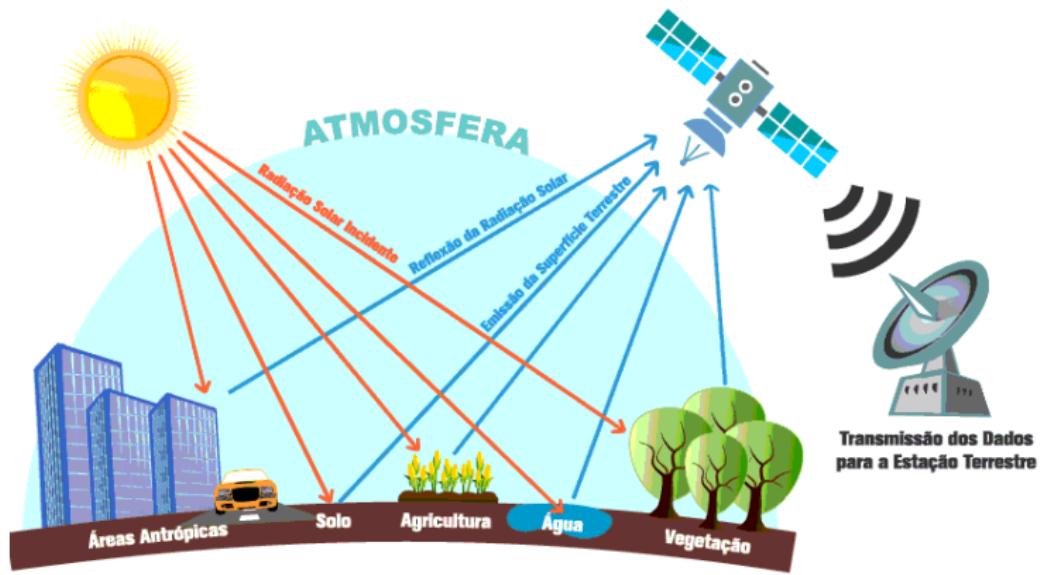


Figura: Sensoriamento remoto Foto:(INSTRUTORGIS, 2022)

Aprendizado de máquina - Definições

“Campo de estudos que visa a dar computadores a habilidade de aprender sem serem explicitamente programados para determinada tarefa.” (SAMUEL, 1959)

“Um algoritmo dito conseguir uma experiência E com respeito a determinada classe de tarefas T e com medidas de desempenho P, se seu desempenho nas tarefas em T, medidas por P, melhoraram a partir da experiência E.” (MITCHELL, 1997)

Aprendizado de máquina - Exemplo

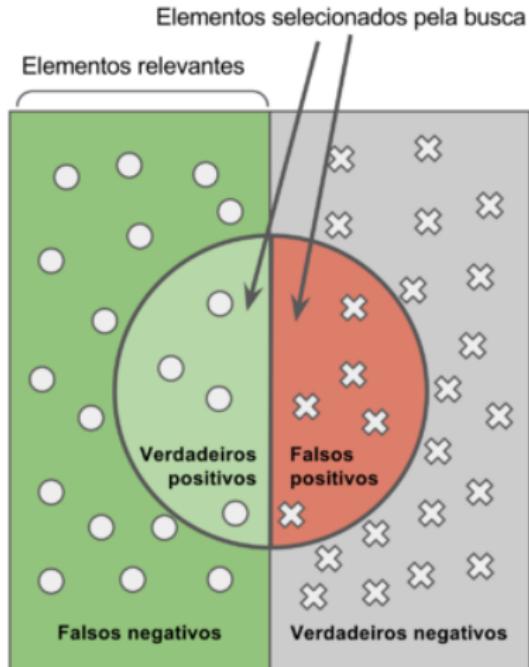
Exemplo:

- ▶ Tarefa T: Problema de classificação
- ▶ Experiência E: Amostras dentre k classes a classificar
- ▶ Desempenho P: acurácia do algoritmo

Producir função:

$$f : \Re^n \rightarrow \{1, \dots, k\}. \text{ Quando } y = f(x)$$

Métricas de desempenho



$$\text{Precisão} = \frac{\text{Verdadeiros positivos}}{\text{Elementos selecionados}}$$

"Quantos elementos
selecionados são
relevantes?"

$$\text{Revocação} = \frac{\text{Elementos relevantes}}{\text{Elementos selecionados}}$$

"Quantos elementos
relevantes foram
selecionados?"

Figura: Precisão e revocação

Métricas de desempenho

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

$$P = \frac{VP}{VP + FN} \quad (1)$$

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

$$AC = \frac{VP + VN}{VP + VN + FN + FP} \quad (3)$$

Figura: Matriz Confusão, Equações de Precisão, Revocação e Acurácia

Métricas de desempenho

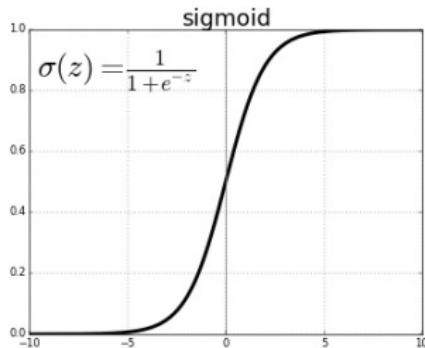
$$F_1 = \frac{P \times R}{P + R} \quad (4)$$

$$F_\beta = (1 + \beta^2) \times \frac{P \times R}{\beta^2 P + R} \quad (5)$$

$$F_2 = 5 \times \frac{P \times R}{5P + R} \quad (6)$$

Figura: Equações de F_β e F_2

Regressão logística



$$\sigma(x) = \frac{1}{1 + \exp(-z(x))} \quad (7)$$

Figura: Função de ativação
Sigmoid.

Otimização

$$w \leftarrow w - \eta \nabla L(w) \quad (8)$$

$$Q(w) = \frac{1}{n} \sum_i Q_i(w) \quad \Rightarrow \quad \nabla Q(w) = \frac{1}{n} \sum_i \nabla Q_i(w) \quad (9)$$

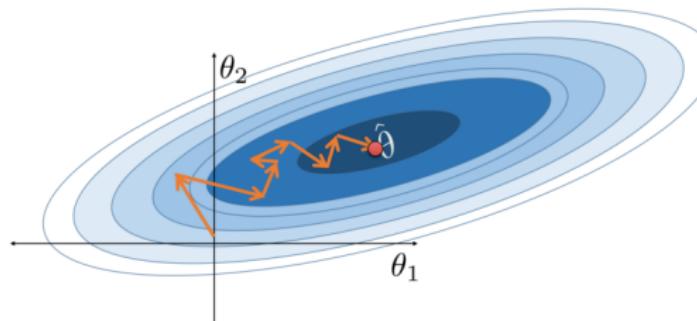


Figura: Ilustração do algoritmo gradiente descendente estocástico.

Função de perda

$$ECB(y_c) = - \sum_{c=0}^M y_c \times \log(\hat{y}_c) \quad (10)$$

$$Focal(y_c) = - \sum_{c=0}^M (1 - w_i \times y_c)^\gamma \times \log(\hat{y}_c) \quad (11)$$

Redes Neurais Artificiais Profundas

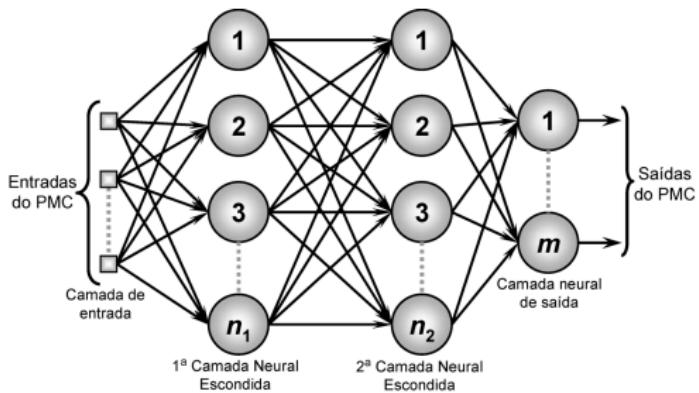


Figura: Exemplo de uma rede de perceptrons de multiplas camadas - MLP.

Redes convolucionais

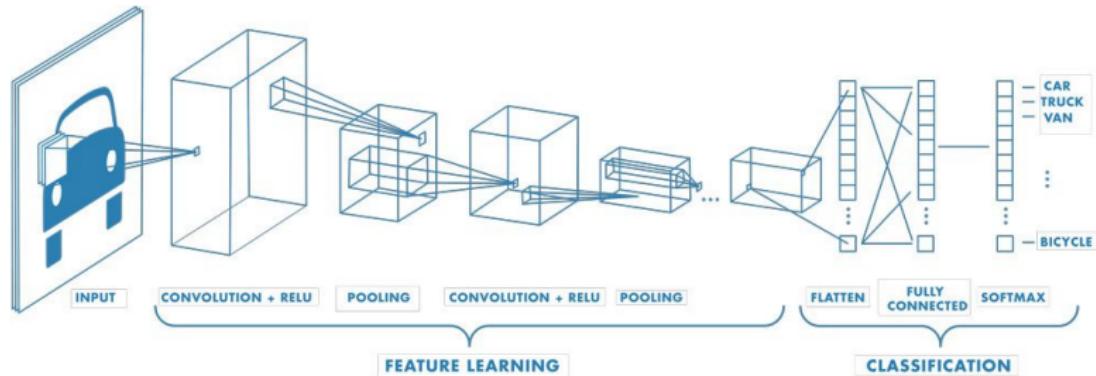


Figura: Arquitetura de uma rede convolucional.

Redes convolucionais

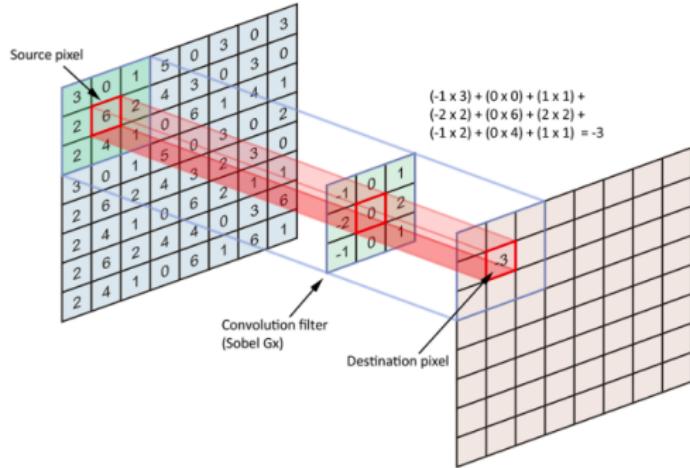


Figura: Filtro convolucional que aplica uma janela deslizante aplicando operação de convolução e pooling.

ResNet

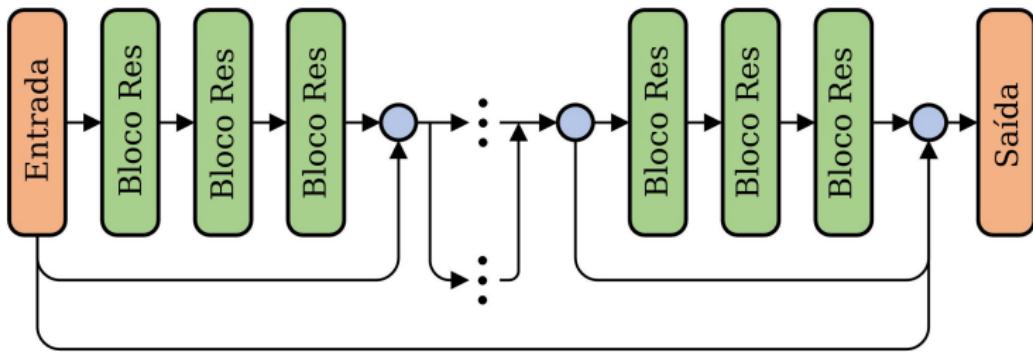


Figura: Arquitetura redes ResNet.

Transformer Visual

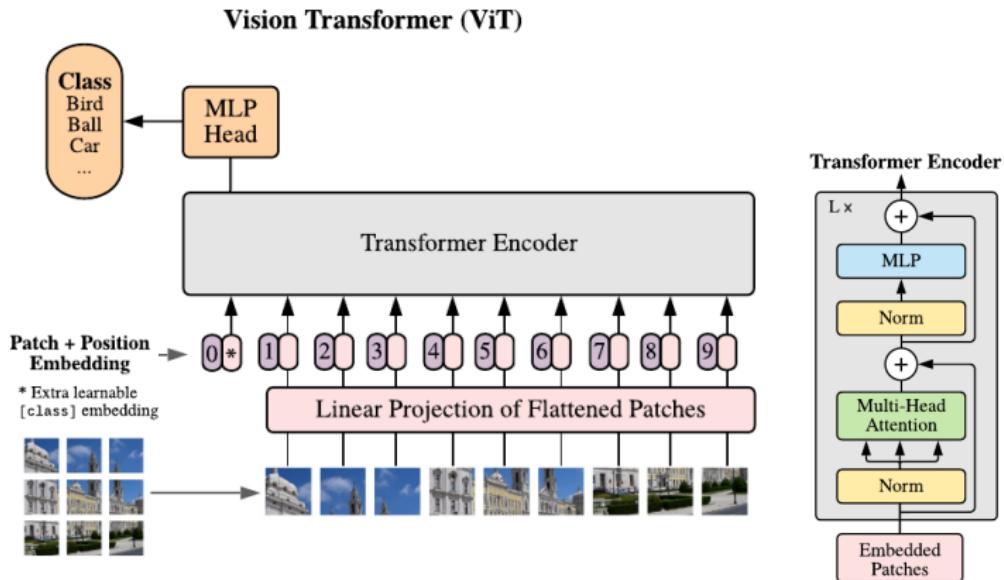


Figura: (DOSOVITSKIY et al., 2020)

Transformer Swin

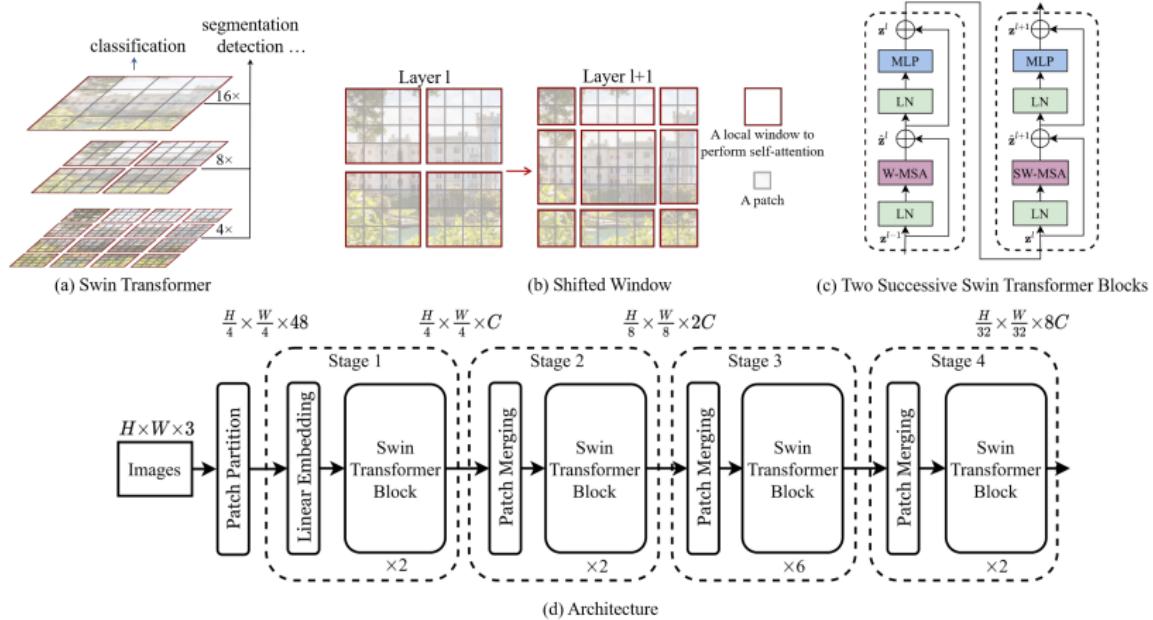


Figura: Arquitetura swin (LIU et al., 2022)

Trabalhos anteriores

- ▶ Em (CAMALAN et al., 2022): Detecção de mudanças para detecção de minas de ouro
- ▶ Em (BALANIUK; ISUPOVA; REECE, 2020): classificação de minas e represas
- ▶ (WANG et al., 2022): Dataset MillionAID e transformers visuais
- ▶ Em (KASELIMI et al., 2022): transformers visuais para detecção de desmatamento

Método



Figura: Amostras de classes do dataset Amazônia do Espaço
Fonte:(PLANETCO, 2016)

Premissas e proposta de solução

- ▶ Explorar viés indutivo dos transformers visuais
- ▶ Comparar desempenho de um modelo Swin e ResNet

Ambiente e ferramentas

- ▶ Jupyter e ambiente cloud Google Collab
- ▶ Pytorch, SciKitLearn e NumPy

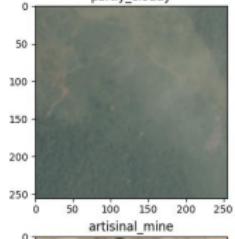
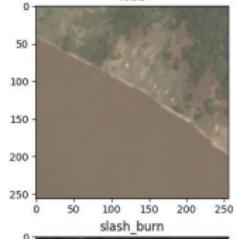
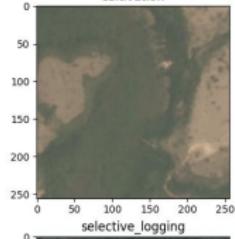
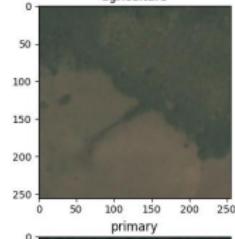
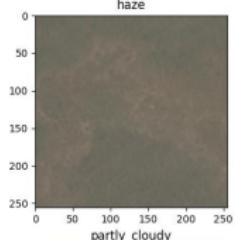
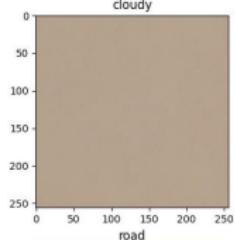
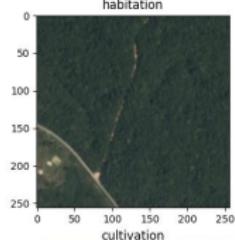
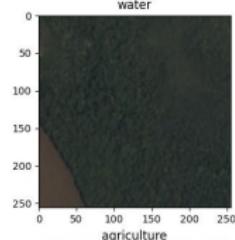
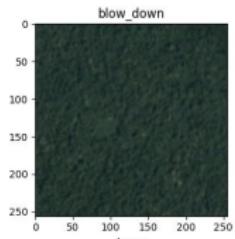
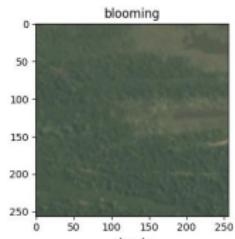
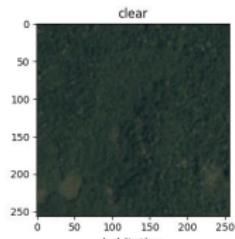
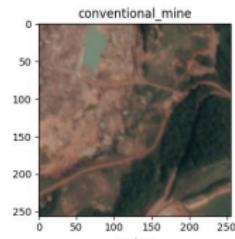
Experimentos

1. Análise exploratória dos dados
2. Disponibilizar pre-processamento do conjunto de dados e de carregadores de amostras.
3. Elaborar um modelo base
4. Utilizar pesos do dos trabalhos disponibilizados para fine-tune.
5. Elaborar modelo proposto baseado em transformers visuais
6. Fixar os hiperparâmetros e metodologia de treino para modelos base e proposto e treiná-los novamente a fim de realizar comparações
7. Realizar análise dos resultados de ambos e comparar desempenho em várias métricas relevantes ao problema

Análise de dados exploratória

Classe	Rótulo	Amostras	Proporção (%)
Mina Convencional	conventional mine	100	0,247
Roça de Ventos	blow down	101	0,250
Queimada	slash burn	209	0,516
Florescimento	blooming	332	0,820
Garimpo	artisinal mine	339	0,837
Desmatamento Seletivo	selective logging	340	0,840
Área Descoberta	bare ground	862	2.129
Nublado	cloudy	2089	5.161
Névoa	haze	2697	6.663
Habitação	habitation	3660	9.042
Cultivação	cultivation	4547	11.233
Parcialmente Nublado	partly cloudy	7261	17.938
Águas	water	7411	18.308
Estrada	road	8071	19.939
Agricultura	agriculture	12315	30.423
Clima Limpo	clear	28431	70.236
Vegetação Primária	primary	37513	92.673

Resultados e discussão



Análise de dados exploratória - Matriz de co-ocorrência

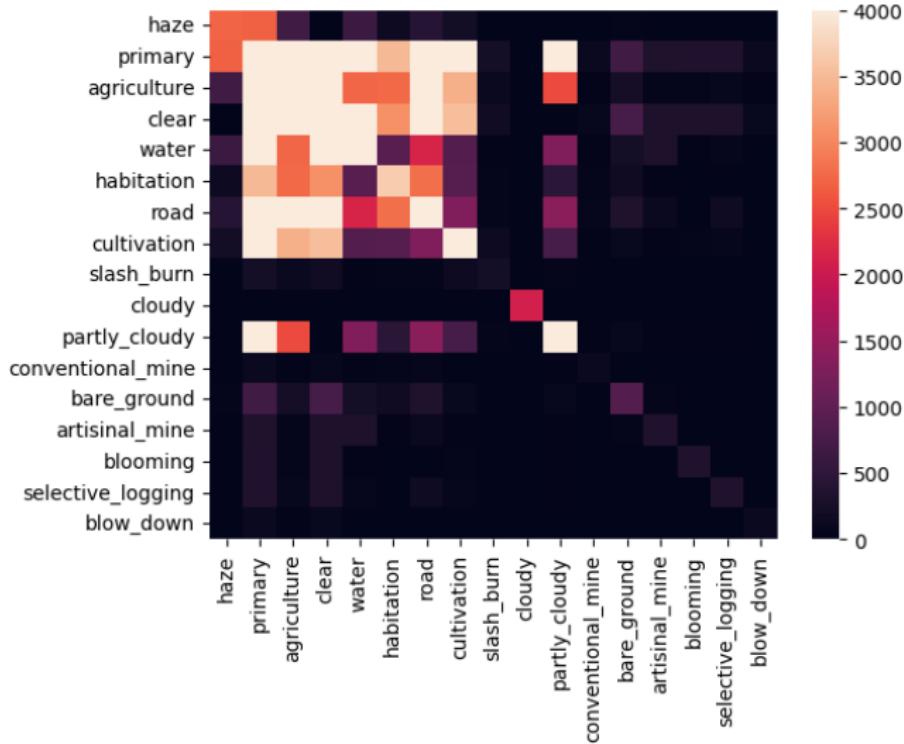


Figura: Matriz de co-ocorrência. Fonte: Autor

Análise de dados exploratória - Agrupamento de amostras



Figura: Agrupamento de amostras via técnica TSNE. Fonte: Autor

Pré-processamento, Transformação e carregamento

1. Carregamento dos canais RGB e descartando o de infravermelho-próximo de cada imagem.
2. Redimensionamento usando interpolação linear de $256 \times 256\text{px}$ para $224 \times 224\text{px}$. Isto se deve ao fato dos modelos já terem sido pré-treinados e configurados para essa dimensão de entrada.
3. Conversão da imagem para estrutura de dados numérica de Tensor
4. Aplicar espelho vertical ou horizontal, cada um com probabilidade de 25
5. Normalizar cada canal de cor RGB usando normalização Gaussiana com médias e desvio padrão do dataset ImageNet.

Definição do modelo

Nome	Pré-Treino	Resolution	Acurácia@1	Acurácia@5	Parâmetros	FLOPs
Swin-T	ImageNet-1K	224x224	81.474	95.5	28.3M	4.5G
Resnet50	ImageNet-1K	224x224	80.858	92.9	25.5M	4.1G

Seleção de modelos

1. Pesos pré-treinados
2. Aumentar capacidade da rede
3. Adicionar regularização
4. Amostrador
5. Aumento de dados aleatória
6. Função de perda
7. Otimizador
8. Taxas de aprendizado
9. Transferência de aprendizado vs Fine Tune
10. Experimentar os mesmos passos para o modelo Swin-T

Resultados e discussão

Resultados e discussão

Conclusão

 BALANIUK, R.; ISUPOVA, O.; REECE, S. Mining and tailings dam detection in satellite imagery using deep learning. *Sensors*, v. 20, n. 23, 2020. ISSN 1424-8220. Disponível em:

<<https://www.mdpi.com/1424-8220/20/23/6936>>. 19

 CAMALAN, S. et al. Change detection of amazonian alluvial gold mining using deep learning and sentinel-2 imagery. *Remote Sensing*, v. 14, n. 7, 2022. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/14/7/1746>>.

19

 DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

17

 INSTRUTORGIS. *QGIS: Satélite Amazonia-1 – Composição Colorida RGB*. 2022. Disponível em: <<https://www.instrutorgis.com.br/qgis-satelite-amazonia1-composicao-colorida-rgb>>.

4

 KASELIMI, M. et al. A vision transformer model for convolution-free multilabel classification of satellite imagery in

deforestation monitoring. *IEEE Transactions on Neural Networks and Learning Systems*, p. 1–9, 2022. ISSN 2162-2388.

19

 LIU, Z. et al. Swin transformer v2: Scaling up capacity and resolution. In: *CVPR 2022*. [s.n.], 2022. Disponível em: <<https://www.microsoft.com/en-us/research/publication/swin-transformer-v2-scaling-up-capacity-and-resolution/>>.

18

 MITCHELL, T. M. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN 978-0-07-042807-2.

5

 PLANETCO, K. *Planet: Understanding the Amazon from Space - Use satellite data to track the human footprint in the Amazon rainforests*. Kaggle, 2016. Disponível em: <<https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/>>.

20

 SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.*, v. 3, p. 210–229, 1959.

5

 WANG, D. et al. An empirical study of remote sensing pretraining. *arXiv preprint arXiv:2204.02825*, 2022.

19